



ANALISIS SENTIMEN PUBLIK TERHADAP AKSI DEMONSTRASI DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE DAN RANDOM FOREST

Ni Made Tara Okta Adriana^{1)*}, I Made Agus Dwi Suarjaya¹⁾, Dwi Putra Githa¹⁾

Universitas Udayana, Badung, Indonesia

Email: taraadriana45@gmail.com

Abstrak

Indonesia adalah sebuah negara demokrasi yang memberikan kebebasan bagi setiap warganya untuk bereksresi dan mengeluarkan pendapat, salah satunya melalui demonstrasi. Namun, kadang kala kebebasan tersebut disalahgunakan oleh pihak-pihak tertentu yang menyebabkan adanya demonstrasi yang diikuti oleh kerusuhan dan aksi anarkis, sehingga menimbulkan berbagai macam kerugian. Penelitian ini bertujuan untuk mengetahui pandangan masyarakat terkait aksi demonstrasi yang terjadi di Indonesia melalui *tweets* pada media sosial Twitter dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest*. Pada penelitian ini terdapat penambahan proses untuk menghilangkan kata-kata yang tidak terdaftar pada Kamus Besar Bahasa Indonesia (KBBI) karena *tweets* pada media sosial Twitter cenderung lebih banyak menggunakan bahasa gaul atau bahasa yang tidak baku, sehingga penambahan proses ini diharapkan dapat memberikan data bersih yang lebih baik. Klasifikasi pada penelitian ini dibagi menjadi 3 kelas, yaitu positif, negatif, dan netral. Pengujian yang telah dilakukan mendapatkan nilai akurasi paling baik dengan menggunakan algoritma SVM, yaitu sebesar 80,3%. Hasil klasifikasi menunjukkan bahwa sebesar 54,98% atau sekitar 960,519 data merupakan klasifikasi negatif yang diikuti oleh sentimen netral sebanyak 29,15% dan sentimen positif sebesar 15,86%.

Kata Kunci: demonstrasi; random forest; support vector machine; twitter.

ANALYSIS OF PUBLIC SENTIMENT ON DEMONSTRATIONS IN INDONESIA USING SUPPORT VECTOR MACHINE AND RANDOM FOREST

Abstract

Indonesia is a democratic country that provides freedom for citizens to express their opinions, one of which is through demonstrations. However, sometimes this freedom is misused by certain parties which causes demonstrations followed by riots and anarchist actions, causing various kinds of losses. The purpose of this study is to determine the public's views regarding demonstrations that occurred in Indonesia through tweets on social media Twitter using the Support Vector Machine (SVM) and Random Forest algorithms. In this study, there is an additional process for removing words that are not listed in the Big Indonesian Dictionary (KBBI) because tweets on social media Twitter tend to use more slang words or non-standard languages, so this additional process is expected to provide better clean data. The classification in this study consists of 3 classes, namely positive, negative, and neutral. The tests that have been carried out get the best accuracy using the SVM algorithm, which is equal to 80.3%. The classification results show that 54.98% or around 960,519 data are classified as negative, followed by neutral sentiment around 29.15% and 15.86% classified as positive sentiment.

Keywords: demonstration; random forest; support vector machine; twitter.

PENDAHULUAN

Pemanfaatan teknologi di era sekarang menjadi salah satu kebutuhan penting bagi setiap bidang kehidupan dalam mendukung setiap proses yang ada di dalamnya. Salah satu teknologi yang menjadi perhatian saat ini adalah teknologi *big data* (Krisnasari et al., 2022). *Big data* merupakan suatu ilmu yang berfokus pada pengolahan data dengan jumlah yang banyak dan waktu yang diperlukan cukup singkat. Salah satu penerapan *big data* adalah pada analisis sentimen (Krisnasari et al., 2022). Analisis sentimen merupakan suatu proses yang dapat dimanfaatkan untuk mengenali maupun mengekspresikan suatu emosi, opini, atau subjektivitas yang ada pada suatu teks. Analisis sentimen akan menemukan nilai polaritas dari sebuah teks dan akan diklasifikasikan ke dalam beberapa kelas sentimen. Pengklasifikasian sentimen tersebut dapat membantu dalam mengevaluasi kinerja atau produk serta membantu dalam hal pengambilan keputusan (Oktaviana et al., 2022). Analisis sentimen juga dapat diartikan sebagai sebuah tahapan yang dilakukan dalam pengolahan data berupa teks dengan tujuan mengambil informasi pada sebuah kalimat opini. Analisis sentimen biasanya digunakan untuk menilai sebuah opini terhadap topik tertentu, seperti isu politik, ekonomi, sosial, hukum, dan lain sebagainya (Arsi & Waluyo, 2021). Perasaan seseorang yang bersifat subjektif atau opini ini dapat diungkapkan melalui *tweets* pada media sosial Twitter (Krisnasari et al., 2022).

