

ANALISIS SENTIMEN TERHADAP TWEET MENGENAI PERISTIWA RUSSIA MELAWAN UKRAINA BERBASIS MACHINE LEARNING

Muhammad Husni Mubarok¹, Jati Sasongko Wibowo²

Universitas Stikubank Semarang, Jawa Tengah, Indonesia

Email: muhammadhusnimubarok@unisbank.ac.id¹, jatisw@edu.unisbank.ac.id²

Abstrak

Perang yang terjadi di Eropa antara Rusia - Ukraina berdampak langsung maupun tidak langsung diseluruh belahan dunia, termasuk Indonesia. Masyarakat banyak yang memberikan pendapat atas peristiwa tersebut, baik itu berupa pujian atau keluh kesah yang dipublikasikan pada media sosial, salah satunya adalah Twitter. Penelitian ini bertujuan untuk mengolah atau mengekstrak tanggapan masyarakat di media Twitter agar menjadi sebuah informasi dengan menggunakan analisis sentiment. Analisis sentiment dengan algoritma klasifikasi *Support Vector Machine* (SVM), *Multinomial Naive Bayes* (MNB), dan *K-Nearest Neighbor* (KKN). Berdasarkan analisis sentimen memberikan gambaran bahwa tanggapan masyarakat Indonesia pengguna twitter pada tanggal 24 Februari 2022 tentang peristiwa perang Rusia - Ukraina didominasi dengan tweet bersentimen negatif (50,1%), kemudian netral (30,5%), dan positif (19,3%). Hasil performa tertinggi model *machine learning* didapat pada model beralgoritma SVM dengan akurasi 90%, diikuti oleh MNB (71%), dan KKN (48%). Masyarakat Indonesia banyak yang mengaitkan peristiwa perang Rusia vs Ukraina dengan peristiwa perang dunia.

Kata kunci : Analisis Sentimen; Twitter; *Support Vector Machine*; *Multinomial Naive Bayes*; *K-Nearest Neighbor*

Abstract

The war that took place in Europe between Russia and Ukraine had a direct or indirect impact on all parts of the world, including Indonesia. Many people gave their opinion on the incident, whether it was in the form of praise or complaints published on social media, one of which was Twitter. This study aims to process or extract public responses on Twitter media to become information using sentiment analysis. Sentiment analysis using the classification algorithm Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), and K-Nearest Neighbor (KKN). Based on sentiment analysis, it illustrates that the response of the Indonesian public on Twitter users on February 24, 2022 about the events of the Russia-Ukraine war was dominated by tweets with negative sentiments (50.1%), then neutral (30.5%), and positive (19.3%). The highest performance results from the machine learning model were obtained on the SVM algorithm model with an accuracy of 90%, followed by MNB (71%), and KKN (48%). Many Indonesian people associate the events of the Russia vs Ukraine war with the events of the world war.

Keywords: *Sentiment Analysis; Twitter; Support Vector Machines; Naive Bayes Multinomial; K-Nearest Neighbor*

How to cite:	Mubarok, M. H., & Wibowo, J. S. (2024). Analisis Sentimen Terhadap Tweet Mengenai Peristiwa Russia Melawan Ukraina Berbasis Machine Learning. <i>Syntax Literate</i> . (9)2. http://dx.doi.org/10.36418/syntax-literate.v9i2
E-ISSN:	2548-1398
Published by:	Ridwan Institute

Pendahuluan

Pada saat ini, perang yang terjadi di Eropa antara Rusia melawan Ukraina berdampak di berbagai negara dibelahan penjurus dunia, tak terkecuali negara Indonesia (Hutabarat, 2022). Sektor ekonomi merupakan sektor yang sangat berdampak akibat perang yang terjadi antara Rusia melawan Ukraina (Bakrie et al., 2022; Rubel & Hossain, 2022). Dampak yang terjadi ialah mulai dari nilai tukar rupiah yang turun terhadap Dollar AS, kehilangan pendapatan dari ekspor terhadap kedua negara tersebut, kenaikan harga impor bahan pokok hingga naiknya harga minyak dunia (Alfiansyah, 2022). Dari fenomena itu, masyarakat banyak yang memberikan berbagai macam pendapat baik itu pujian atau keluh kesah yang dipublikasikan di berbagai media sosial, salah satunya adalah Twitter (Pamungkas & Kharisudin, 2021). Dari berbagai tanggapan tersebut perlu digolongkan ke dalam bentuk tanggapan positif, netral atau negatif sehingga di dapatkan suatu informasi dari tanggapan-tanggapan tersebut.

Salah satu cara mengolah atau mengekstrak tanggapan agar menjadi sebuah informasi adalah menggunakan analisis sentimen (Vidyadhari, 2022). Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tesktual secara otomatis untuk mendapat informasi (Pang & LillianLee, 2008). Dalam menerapkan analisis sentimen diperlukan adanya sebuah klasifikasi. Klasifikasi digunakan untuk mengelompokkan tanggapan mana yang memiliki sentimen positif, sentiment netral, dan sentimen negative (Salam et al., 2018). Penelitian mengenai analisis sentimen di Indonesia telah banyak dilakukan dengan berbagai metode. (Setiawan et al., 2021) menyimpulkan bahwa *multinomial naive bayes* dapat digunakan untuk klasifikasi teks dengan akurasi yaitu sebesar 82%. Namun dalam penelitian (Shofiya et al., 2020) menyimpulkan bahwa algoritma *support vector machine* memiliki nilai akurasi sebesar 82% sedangkan *multinomial naive bayes* memiliki nilai akurasi sebesar 64%. Sedangkan menurut (Rezwanul et al., 2017) dalam pengklasifikasian data tweet menyimpulkan bahwa algoritma *K-nearest neighbor* memiliki tingkat akurasi tertinggi dibandingkan dengan algoritma *support vector machine*.