Twitter merupakan salah satu media sosial yang cukup populer karena menjadi sebuah media yang dapat digunakan untuk saling berbagi informasi, berdiskusi, dan menyampaikan pendapat secara bebas dan terbuka terhadap isu tertentu (Mulya & Sujaini, 2022). Salah satu topik yang menjadi perbincangan di media sosial Twitter adalah aksi demonstrasi yang terjadi di Indonesia. Indonesia adalah sebuah negara demokrasi yang menganut sistem pemerintahan dari rakyat, oleh rakyat, dan untuk rakyat. Setiap warga negara Indonesia memiliki hak yang sama untuk menyampaikan pikiran maupun aspirasinya, baik secara lisan maupun tulisan. Sebagai negara hukum, kaidah hukum tertulis sudah mengatur perihal kebebasan untuk menyampaikan pendapat di hadapan umum yang termuat dalam Undang-Undang No. 9 Tahun 1998 dengan demonstrasi adalah salah satu cara untuk menyampaikan pendapat (Laila, 2019). Namun, adanya ketentuan hukum yang mengatur perihal kebebasan berpendapat tersebut dapat memunculkan terjadinya aksi yang terlalu bebas tanpa adanya pengontrolan diri dan juga sering disalahgunakan oleh oknum tidak bertanggung jawab, sehingga memicu aksi yang bersifat anarkis, seperti kerusuhan, pembakaran, penjarahan, bahkan pembunuhan (Saputra & Rosadi, 2019). Berdasarkan permasalahan tersebut, maka penulis akan melakukan sebuah penelitian untuk mengetahui pandangan masyarakat terhadap aksi demonstrasi yang terjadi di Indonesia menggunakan data Twitter.

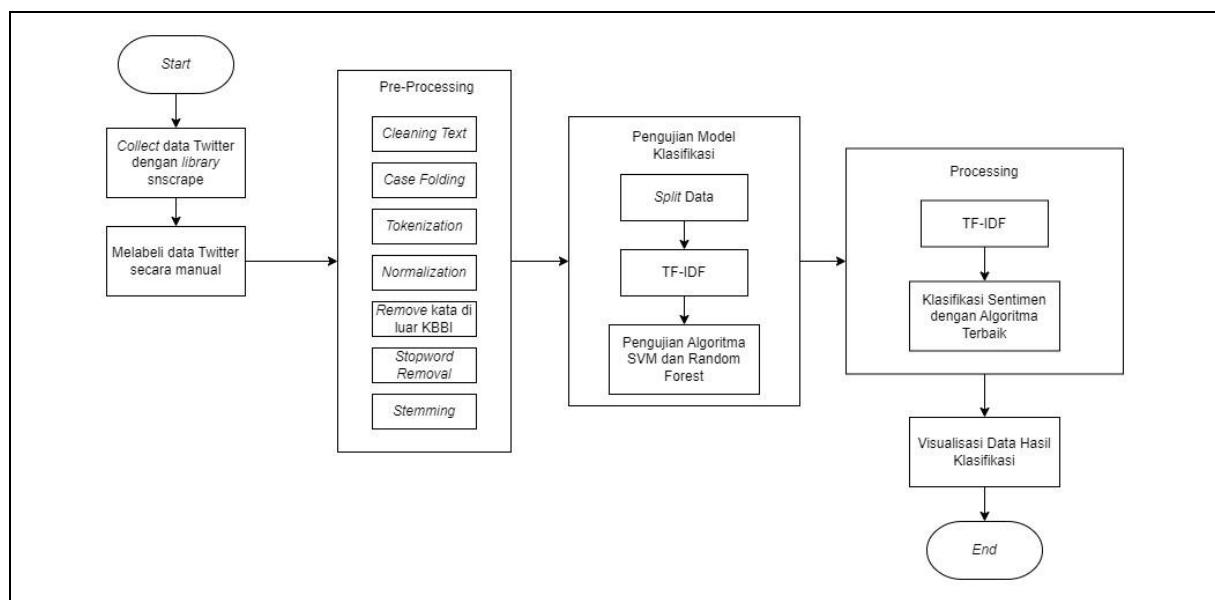
Penelitian terdahulu mengenai analisis sentimen pernah dilakukan oleh (Pane et al., 2021) menggunakan data Twitter dan algoritma *Support Vector Machine* (SVM) terkait topik Undang-Undang Cipta Kerja. Hasil penelitian tersebut menunjukkan bahwa masyarakat Indonesia lebih banyak berpendapat positif terhadap topik tersebut dan akurasi dari algoritma SVM yang digunakan adalah sebesar 83%. Kemudian, penelitian lain dari data Twitter juga dilakukan dengan menggunakan sebanyak 1767 opini untuk melakukan analisis sentimen terhadap usaha waralaba. Hasil yang didapatkan adalah algoritma SVM dan *Neural Network* memiliki akurasi terbaik, yaitu sebesar 83%. Kemudian, algoritma *Decision Tree* sebesar 81%, *K-Nearest Neighbor* sebesar 52%, dan *Naïve Bayes* sebesar 80%. Penelitian tersebut juga mendapatkan hasil bahwa usaha waralaba lebih banyak mendapatkan dukungan dari masyarakat (Mardiana et al., 2019). Penelitian mengenai analisis sentimen dari data Twitter juga pernah diteliti oleh (Herdiyani & Zailani, 2022) terkait dengan pemindahan Ibu Kota Indonesia. Berdasarkan penelitian tersebut, didapatkan hasil bahwa masyarakat Indonesia mendukung pemindahan Ibu Kota Indonesia dan hasil akurasi yang didapatkan adalah sebesar

76% dengan menggunakan algoritma *Random Forest*. Selanjutnya, penelitian mengenai analisis sentimen menggunakan algoritma *Random Forest* juga pernah dilakukan pada tahun 2022 dengan menggunakan sebanyak 1500 data *tweets* dari media sosial Twitter. Penelitian tersebut mendapatkan hasil bahwa akurasi dengan algoritma *Random Forest* yang digunakan adalah sebesar 79% dan masyarakat lebih banyak berpandangan positif terhadap topik vaksin *sinovac* yang diteliti (Aldean et al., 2022).

Adapun perbedaan atau kebaharuan pada penelitian ini adalah adanya penambahan proses pada tahapan *pre-processing*, yaitu menambahkan proses untuk menghilangkan kata-kata yang tidak terdaftar pada Kamus Besar Bahasa Indonesia (KBBI). Hal ini bertujuan karena pengguna media sosial lebih banyak menggunakan bahasa atau kata-kata gaul dan tidak baku yang kadang kala tidak terdaftar pada KBBI dalam membuat sebuah *tweets*, sehingga perlu untuk menambahkan proses ini dan diharapkan bisa mendapatkan data *tweets* hasil *pre-processing* yang lebih bersih dan juga dapat meningkatkan hasil akurasi. Penelitian yang akan dilakukan ini diharapkan dapat memberikan pengetahuan dan informasi kepada pembaca mengenai pandangan masyarakat terhadap aksi demonstrasi yang terjadi di Indonesia. Selain itu, pembaca juga diharapkan dapat mengetahui *trend* sentimen masyarakat terkait topik tersebut setiap tahunnya.

METODE

Penelitian mengenai analisis sentimen terkait aksi demonstrasi di Indonesia ini dilakukan dari Bulan September 2022 yang berlokasi di Gedung Teknologi Informasi, Fakultas Teknik, Universitas Udayana. Penelitian yang akan dilakukan ini menggunakan bahasa pemrograman Python, *database* mongoDB, serta *tools* visualisasi Tableau. Adapun alur penelitian dalam melakukan analisis sentimen masyarakat terhadap aksi demonstrasi di Indonesia dari data Twitter akan dijelaskan pada Gambar 1 di bawah ini.



Gambar 1. Alur Penelitian

Gambar 1 di atas adalah alur atau metode yang akan digunakan pada penelitian ini. Adapun tahapan dari penelitian ini diawali dengan pengumpulan data *tweets* dari media sosial Twitter dengan memanfaatkan *library* snscreape. Data *tweets* terkait dengan topik aksi demonstrasi di Indonesia dikumpulkan dari Januari 2010 sampai dengan Mei 2022.

Selanjutnya, data Twitter tersebut akan dibagi menjadi dua data, yaitu data yang diberikan label dengan cara manual dan data yang tidak berisi label. Pelabelan data secara manual ini dilakukan sendiri oleh penulis. Data dengan label akan digunakan untuk proses pengujian algoritma dengan jumlah data sebanyak 3000, di mana jumlah masing-masing data label untuk setiap kelas adalah 1000 data. Sedangkan, data yang tidak diberi label akan digunakan pada tahapan *processing*. Langkah selanjutnya adalah *pre-processing* untuk membersihkan data *tweets*. Pada tahapan ini terdapat beberapa proses, yaitu *cleaning text* untuk menghilangkan karakter yang tidak dibutuhkan, *case folding* untuk menyamakan kata menjadi bentuk *lower case*, *tokenization* untuk memisahkan kalimat menjadi beberapa kata, *normalization* untuk mengubah bahasa gaul menjadi bentuk baku, menghapus kata yang tidak ada di KBBI (Kamus Besar Bahasa Indonesia), *stopword removal* untuk menghapus kata tidak penting, serta *stemming* untuk mengubah suatu kata menjadi bentuk dasarnya dengan cara menghilangkan imbuhan maupun kata depan (Naufal & Kusuma, 2022). Kemudian, data dengan label yang sudah bersih akan diuji menggunakan algoritma SVM dan *Random Forest* untuk menentukan algoritma yang memiliki nilai akurasi terbaik.

Pemilihan algoritma SVM dan Random Forest pada penelitian ini didasarkan pada penelitian sebelumnya yang juga menggunakan kedua algoritma ini dan memberikan hasil akurasi yang cukup baik. Penelitian yang dilakukan oleh (Pane et al., 2021) tentang sentimen masyarakat pada topik Undang-Undang Cipta Kerja mendapatkan nilai akurasi paling baik dengan menggunakan algoritma SVM, yaitu sebesar 83%. Kemudian, penelitian yang dilakukan pada tahun 2022 menggunakan algoritma Random Forest mendapatkan hasil akurasi paling baik sebesar 76% terkait topik pemindahan Ibu Kota Indonesia (Herdiani & Zailani, 2022), sehingga pada penelitian ini memilih untuk menggunakan kedua algoritma tersebut pada topik yang berbeda, yaitu terkait topik aksi demonstrasi yang terjadi di Indonesia. Setelah tahap pengujian algoritma, langkah selanjutnya adalah tahapan *processing* untuk mengklasifikasikan data yang tidak berisi label menggunakan algoritma yang memiliki performa terbaik, sehingga bisa ditentukan kelas untuk masing-masing data. Kemudian, akan dilanjutkan dengan membuat visualisasi berdasarkan data hasil klasifikasi agar data lebih mudah dipahami.