Dalam penerapannya algoritma pada *support vector machine* memiliki keunggulan dalam mengklasifikasikan data berbentuk teks, diantara *Support Vector Machine*, *Multinomial Naive Bayes*, serta *K-Nearest Neighbor* algoritma mana yang paling tepat mengklasifikasikan data tanggapan masyarakat Indonesia pada media sosial twitter (Devi, 2023).

Pada penelitian ini akan melakukan analisis sentimen tanggapan masyarakat terhadap perang antara Rusia melawan Ukraina pada media sosial Twitter dengan algoritma klasifikasi *Support Vector Machine*, *Multinomial Naive Bayes*, serta *K-Nearest Neighbor*.

Metode Penelitian

Pada tahapan yang pertama, penulis melakukan pembelajaran beberapa literatur yang ada (Creswell & Poth, 2016). Kemudian peneliti melakukan pengumpulan data, data yang dikumpulkan merupakan data yang berasal dari twitter yaitu data tweets. Setelah melalui tahapan pengumpulan data, peneliti melanjutkan ke tahap pre-processing data, pemberian bobot, visualisasi data, dan transformasi ke bentuk TF-IDF (Sugiyono, 2019). Setelah itu data yang telah didapatkan akan diolah menggunakan metode klasifikasi *Support Vector Machine*, *K-Nearest Neighbor*, dan *Multinomial Naive Bayes*.

Pada tahap pengumpulan data peneliti mengumpulkan data berupa tweet dari media sosial yaitu *twitter* terkait dengan perang yaitu perang antara rusia vs ukraina dengan keyword “Rusia vs Ukraina”, “Russia”, “Rusia”, “Ukraina”, “Ukraine”, “Putin”, dan “Zelensky”, “Kyiv”. Tweet yang dikumpulkan merupakan tweet berbahasa Indonesia dengan batasan tweet dibuat hanya pada tanggal 24 Februari 2022. Proses pencarian dan pengumpulan data dapat dilakukan dengan menggunakan fitur API *twitter* (*application programming interface*).

Setelah menyelesaikan tahap proses pengambilan dan pengumpulan data, peneliti selanjutnya melanjutkan ke tahap *pre-processing* data. Pada tahap *pre-processing* data menyiapkan data mentah yang diperoleh kembali dalam format yang sesuai atau data untuk melakukan analisis. Tahapan yang harus dilakukan saat melakukan *pre-processing* data adalah *cleaning text*, *to lower case text*, *tokenize text*, *slang word transform*, *remove stopword*, *stemming*, dan *remove duplicates*.

Pada tahap pemberian pembobotan, peneliti menentukan label dari tweet yang lolos pada tahap *pre-processing* sebelumnya. Peneliti menggunakan metode perhitungan berbasis *lexicon* untuk melakukan pelabelan yang mencakup negatif, netral, dan positif. Untuk menentukan sentimen positif atau negatif dari sebuah kalimat menggunakan teknik menghitung *polarity score* berdasarkan nilai-nilai sentimen dari kamus *lexicon*.

Setelah data sudah diberi pembobotan menggunakan metode *lexicon*, selanjutnya data akan diubah bentuknya kedalam bentuk TF-IDF untuk keperluan ekstraksi fitur. Setelah itu data akan digunakan untuk proses klasifikasi atau proses training model.. Proses ini dilakukan dengan tiga metode, yaitu *Support Vector Machine*, *K-Nearest Neighbor*, dan *Multinomial Naïve Bayes*. Setelah itu model di uji dengan data test serta menampilkan *performance matrix* dan *confusion matrix*nya.

Hasil dan Pembahasan

Data yang dicari dan digunakan untuk referensi terkait dengan perang yaitu perang antara rusia vs ukraina dengan kata “Rusia vs Ukraina”, “Russia”, “Rusia”, “Ukraina”, “Ukraine”, “Putin”, dan “Zelensky”, “Kyiv”, dan hanya pada tweet yang dibuat pada tanggal 24 Februari 2022 saja bertepatan Rusia melakukan invasi terhadap Ukraina. Pengambilan data dilakukan dengan menggunakan fitur API *twitter* versi 2 yang didapatkan dengan mengajukan status *academic research* pada *twitter*. Untuk mengatasi perihal limit dalam pemanggilan API *twitter*, penelitian ini menggunakan library *twar2* agar dapat memudahkan dalam menyimpan data tweet yang didapat kedalam file CSV. Dengan keyword yang ada didapat data tweet sebanyak 78999 dapat dilihat pada Gambar 1.

	id	created_at	text
0	1.496998e+18	2022-02-24 00:00:00	@jokowi Rasah melu melunRakyatmu ngantri minya...
1	1.496998e+18	2022-02-24 00:00:00	KRONOLOGI RINGKAS APA YANG BERLAKU ANTARA RUSI...
2	1.496998e+18	2022-02-24 00:00:00	Krisis Ukraine-Rusia: Apa punca sebenar?nn[Ini...
3	1.496998e+18	2022-02-24 00:00:00	Kiamat Urusan TuhannTapi Jika Barat Ingin Kiam...
4	1.496998e+18	2022-02-24 00:00:00	@CNNIndonesia Emang dianggap sm mr Putin nguru...
...
78994	1.496636e+18	2022-02-24 00:00:00	@Green_Nord27 @AremafcOfficial Waduu kenapa ga...
78995	1.496636e+18	2022-02-24 00:00:00	Berita negatif bursa:nnRusia invasi Ukraina, t...
78996	1.496636e+18	2022-02-24 00:00:00	@GlcnC Putin dedi:(
78997	1.496636e+18	2022-02-24 00:00:00	Menurut gue ini pertanyaan penting karena kalo...
78998	1.496636e+18	2022-02-24 00:00:00	Saat Conor McGregor Salah Tingkah karena Terla...

78999 rows × 3 columns

Gambar 1. Data Mentah Hasi Pencarian dari Twarc2

Setelah menyelesaikan tahap proses pengambilan dan pengumpulan data, peneliti selanjutnya melanjutkan ke tahap *pre-processing* data. Pada tahap *pre-processing* data menyiapkan data mentah yang diperoleh kembali dalam format yang sesuai atau data untuk melakukan analisis. Dengan beberapa tahapan meliputi *cleaning text*, *to lower case text*, *tokenize text*, *slang word transform*, *remove stopword*, *stemming*, dan *remove duplicates*. Hasil dari proses dapat dilihat pada Gambar 2 dibawah ini. Dari data awal sebanyak 78999 setelah dilakukan preprocessing, data berkurang menjadi 73628.

	id	created_at	text_clean	text_preprocessed
0	1,50E+29	2022-02-24T14:16:28.000Z	rusia dah perang kt msh kepopan	[rusia, deh, perang, pop]
1	1,50E+29	2022-02-24T15:07:56.000Z	kalian pd panik perang rusiaukraine pdhl sampe...	[panik, perang, rusiaukraine, pdhl, palestina...
2	1,50E+29	2022-02-24T06:15:17.000Z	via putin mengancam negara lain yang berniat...	[via, putin, ancam, negara, niat, intervensi, ...
3	1,50E+30	2022-02-24T06:36:52.000Z	gsp perspektif dari kubu rusia	[gsp, perspektif, kubu, rusia]
4	1,50E+30	2022-02-24T08:11:24.000Z	sumpahh udh meledak di ukraina gila bgt iniii	[sumpahh, ledak, ukraina, gila, banget, iniii]
...
73623	1,50E+34	2022-02-24T20:39:10.000Z	ini pernyataan pemerintah indonesia atas sera...	[nyata, perintah, indonesia, serang, rusia, uka...
73624	1,50E+34	2022-02-24T20:39:10.000Z	emang dianggap sm mr putin ngurus minyak goreng...	[emang, anggap, mr, putin, urus, minyak, goreng...
73625	1,50E+34	2022-02-24T20:39:21.000Z	kiamat urusan tuhan tapi jika barat ingin kiam...	[kiamat, urus, tuhan, barat, kiamat, urus, vla...
73626	1,50E+34	2022-02-24T20:39:24.000Z	krisis ukrainerusia apa punca sebenar ini ada...	[krisis, ukrainerusia, punca, benar, bebenang]
73627	1,50E+34	2022-02-24T20:39:25.000Z	kronologi ringkas apa yang berlaku antara rusi...	[kronologi, ringkas, laku, rusia, ukraina, iku...

73628 rows × 4 columns

Gambar 2. Hasil Preprocessing Data Mentah

Data hasil preprocessing akan diberi pembobotan yang bertujuan untuk menilai nilai sentiment dan memberi label yang terbagi menjadi positif, negatif, dan netral. Memberi bobot sentiment menggunakan metode perhitungan berbasis *lexicon* dengan konsep menghitung *score* per kata berdasarkan kamus *lexicon* dari InSet. *Score* pada kamus *lexicon* InSet memiliki range -5 sampai +5 dan kamus terdiri dari 3.609 kata positif dan 6.609 kata negatif.

Setiap data tweet akan dihitung total *polarity score*-nya berdasarkan hasil jumlah nilai setiap kata pada tweet yang ada pada kamus *lexicon* dari InSet. Di sini S diartikan

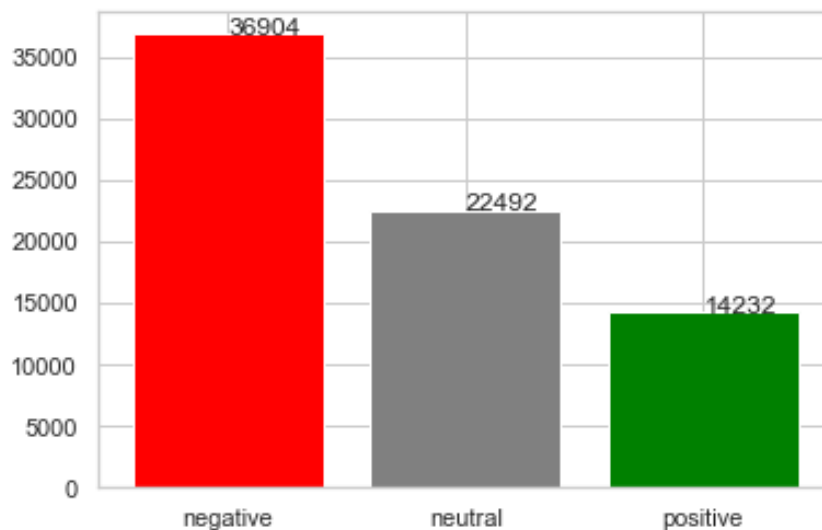
sebagai *polarity score*, jadi jika S lebih besar dari 0 maka S positif, jika S lebih kecil dari 0 maka S negatif dan jika S sama dengan 0 maka bersifat netral. Hasil dari proses ini dapat dilihat pada Gambar 3 dibawah ini.

id	text_preprocessed	polarity_score	polarity
0	[rusia, deh, perang, pop]	2	positive
1	[panik, perang, rusiaukraine, pdhl, palestina...]	-13	negative
2	[via, putin, ancam, negara, niat, intervensi, ...]	-5	negative
3	[gsp, perspektif, kubu, rusia]	0	neutral
4	[sumpahh, ledak, ukraina, gila, banget, iniii]	-5	negative
...
73623	[nyata, perintah, indonesia, serang, rusia, uk...]	-2	negative
73624	[emang, anggap, mr, putin, urus, minyak, goren...]	-4	negative
73625	[kiamat, urus, tuhan, barat, kiamat, urus, vla...]	-3	negative
73626	[krisis, ukrainerusia, punca, benar, bebenang]	-6	negative
73627	[kronologi, ringkas, laku, rusia, ukraine, iku...]	-6	negative

73628 rows x 4 columns

Gambar 3. Hasil Pembobotan dan Pelabelan

Setelah dilakukan tahap preprocessing dan pelabelan, data mentah yang didapat dari API Twitter dapat dianggap sebagai dataset. Dalam tahap ini dataset akan diekstrak seluruh informasi sehingga dapat diketahui kata yang paling sering dibahas pada ulasan pengguna. Selain itu juga dapat memberikan gambaran tentang dominasi sentimen yang dipaparkan pengguna media sosial twitter terhadap peristiwa Rusia vs Ukraina.