HASIL DAN PEMBAHASAN

Adapun proses awal dalam penelitian ini adalah mengumpulkan data *tweets* yang diambil dari media sosial Twitter. Data *tweets* ini diambil melalui proses *collect* data dengan memanfaatkan *library snscreape*. Proses *collect* data *tweets* pada penelitian ini menggunakan beberapa *keywords*, seperti “demonstrasi”, “demo buruh”, dan “demo mahasiswa”. Adapun data *tweets* yang berhasil diambil berjumlah 2.036.786 data dan seluruh *field* yang didapatkan pada proses *collect* data Twitter disimpan dalam *database* secara lengkap. Adapun contoh *list* data *tweets* dapat dilihat pada Tabel 1 berikut.

Tabel 1. Contoh Data *Tweets*

| Content |
|---|
| .@ShamsiAli2: Demonstrasi ini merupakan ekspresi kekecewaan yg diungkap secara demokratis. #IndonesiaSatu #BreakingNews #AksiDamai411 https://t.co/o7NafEdpQo |
| Malu sama mahasiswa yg pada demo,bikin rusuh aj ngerusak fasilitas umum,itu buatny pke duit elo jg!tolol |
| @cacinksarapp demo boleh, tp melakukan vandalisme itu kurang ajar, mahasiswa anj*ng :p |
| @yunsitroh2014 bagus nih demonstrasi begini. Salut! :) |
| 7. Saya kesal dgn demo mahasiswa di Jkt akhir2 ini. mrk ganggu lalu lintas, bikin macet tiap hari, bikin rusuh. itu teror bukan orasi. |

@kepingembul kalo korup gak usah d bahas, mahasiswa kan? berpendidikan kan? demo boleh tp harus yg berkelas.

Data *tweets* yang sudah didapatkan akan dibagi menjadi 2 data, yaitu data dengan label dan data yang belum berisikan label. Data label merupakan data *tweets* yang sudah diberikan label secara manual yang akan digunakan untuk melatih model menggunakan algoritma SVM dan *Random Forest*. Adapun pada penelitian ini digunakan tiga *class*, yaitu “*positive*”, “*negative*”, dan “*netral*”. Data yang diberikan label secara manual ini berjumlah 3000 data dengan rincian 1000 data berlabel *positive*, 1000 data berlabel *negative*, dan 1000 data berlabel netral. Adapun contoh data yang sudah diberikan label dapat dilihat pada Tabel 2 berikut.

Tabel 2. Contoh Data Label

| Content | Class |
|--|----------|
| ekspresi kecewa ungkap demokrasi | positive |
| mahasiswa bikin rusuh | negative |
| mahasiswa boleh asal tertib tidak rugi orang | netral |
| bagi hak warga negara damai | positive |
| mahasiswa rugi masyarakat calon suram | negative |
| buat mahasiswa boleh asal jangan anarkis | netral |

Langkah selanjutnya adalah tahapan *pre-processing*, yaitu tahapan untuk membersihkan data yang terdiri dari proses penghilangan karakter (*cleaning text*), *case folding*, *tokenization*, *normalization*, penghapusan kata yang tidak ada pada KBBI, *stopword removal*, serta *stemming*. Adapun hasil dari masing-masing proses *pre-processing* dapat dilihat pada Tabel 3 berikut.

Tabel 3. Contoh Hasil *Pre-Processing*

| Content | Tahapan Pre-Processing | Hasil Pre-Processing |
|--|-------------------------------------|--|
| | <i>Cleaning Text</i> | Mahasiswa apa Mahasewa nie Kalau mahasiswa yg waras gak bakalan mau demo yg merugikan org bnyk |
| | <i>Case Folding</i> | mahasiswa apa mahasewa nie kalau mahasiswa yg waras gak bakalan mau demo yg merugikan org bnyk |
| @alextham878 Mahasiswa apa Mahasewa nie .. | <i>Tokenization</i> | ['mahasiswa', 'apa', 'mahasewa', 'nie', 'kalau', 'mahasiswa', 'yg', 'waras', 'gak', 'bakalan', 'mau', 'demo', 'yg', 'merugikan', 'org', 'bnyk'] |
| Kalau mahasiswa yg waras gak bakalan mau demo yg merugikan org bnyk | <i>Normalization</i> | ['mahasiswa', 'apa', 'mahasewa', 'ini', 'kalau', 'mahasiswa', 'yang', 'waras', 'tidak', 'akan', 'mau', 'demo', 'yang', 'merugikan', 'orang', 'banyak'] |
| | <i>Remove Kata di Luar KBBI</i> | ['mahasiswa', 'apa', 'ini', 'kalau', 'mahasiswa', 'yang', 'waras', 'tidak', 'akan', 'mau', 'demo', 'yang', 'merugikan', 'orang', 'banyak'] |
| | <i>Stopword Removal</i> | ['mahasiswa', 'mahasiswa', 'waras', 'tidak', 'merugikan'] |
| | <i>Stemming</i> | mahasiswa mahasiswa waras tidak rugi |