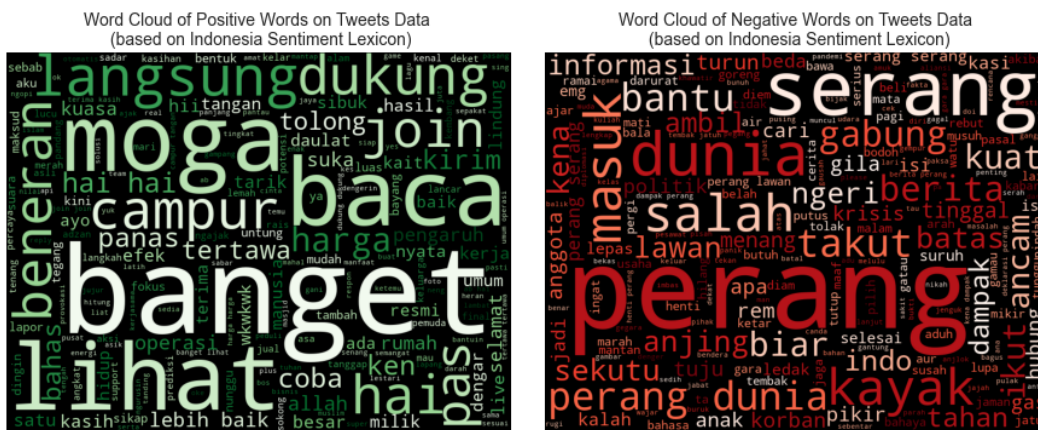


Gambar 4. Visualisasi Bar Chart Dataset

Pada Gambar 4 dari total data sebanyak 73628 data, dapat terlihat bahwa data tweets yang memiliki sentiment negatif lebih banyak jika dibandingkan dengan data yang memiliki sentimen netral dan positif. Dengan rincian data yang memiliki sentiment negatif berjumlah sebanyak 36904 data tweets, kemudian untuk data yang memiliki sentiment netral berjumlah sebanyak 22492 data tweets, dan yang terakhir adalah data yang memiliki sentimen positif berjumlah sebanyak 14232 data tweets.

Berdasarkan Gambar 6, tweet dari pengguna media sosial twitter terhadap peristiwa perang Rusia vs Ukraina dikaitkan dengan beberapa kata atau term. Kata atau term yang dikaitkan antara lain ‘perang dunia’, ‘world war’, ‘us’, ‘nato’, ‘belarus’, ‘china’, ‘uni soviet’, dan ‘indonesia’. Tidak secara langsung pengguna twitter mengkaitkan peristiwa perang Rusia vs Ukraina dengan peristiwa perang dunia. Selain itu terdapat pihak-pihak yang menurut tweet yang didapat, dikaitkan dengan peristiwa ini yaitu antara lain Rusia, Ukraina, NATO, Amerika, Uni Soviet, Belarus, China, dan Indonesia.

Word cloud juga dapat divisualisasikan secara terpisah berdasarkan data tweet bersentimen positif dan negatif. Dengan memisahkan data tweet berdasarkan sentimennya, dapat memberikan gambaran tentang apa saja yang sering muncul dari tweet pengguna media sosial twitter bersentimen positif maupun negative. Berikut word cloud sentimen positif dan negatif dari dataset dapat dilihat pada Gambar 7 dibawah ini.



Gambar 7. Word Cloud Dataset Bersentimen Positif dan Negatif

Proses selanjutnya adalah proses transformasi TF-IDF, proses ini dilakukan dengan beberapa proses. Tujuan proses ini adalah mengubah format dataset yang dapat dipelajari oleh algoritma machine learning. Konsep TF-IDF secara garis besar mengukur pentingnya atau relevansi representasi dari sebuah kata atau terms dalam dokumen atau kumpulan dokumen yang biasa dikenal sebagai corpus, dalam kasus ini corpus adalah representasi dari tweet. Sebelum ke proses transformasi TF-IDF, dataset dipecah terlebih dahulu menjadi data train dan data test. Data train berguna sebagai data untuk melatih model, sedangkan data test berguna sebagai data untuk menguji atau mengevaluasi model. Pembagian data train dan data test adalah 70:30, dimana data train berisi 70% (49330 data) dan data test berisi 30% (24298 data) dari total data di dataset. Pembagian data menggunakan fungsi `train_test_split` dari library `sklearn` agar pembagian data akan merata berdasarkan labelnya.

Untuk transformasi bentuk dataset menjadi bentuk TF-IDF, menggunakan fungsi `TfidfVectorizer` dari library `sklearn` bagian feature extraction. Dengan parameter `max_feature` bernilai 5000, yang berarti dimensi data yang dibuat adalah matrix berukuran $n \times 5000$, dimana n adalah total corpus atau tweet. Pada fungsi `TfidfVectorizer`, `max_feature` merupakan limitasi total kata atau term dari frekuensi terbanyak, jika `max_feature` bernilai 5000 maka TF-IDF hanya memilih 5000 kata yang sering muncul pada keseluruhan corpus atau tweet sebagai fitur. Jadi setelah proses

ekstraksi fitur TF-IDF, dimensi data train akan menjadi 49330 x 5000 dan dimensi data test akan menjadi 24298 x 5000. Hasil ekstraksi fitur menggunakan TF-IDF terhadap data train dapat dilihat pada Gambar 8.

	aa	aache	aaj	aamiin	aan	aap	aapke	aapki	aapko	aaya	...	zenit	zetu	zile
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
...
49325	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
49326	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
49327	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
49328	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
49329	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0

49330 rows x 5000 columns

Gambar 8. Hasil Transformasi TF-IDF pada Data Train

Klasifikasi pada penelitian ini melakukan training pada model dengan data train yang berjumlah 49330 data. Terdapat tiga metode yang digunakan pada penelitian ini antara lain *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), dan *Multinomial Naïve Bayes* (MNB). Model akan ditrain dengan parameter default dari library yang digunakan yaitu *sklearn*. Setelah itu Model akan diuji menggunakan data test berjumlah 24298 data yang bertujuan untuk menebak label dari data test tersebut. Untuk menghitung performa, label data test akan kalkulasi dengan label hasil prediksi sehingga bisa mendapatkan metrik performa seperti recall, precision, f1-score dan accuracy.

Proses melatih model SVM dengan parameter default menggunakan data train sebanyak 49330 data memerlukan waktu 314 detik. Setelah model dilatih, model akan diuji untuk dapat memprediksi data, pada model SVM waktu untuk melakukan prediksi terhadap data test sebanyak 24298 data adalah 53 detik dapat dilihat pada Gambar 9.


```

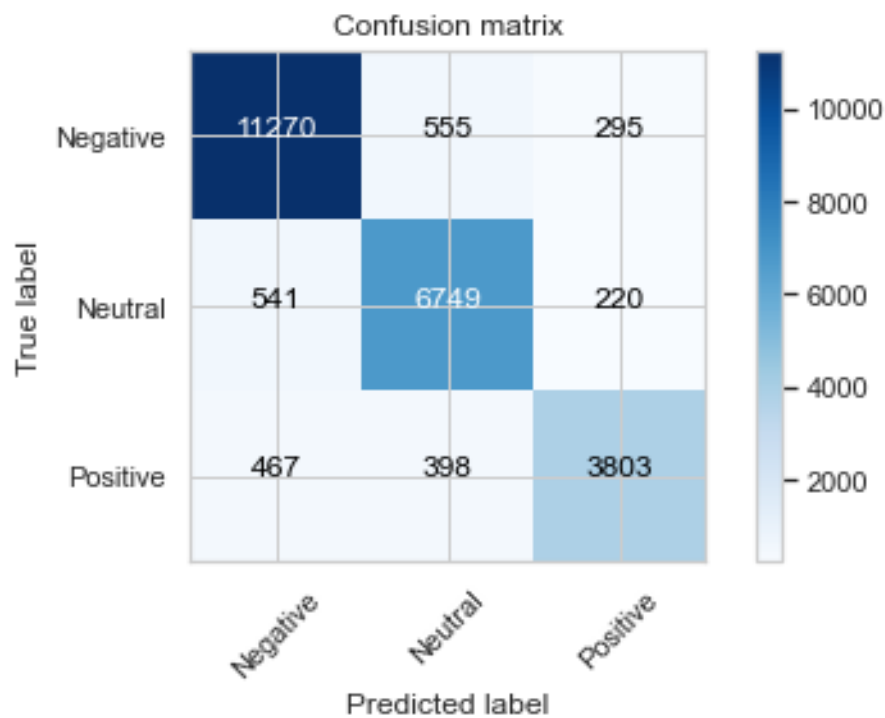
Training time: 314.516204s; Prediction time: 53.713498s
      precision    recall  f1-score   support

negative    0.92     0.93     0.92    12120
neutral     0.88     0.90     0.89     7510
positive    0.88     0.81     0.85     4668

accuracy                0.90    24298
macro avg    0.89     0.88     0.89    24298
weighted avg 0.90     0.90     0.90    24298
    
```

Gambar 9. Metrik Performa Model SVM

Pada Gambar 9 memperlihatkan tentang hasil metrik performa yang diperoleh oleh model SVM dari pengujian dengan data test. Model SVM dapat memprediksi label dari ketiga kelas dengan akurasi 90%, bermakna 10% dari total 24298 data test telah terprediksi dengan label yang salah dan sisanya benar. Dengan detail untuk metrik performa f1-score yang dihasilkan dari perhitungan harmonic mean dari precision yang dan recall, dari label negatif bernilai 92%, netral 89%, dan positif 85%. Prediksi label negatif memiliki f-1 score tertinggi yaitu 92%, dan prediksi label positif memiliki f-1 score terendah yaitu 85%. Menandakan model SVM sangat bagus dalam memprediksi tweet bersentimen negatif dibanding tweet bersentimen positif maupun netral. Untuk detail hasil prediksi dari data test dapat dilihat dengan confusion matrix yang terdapat pada Gambar 10 dibawah ini.



Gambar 10. Confusion Matrix Model SVM

Pada Gambar 10 memperlihatkan detail nilai hasil prediksi yang terbukti benar dan yang salah dari model SVM. Dari 12120 data test berlabel negatif, model dapat memprediksi dengan benar sebanyak 11270 data, terprediksi salah sebagai label netral sebanyak 555 data, dan terprediksi salah sebagai label positif sebanyak 295 data. Dari 7510 data test berlabel netral, model dapat memprediksi dengan benar sebanyak 6749 data, terprediksi salah sebagai label negatif sebanyak 541 data, dan terprediksi salah sebagai label positif sebanyak 220 data. Dari 4668 data test berlabel positif, model dapat memprediksi dengan benar sebanyak 3803 data, terprediksi salah sebagai label negatif sebanyak 467 data, dan terprediksi salah sebagai label netral sebanyak 398 data.