Setelah data melalui tahapan *pre-processing*, langkah selanjutnya adalah proses pengujian model menggunakan dua algoritma, yaitu SVM dan *Random Forest* serta

digunakan juga empat skenario untuk menguji kedua algoritma tersebut. Adapun perbandingan hasil pengujian dari algoritma SVM dan *Random Forest* dapat dilihat pada Tabel 4 berikut.

Tabel 4. Hasil Pengujian Algoritma SVM dan Random Forest

| Algoritma | Skenario | Accuracy | Precision | Recall | F1-Score |
|-------------------------------|----------|----------|-----------|--------|----------|
| Support Vector Machine | 90:10 | 80.3% | 82.1% | 80.3% | 80.3% |
| | 80:20 | 79.8% | 80.0% | 79.8% | 79.8% |
| | 70:30 | 75.7% | 76.7% | 75.7% | 75.6% |
| | 60:40 | 74.6% | 74.9% | 74.6% | 74.5% |
| Random Forest | 90:10 | 77.7% | 79.2% | 77.7% | 77.7% |
| | 80:20 | 77.5% | 78.5% | 77.5% | 77.5% |
| | 70:30 | 72.8% | 73.8% | 72.8% | 72.8% |
| | 60:40 | 74.4% | 74.8% | 74.4% | 74.4% |

Tabel 4 di atas merupakan perbandingan hasil pengujian dari masing-masing algoritma yang digunakan. Evaluasi dari kinerja model algoritma SVM dan Random Forest di atas menggunakan *confusion matrix*. Dalam *confusion matrix*, terdapat beberapa metriks yang dapat digunakan sebagai indikator evaluasi, yaitu nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Penelitian ini akan berfokus pada metriks *accuracy*, yaitu perbandingan dari banyaknya data yang diprediksi secara benar dengan jumlah keseluruhan data yang ada (Fitriyyah et al., 2019). Pemilihan metriks *accuracy* dikarenakan pada penelitian ini akan memfokuskan agar algoritma yang digunakan mampu memberikan hasil prediksi yang sesuai dengan kelas sentimen yang seharusnya. Nilai *accuracy* juga dapat mempengaruhi algoritma dalam melakukan proses klasifikasi dengan tepat, sehingga metriks *accuracy* menjadi fokus utama pada evaluasi kinerja algoritma (Ainurrohmah, 2021). Selain itu, pada pengujian algoritma ini juga menggunakan beberapa skenario pembagian *dataset* untuk *training* dan *dataset* yang berperan sebagai *data testing*. Hal ini bertujuan agar dapat menentukan pembagian *data training* dan *data testing* yang sesuai, sehingga dapat menghasilkan nilai akurasi paling baik.

Berdasarkan Tabel 4 di atas dapat dilihat bahwa algoritma SVM memiliki nilai akurasi terbaik, yaitu sebesar 80.3%. Hasil tersebut didapatkan dari skenario 90:10, yaitu 90% sebagai *data training* dan 10% sebagai *data testing*. Penggunaan skenario 90:10 untuk proses *split data* juga menjadi penyebab nilai akurasi yang meningkat. Hal ini disebabkan karena penggunaan *data training* yang lebih banyak, sehingga algoritma memiliki banyak variasi data dari *dataset*. Lalu, lebih banyaknya jumlah *data training* juga dapat membantu model untuk mempelajari data dengan lebih baik, sehingga dapat meningkatkan kemampuan model dalam proses klasifikasi (A.M & Yaqin, 2022). Selain itu, algoritma SVM juga memiliki kinerja yang baik dalam melakukan analisis pada data dan pengenalan pola, sehingga algoritma ini akan bekerja lebih baik pada proses *text mining* atau pengklasifikasian berbasis teks (Natasuwarna, 2020). Berdasarkan hasil pengujian model tersebut, pada penelitian ini akan menggunakan algoritma SVM dengan skenario 90:10 untuk tahapan *processing* sesuai dengan hasil pengujian.

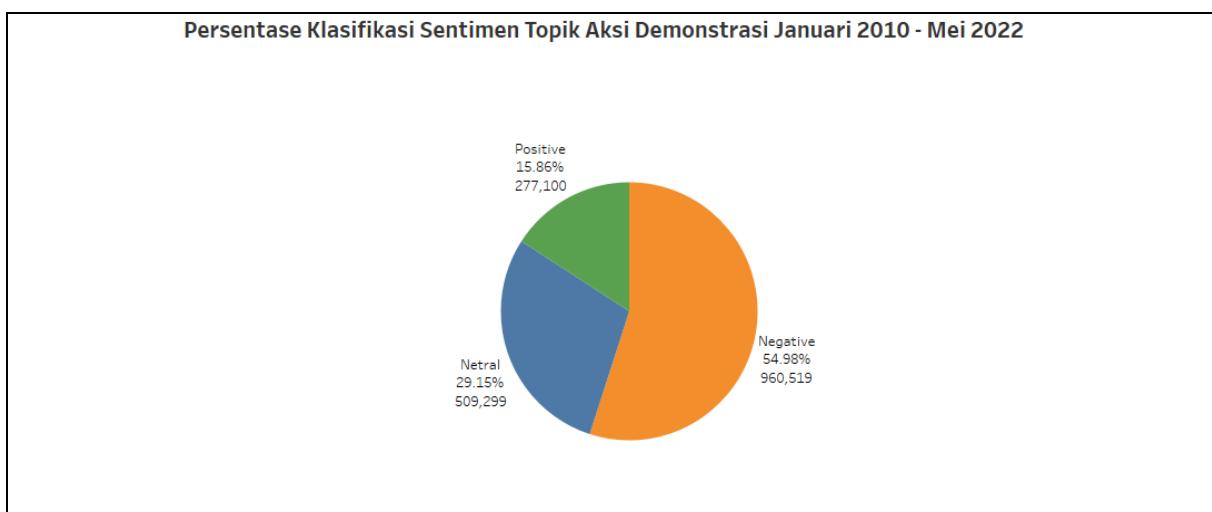
Adapun langkah selanjutnya adalah tahapan *processing*, yaitu tahapan yang bertujuan untuk memberikan label pada data yang belum diberikan label. Proses pelabelan ini menggunakan algoritma yang telah dilatih pada proses pengujian model klasifikasi dan mampu memberikan performa yang baik, yaitu algoritma SVM dengan skenario 90:10. Adapun contoh hasil dari tahapan *processing* dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil Tahapan *Processing*

| Content Clean | Class |
|---------------|-------|
|---------------|-------|

| | |
|--|----------|
| macet gara gara buruh | Negative |
| hak warga negara jamin kecuali negara otoriter | Positive |
| asal tertib hak warga negara | Netral |
| buruh jangan rugi orang | Negative |
| hak warga negara tema cara bisa tahu kualitas karakter | Positive |
| mahasiswa pikir jernih jangan anarkis | Netral |

Tahapan selanjutnya adalah visualisasi data, yaitu proses untuk mengubah data hasil klasifikasi ke dalam bentuk visualisasi data, seperti *line chart*, *geo chart*, maupun dalam bentuk persentase. Tujuan dari visualisasi data ini adalah agar data hasil klasifikasi dapat lebih mudah untuk dipahami. Adapun contoh hasil dari visualisasi data dapat dilihat pada Gambar 2 berikut.



Gambar 2. Hasil Visualisasi Data

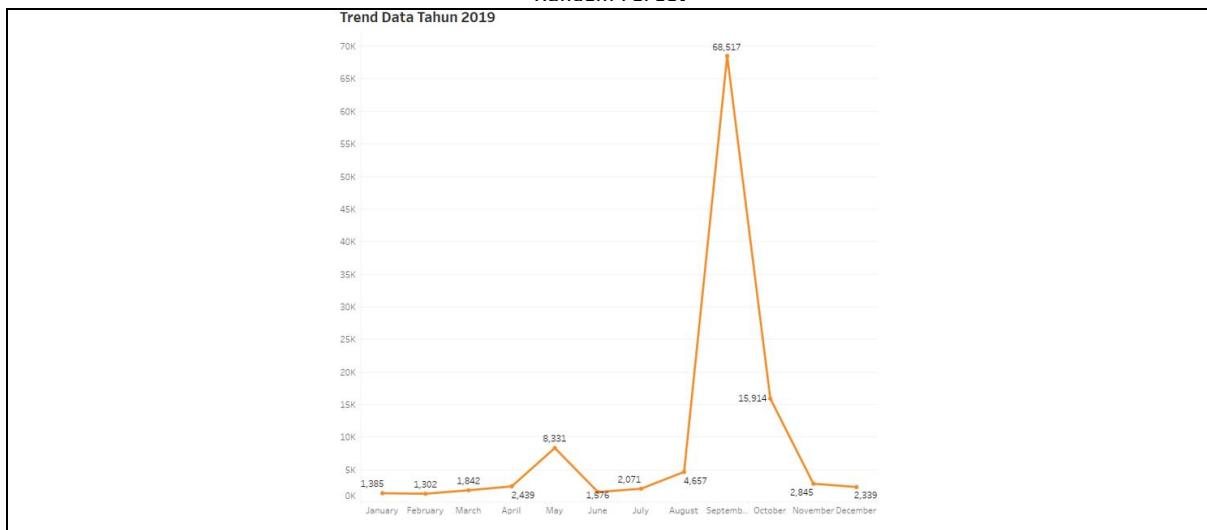
Gambar 2 di atas merupakan visualisasi hasil dari proses pelabelan data *tweets* dalam bentuk persentase. Data di atas menggambarkan bahwa masyarakat Indonesia lebih banyak berpandangan negatif terkait dengan aksi demonstrasi yang terjadi dengan total 54,98% atau sekitar 960.519 data. Kemudian, jumlah masyarakat yang mendukung terjadinya aksi demonstrasi ini adalah 277.100 orang atau sekitar 15,86%. Sedangkan, sebanyak 509.299 atau 29,15% masyarakat Indonesia memilih untuk bersikap netral terkait aksi demonstrasi yang pernah terjadi. Adapun visualisasi *trend* sentimen setiap tahunnya dapat dilihat pada Gambar 3 berikut.

DECODE: Jurnal Pendidikan Teknologi Informasi, 3 (2) (2023): 257-267

Ni Made Tara Okta Adriana, I Made Agus Dwi Suarjaya, Dwi Putra Githa



DECODE: Jurnal Pendidikan Teknologi Informasi, 3 (2) (2023): 257-267
Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan Support Vector Machine dan Random Forest



Gambar 3. Visualisasi *Trend* Sentimen per Tahun

Gambar 3 di atas merupakan visualisasi terkait *trend* sentimen masyarakat terhadap topik aksi demonstrasi setiap tahunnya, yaitu dari Januari 2010 sampai dengan Mei 2022. Gambar di atas menunjukkan bahwa *trend* sentimen masyarakat berubah-ubah setiap tahunnya dan cenderung menunjukkan grafik naik turun. Hal ini bisa disebabkan karena adanya aksi demonstrasi dalam satu waktu tertentu, sehingga menyebabkan *trend* akan mengalami peningkatan. Sebagai contoh, pada tahun 2019 terlihat bahwa di awal bulan grafik menunjukkan data yang cukup stagnan dan terjadi peningkatan di bulan Mei. Hal ini terjadi karena adanya aksi demonstrasi oleh massa pendukung salah satu pasangan calon presiden yang merasa kecewa dengan hasil perhitungan suara pada pemilihan umum pada saat itu. Aksi yang semulanya damai harus berakhir rusuh karena adanya provokator yang menyusup dan melempari aparat dengan kayu atau batu, sehingga pihak kepolisian harus mengambil tindakan dengan menembakkan gas air mata ke massa aksi (Indonesia, 2020).

Kemudian, masih di tahun yang sama terjadi lagi peningkatan sentimen yang cukup signifikan dibandingkan bulan sebelumnya, yaitu pada bulan September 2019. Pada bulan tersebut, ribuan mahasiswa dari berbagai perguruan tinggi di Indonesia melakukan aksi unjuk rasa secara besar-besaran. Bahkan, aksi ini tercatat sebagai aksi demonstrasi besar yang pernah terjadi di Indonesia. Adapun yang menjadi permasalahan dalam aksi ini adalah penolakan mahasiswa terkait dengan adanya revisi Undang-Undang Komisi Pemberantasan Korupsi dan Rancangan Undang-Undang Kitab Undang-Undang Hukum Pidana (RUU KUHP) yang dianggap dapat mengancam hak asasi manusia sekaligus pelemahan terhadap lembaga KPK (Nabilla, 2022), sehingga pada bulan ini sentimen masyarakat terkait dengan aksi demonstrasi yang terjadi mengalami peningkatan yang cukup tinggi.

KESIMPULAN DAN SARAN

Penelitian mengenai analisis sentimen masyarakat terkait dengan aksi demonstrasi yang terjadi di Indonesia ini menggunakan sebanyak 2.036.786 data. Data didapatkan dari opini masyarakat pada media sosial Twitter yang diambil dari Januari 2010 sampai dengan Mei 2022. Penelitian yang telah dilakukan menggambarkan bahwa algoritma SVM memiliki nilai akurasi yang lebih baik daripada algoritma *Random Forest*, yaitu sebesar 80,3% dengan menggunakan skenario 90:10. Proses klasifikasi menunjukkan bahwa sentimen *negative* lebih mendominasi dibandingkan sentimen lainnya, yaitu sebesar 54,98% atau sekitar 960.519 data. Kemudian, jumlah masyarakat yang mendukung terjadinya aksi demonstrasi ini adalah sekitar 15,86%. Sedangkan, sebanyak 29,15% masyarakat Indonesia memilih untuk bersikap netral

terkait aksi demonstrasi yang pernah terjadi. Kemudian, *trend* sentimen masyarakat terkait dengan aksi demonstrasi di Indonesia ini menunjukkan *trend* yang berubah-ubah setiap tahunnya dan cenderung menunjukkan grafik naik turun. *Trend* akan mengalami peningkatan apabila terjadi aksi demonstrasi terkait isu tertentu dalam kurun waktu tertentu. Apabila aksi yang terjadi merupakan aksi demonstrasi besar yang melibatkan banyak massa dan menjadi perbincangan yang ramai di kalangan pengguna media sosial, khususnya Twitter, maka sentimen masyarakat juga akan mengalami peningkatan yang cukup signifikan.

Adapun saran untuk penelitian mengenai analisis sentimen yang selanjutnya adalah dapat menggunakan kombinasi data dari *platform* media sosial yang lain, seperti Facebook agar data yang didapatkan lebih banyak dan beragam. Kemudian, pada proses *cleaning text* juga dapat lebih dimaksimalkan lagi agar data yang didapatkan dari tahapan ini menjadi lebih bersih. Selain itu, juga dapat menggunakan jenis algoritma *machine learning* maupun *deep learning* yang lain untuk mendapatkan algoritma yang memiliki performa dan akurasi yang lebih baik lagi, sehingga bisa mendapatkan hasil klasifikasi yang lebih akurat.

DAFTAR PUSTAKA

- Ainurrohmah, A. (2021, February). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. *PRISMA, Prosiding Seminar Nasional Matematika*, 493-499.
- Aldean, M. Y., Paradise, P., & Nugraha, N. A. S. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac). *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 4(2), 64-72.
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147-156. <https://doi.org/10.25126/jtiik.0813944>
- Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(3), 279-285. <https://doi.org/10.26418/jp.v5i3.34368>
- Herdiyani, T. C., & Zailani, A. U. (2022). *Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia*. 3(2), 154-165.
- Indonesia, C. (2020). *22 Mei Setahun yang Lalu, Jakarta Membara di Masa Pemilu*. <https://www.cnnindonesia.com/nasional/20200522051704-20-505747/22-meい-setahun-yang-lalu-jakarta-membara-di-masa-pemilu>
- Krisnasari, N. K. A., Suarjaya, I. M. A. D., & Raharja, I. M. S. (2022). Classification of Public Figures Sentiment on Twitter using Big Data Technology. *Journal of Informatics and Telecommunication Engineering*, 6(1), 157-169. <https://doi.org/10.31289/jite.v6i1.7329>
- Laila, K. (2019). Hukum Progresif sebagai Solusi Kebebasan Berpendapat dengan Asas Demokrasi Pancasila. *Jurnal Cakrawala Hukum*, 10(2), 177-186. <https://doi.org/10.26905/idjch.v10i2.3546>
- Mardiana, T., Syahreva, H., & Tuslaela, T. (2019). Komparasi Metode Klasifikasi Pada Analisis Sentimen Usaha Waralaba Berdasarkan Data Twitter. *Jurnal Pilar Nusa Mandiri*, 15(2), 267-274. <https://doi.org/10.33480/pilar.v15i2.752>

DECODE: Jurnal Pendidikan Teknologi Informasi, 3 (2) (2023): 257-267

Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan Support Vector Machine dan Random Forest

- Mulya, S., & Sujaini, H. (2022). *Analisis Sentimen Tren Olahraga di Masa Pandemi COVID-19 pada Twitter dengan Metode Naïve*. 8(2), 284-291.
- Nabilla, F. (2022). *4 Aksi Demo Mahasiswa Terbesar di Indonesia Sepanjang Sejarah, Terbaru Tahun 2019*. <https://www.suara.com/news/2022/04/10/172437/4-aksi-demo-mahasiswa-terbesar-di-indonesia-sepanjang-sejarah-terbaru-tahun-2019?page=2>
- Natasuwarna, A. P. (2020). Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring. *Techno.Com*, 19(4), 437-448. <https://doi.org/10.33633/tc.v19i4.4044>
- Naufal, M. F., & Kusuma, S. F. (2022). Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis Deep Learning. *Jurnal Edukasi dan Penelitian Informatika*, 8(1), 44-49.
- Oktaviana, N. E., Sari, Y. A., & Indriati. (2022). Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features Dan Support Vector Machine. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, 9(2), 357-362. <https://doi.org/10.25126/jtiik.202295625>
- Pane, S. F., Owen, A., & Prianto, C. (2021). Analisis Sentimen UU Omnibus Law pada Twitter Menggunakan Metode Support Vector Machine. *Jurnal Telekomunikasi Dan Komputer*, 11(2), 130-142. <https://doi.org/10.22441/incomtech.v11i2.10874>
- Saputra, S., & Rosadi, O. (2019). Efektivitas Penggunaan Kekuatan Oleh Kepolisian Dalam Menangani Aksi Unjuk Rasa Anarkis Di Kota Padang (Studi Pada Satsabhara Polresta Padang). *UNES Law Review*, 1(3), 309-317. <https://doi.org/10.31933/law.v1i3.42>
- Tamrizal, A. M., & Yaqin, A. Perbandingan Algoritma Naïve Bayes, K-Nearest Neighbors dan Random Forest untuk Klasifikasi Sentimen Terhadap BPJS Kesehatan pada Media Twitter. *InComTech : Jurnal Telekomunikasi Dan Komputer*, 12(1), 01. <https://doi.org/10.22441/incomtech.v12i1.13642>

How to cite:

- Adriana, N. M. T. D., Suarjaya, I. M. A. D., & Githa, D. P. (2023). Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan Support Vector Machine dan Random Forest. *DECODE: Jurnal Pendidikan Teknologi Informasi*, 3(2), 257-267. <http://dx.doi.org/10.51454/decode.v3i2.187>