Proses melatih model KNN dengan parameter default menggunakan data train sebanyak 49330 data memerlukan waktu 0.06 detik. Setelah model dilatih, model akan diuji untuk dapat memprediksi data, pada model KNN waktu untuk melakukan prediksi terhadap data test sebanyak 24298 data adalah 32 detik dapat dilihat pada Gambar 11.

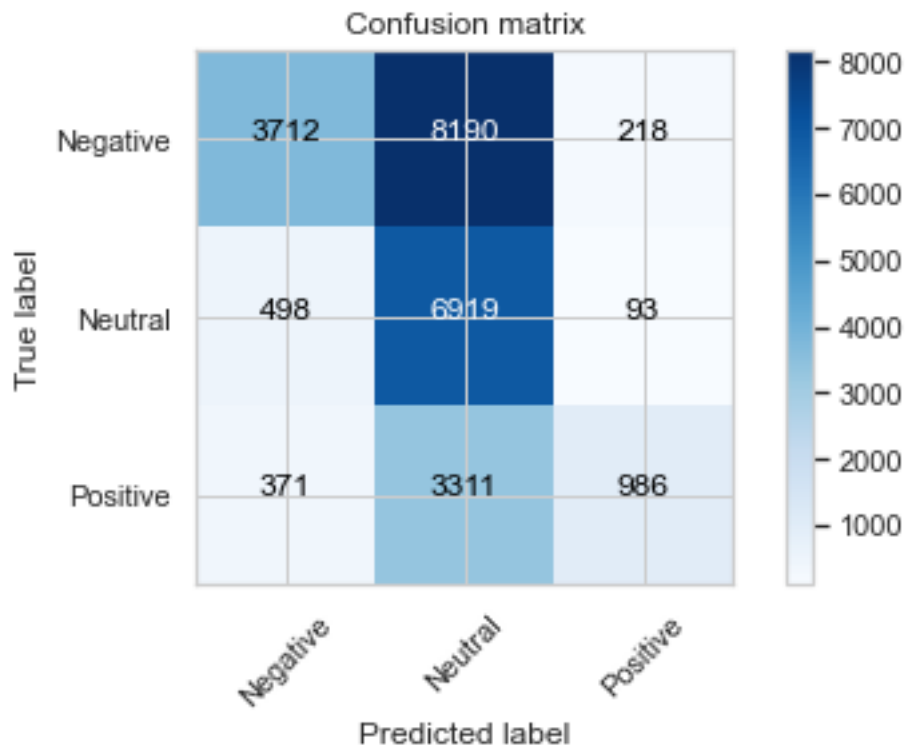
```
Training time: 0.066017s; Prediction time: 32.873969s
      precision    recall  f1-score   support

negative    0.81     0.31     0.44     12120
neutral     0.38     0.92     0.53     7510
positive    0.76     0.21     0.33     4668

accuracy                    0.48     24298
macro avg    0.65     0.48     0.44     24298
weighted avg 0.67     0.48     0.45     24298
```

Gambar 11. Metrik Performa Model KNN

Pada Gambar 11 memperlihatkan tentang hasil metrik performa yang diperoleh oleh model KNN dari pengujian dengan data test. Model KNN dapat memprediksi label dari ketiga kelas dengan akurasi 48%, bermakna 52% dari total 24298 data test telah terprediksi dengan label yang salah dan sisanya benar. Dengan detail untuk metrik performa f1-score yang dihasilkan dari perhitungan harmonic mean dari precision yang dan recall, dari label negatif bernilai 44%, netral 53%, dan positif 33%. Prediksi label netral memiliki f-1 score tertinggi yaitu 53%, dan prediksi label positif memiliki f-1 score terendah yaitu 33%. Menandakan model KNN sangat bagus dalam memprediksi tweet bersentimen netral dibanding tweet bersentimen positif maupun negatif. Untuk detail hasil prediksi dari data test dapat dilihat dengan confusion matrix yang terdapat pada Gambar 12 dibawah ini.



Gambar 12. Confusion Matrix Model KNN

Pada Gambar 12 memperlihatkan detail nilai hasil prediksi yang terbukti benar dan yang salah dari model KNN. Dari 12120 data test berlabel negatif, model dapat memprediksi dengan benar hanya sebanyak 3712 data, terprediksi salah sebagai label netral sebanyak 8190 data, dan terprediksi salah sebagai label positif sebanyak 218 data. Dari 7510 data test berlabel netral, model dapat memprediksi dengan benar sebanyak 6919 data, terprediksi salah sebagai label negatif sebanyak 498 data, dan terprediksi salah sebagai label positif sebanyak 93 data. Dari 4668 data test berlabel positif, model dapat memprediksi dengan benar hanya sebanyak 966 data, terprediksi salah sebagai label negatif sebanyak 371 data, dan terprediksi salah sebagai label netral sebanyak 3311 data.

Proses melatih model MNB dengan parameter default menggunakan data train sebanyak 49330 data memerlukan waktu 0.16 detik. Setelah model dilatih, model akan diuji untuk dapat memprediksi data, pada model MNB waktu untuk melakukan prediksi terhadap data test sebanyak 24298 data adalah 0.0004 detik dapat dilihat pada Gambar 13.

```

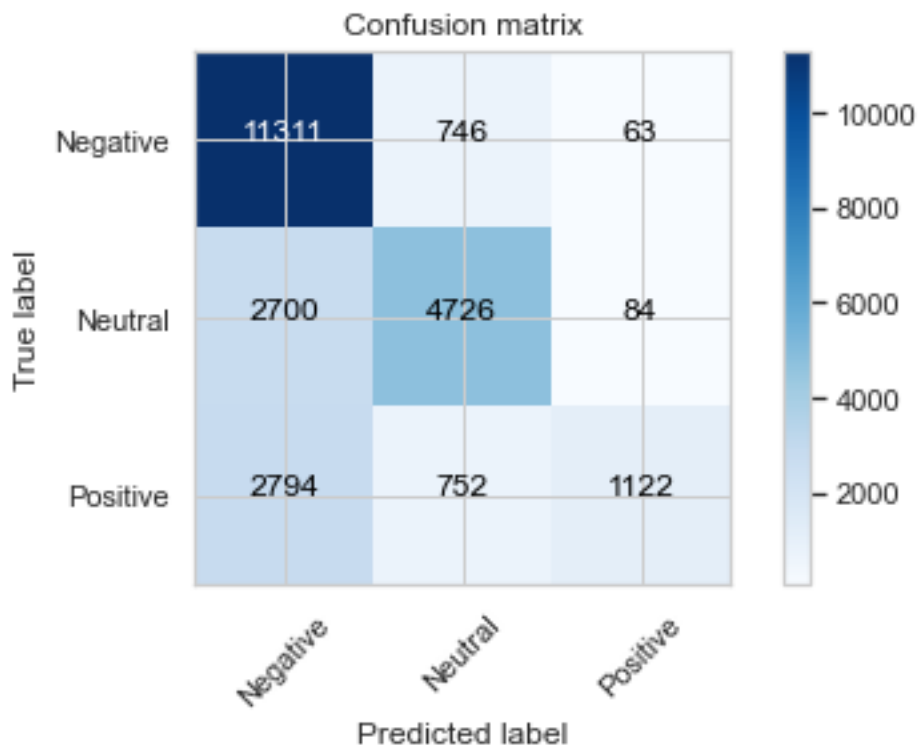
Training time: 0.162035s; Prediction time: 0.004006s
      precision    recall  f1-score   support

negative    0.67     0.93     0.78     12120
neutral     0.76     0.63     0.69      7510
positive    0.88     0.24     0.38      4668

accuracy                0.71     24298
macro avg    0.77     0.60     0.62     24298
weighted avg 0.74     0.71     0.68     24298
    
```

Gambar 13. Metrik Performa Model MNB

Pada Gambar 13 memperlihatkan tentang hasil metrik performa yang diperoleh oleh model MNB dari pengujian dengan data test. Model MNB dapat memprediksi label dari ketiga kelas dengan akurasi 71%, bermakna 29% dari total 24298 data test telah terprediksi dengan label yang salah dan sisanya benar. Dengan detail untuk metrik performa f1-score yang dihasilkan dari perhitungan harmonic mean dari precision yang dan recall, dari label negatif bernilai 78%, netral 69%, dan positif 39%. Prediksi label negatif memiliki f-1 score tertinggi yaitu 78%, dan prediksi label positif memiliki f-1 score terendah yaitu 33%. Menandakan model MNB sangat bagus dalam memprediksi tweet bersentimen negatif dibanding tweet bersentimen positif maupun netral. Untuk detail hasil prediksi dari data test dapat dilihat dengan confusion matrix yang terdapat pada Gambar 14 dibawah ini.



Gambar 14. Confusion Matrix Model MNB

Pada Gambar 14 memperlihatkan detail nilai hasil prediksi yang terbukti benar dan yang salah dari model MNB. Dari 12120 data test berlabel negatif, model dapat memprediksi dengan benar sebanyak 11311 data, terprediksi salah sebagai label netral sebanyak 746 data, dan terprediksi salah sebagai label positif sebanyak 63 data. Dari 7510 data test berlabel netral, model dapat memprediksi dengan benar sebanyak 4726 data, terprediksi salah sebagai label negatif sebanyak 2700 data, dan terprediksi salah sebagai label positif sebanyak 84 data. Dari 4668 data test berlabel positif, model dapat memprediksi dengan benar sebanyak 1122 data, terprediksi salah sebagai label negatif sebanyak 2794 data, dan terprediksi salah sebagai label netral sebanyak 752 data.

Pembahasan

Setelah melakukan penelitian serta analisis sentimen terhadap peristiwa perang Rusia vs Ukraina berbasis lexicon dan menggunakan metode machine learning *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), dan *Multinomial Naïve Bayes* (MNB), maka pada bagian yang terakhir ini, peneliti akan melakukan penjelasan tentang kesimpulan yang terkait dengan analisis sentimen yang telah dikerjakan.

Hasil berdasarkan analisis sentimen memberikan gambaran bahwa tanggapan masyarakat Indonesia pengguna twitter pada tanggal 24 Februari 2022 tentang peristiwa perang Rusia vs Ukraina didominasi dengan tweet bersentimen negatif (Aditya & Wibowo, 2022). Dengan tweet bersentimen negatif memiliki presentase terbanyak yaitu 50.1%, tweet bersentimen netral memiliki persentase 30.5%, dan tweet bersentimen positif memiliki persentase 19.3%. Hasil persentase sentimen diatas didapat dari perhitungan nilai sentimen berbasis lexicon dengan total data tweet terkumpul sebanyak 73628 data.

Dari data tweet yang didapat, masyarakat Indonesia banyak yang mengaitkan peristiwa perang Rusia vs Ukraina dengan peristiwa perang dunia. Terdapat juga beberapa pihak yang sering disebut pada tweet yang didapat, seperti Rusia, Ukraina, Amerika, Nato, Belarus, Uni Soviet, China dan Indonesia. Keterkaitan ini didapat dari hasil perhitungan setiap kata atau term yang ada pada keseluruhan data tweet. Sehingga subjek yang diduga dikaitkan dengan peristiwa ini merupakan kata atau term yang memiliki frekuensi tinggi dan tampil pada visualisasi word cloud.

Hasil performa tertinggi model machine learning didapat pada model beralgoritma SVM dengan akurasi 90%, posisi kedua didapat pada model MNB dengan akurasi 71%, dan diposisi terakhir didapat pada model KNN dengan akurasi hanya 48%. Hasil metrik performa khususnya akurasi didapat dari model yang sudah di train dengan data train dan diuji dengan data test. Data tersebut sebelumnya diolah dengan menggunakan beberapa preprocessing dan juga menerapkan ekstraksi fitur menggunakan TF-IDF (Gifari et al., 2022).

Dari waktu komputasi melatih model dengan menggunakan data train sebanyak 49330 data, model KNN memiliki waktu komputasi tercepat yaitu hanya 0.06 detik. Sedangkan model MNB memerlukan 0.16 detik dan model SVM memerlukan waktu terlama yaitu 314 detik. Sedangkan Dari waktu komputasi menguji model dengan menggunakan data test sebanyak 24298 data, model MNB memiliki waktu komputasi tercepat yaitu hanya 0.004 detik. Sedangkan model KNN memerlukan 32 detik dan model SVM memerlukan waktu terlama yaitu 53 detik.

Kesimpulan

Berdasarkan analisis sentimen terhadap peristiwa perang Rusia vs Ukraina berdasarkan data tweet dari masyarakat Indonesia pada tanggal 24 Februari 2022, hasil menunjukkan dominasi tweet bersentimen negatif dengan presentase tertinggi mencapai 50.1%, diikuti oleh tweet bersentimen netral (30.5%) dan positif (19.3%). Banyak masyarakat mengaitkan peristiwa tersebut dengan perang dunia, dan beberapa pihak seperti Rusia, Ukraina, Amerika, dan lainnya sering disebut dalam tweet. Model machine learning menunjukkan performa tertinggi pada SVM dengan akurasi 90%, disusul oleh MNB (71%), dan KNN (48%). Waktu komputasi tercepat untuk melatih model dimiliki oleh KNN (0.06 detik), sedangkan MNB (0.16 detik) dan SVM (314 detik) memerlukan waktu lebih lama. Untuk pengujian model, MNB juga memiliki waktu komputasi tercepat (0.004 detik), diikuti oleh SVM (53 detik) dan KNN (32 detik). Dengan demikian, kesimpulannya adalah model SVM memiliki kinerja terbaik dalam analisis sentimen terhadap peristiwa tersebut, meskipun waktu komputasinya lebih lama dibandingkan dengan model lainnya.

BIBLIOGRAFI

- Aditya, A., & Wibowo, A. (2022). Analisis Sentimen Menggunakan Metode Naïve Bayes Berdasarkan Opini Masyarakat Dari Twitter Terhadap Perang Rusia dan Ukraina. *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 1(1), 551–558.
- Alfiansyah, T. R. (2022). *5 Dampak Perang Rusia-Ukraina Bagi Indonesia*. Kompas.Com.
- Bakrie, C. R., Delanova, M. O., & Yani, Y. M. (2022). Pengaruh perang Rusia dan Ukraina terhadap perekonomian negara kawasan Asia Tenggara. *Caraka Prabhu: Jurnal Ilmu Pemerintahan*, 6(1), 65–86.
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- Devi, R. F. (2023). *Analisis Sentimen Terhadap Kinerja Pelayanan Di Pt Bank Rakyat Indonesia (Persero) Tbk. Menggunakan Metode Support Vector Machine, Naive Bayes, Dan K-Nearest Neighbors*.
- Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40.
- Hutabarat, G. F. I. (2022). Arah Kebijakan Luar Negeri Indonesia Pasca Perang Rusia-Ukraina Berdasarkan Perspektif National Interest. *Jurnal Al Azhar Indonesia Seri Ilmu Sosial E-ISSN*, 2745, 5920.
- Kusumawati, E. (2023). Kepemimpinan Digital dalam Pendidikan: Sebuah Analisis Bibliometrik. *Journal of Education and Teaching (JET)*, 4(2), 252-260.
- Nandana, Y., Runturambi, A. J. S., & Daryanto, E. (2023). The Turmoil of Europe's Economy, Food and Energy Stability: Will It Ends Western Sanctions on Russia?. *International Journal of Social Service and Research*, 3(5), 1290-1302.
- Pamungkas, F. S., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 628–634.

- Pang, B., & LillianLee. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 1–2(2), 1–135. <https://doi.org/10.1561/1500000001>
- Rezwanul, M., Ali, A., & Rahman, A. (2017). Sentiment Analysis on Twitter Data using KNN and SVM. *International Journal of Advanced Computer Science and Applications*, 8(6), 19–25. <https://doi.org/10.14569/ijacsa.2017.080603>
- Rubel, M., & Hossain, M. S. (2022). Analyze The Causes of The Russian-Ukraine War with Waltz's Three Images (Individual, Domestic Politics, International System). *Journal of Social Science*, 3(5), 1113-1121.
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). *Analisis Sentimen Data Komentar Sosial Media Facebook Dengan k-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)*.
- Setiawan, A. R., Rintyama, B. S., & Cahyanto, T. A. (2021). *Preferensi Konsumen Terhadap Produk by.U Dan MPWR Dengan Analisis Sentimen berbasis Multinomial Naive Bayes*.
- Shofiya, F., Arifianto, D., & Faruq, H. A. Al. (2020). *Perbandingan Algoritma Support Vector Machine (SVM) Dan Multinomial Naive Bayes (MNB) Dalam Klasifikasi Abstrak Tugas Akhir*.
- Sugiyono. (2019). *Metode Penelitian*. CV Alfabeta.
- Vidyadhari, D. C. (2022). *Analisis Sentimen Konflik Rusia-Ukraina dalam Kanal Berita Online Indonesia Menggunakan Pendekatan Berbasis Lexicon*.

Copyright holder:

Muhammad Husni Mubarak, Jati Sasongko Wibowo (2024)

First publication right:

Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under:

