

Penerapan Metode *Support Vector Machine* Analisis Sentimen Tweet Pergantian Logo Halal di Indonesia

Widi Afandi¹, Tri Ginanjar Laksana², Nia Annisa Ferani Tanjung³

^{1,2,3}Institut Teknologi Telkom Purwokerto

Jl. DI Panjaitan No.128, Purwokerto Selatan, Kabupaten Banyumas 53147, Jawa Tengah, Indonesia

19102127@ittelkom-pwt.ac.id, anjarlaksana@ittelkom-pwt.ac.id, nia@ittelkom-pwt.ac.id

ARTICLE INFO

Article history:

Received Oktober 28 2022

Received in revised form November 12 2022

Accepted Desember 29 2022

Available online Juli 2022

ABSTRACT

The Halal Product Assurance Agency (BPJPH) is an agency under the auspices of the Ministry of Religion with the task of ensuring the halalness of products in Indonesia. BPJPH has become a public concern after establishing the new halal logo. On February 10, 2022 the new halal logo was ratified by the Head of BPJPH, Muhammad Aqil Irham. This has become a topic of public discussion either directly or through social media, one of which is social media twitter. The number of opinion tweets about the change of the halal logo can be used as a data source to obtain information about public opinion on the change of the halal logo through sentiment analysis. Sentiment analysis can be done by machine learning approach, one of these is the SVM algorithm . In this research, oversampling and undersampling are applied to handle data that has an unbalanced sentiment class. The results showed that the Support Vector Machine (SVM) model using oversampling training data got the highest accuracy, recall, precision, and f1-score, namely 71% accuracy, 67% precision, 61% recall, and 61% f1-score while training using undersampling training data has the lowest performance, namely getting 56% accuracy, 51% precision, 57% recall, and 52% f1-score.

Keywords: *Sentiment Analysis, Support Vector Machine, Text Classification, Twitter*

1. Pendahuluan

Pangan adalah salah satu kebutuhan primer bagi manusia. Pangan yang baik haruslah tersedia, aman, berkualitas, memiliki nutrisi serta gizi, dan tidak bertentangan dengan agama, keyakinan, serta budaya masyarakat. Dalam agama islam, umat muslim diwajibkan mengonsumsi makanan dan minuman yang halal. Pangan yang sudah terverifikasi kehalalannya memiliki sebuah logo yang terdapat tulisan "halal" dalam bahasa arab pada kemasannya sebagai penanda bahwa pangan tersebut sudah melalui sertifikasi halal dan layak dikonsumsi oleh umat muslim[1]. Di Indonesia terdapat lembaga khusus untuk mengurus aturan dalam memberikan label halal pada pangan yaitu Badan Penyelenggara Jaminan Produk Halal (BPJPH)[2].

Received Oktober, 28 2022; Revised November, 12 2022; Accepted Desember, 29 2022

Badan Penyelenggara Jaminan Produk Halal (BPJPH) adalah sebuah badan yang berada dalam naungan Kementerian Agama dengan tugas dan fungsinya menjamin kehalalan produk yang masuk, diedarkan, dan diperjualbelikan di Indonesia sesuai Undang-Undang Nomor 13 Tahun 2014. Badan Penyelenggara Jaminan Produk Halal juga didukung oleh tugas dan fungsi yang ada pada Undang-Undang Nomor 33 Tahun 2014 mengenai registrasi halal, sertifikasi halal, verifikasi halal, pembinaan dan pengawasan kehalalan produk, bekerjasama dengan stakeholder terkait, dan menetapkan standar kehalalan suatu produk di Indonesia. BPJPH menjadi perhatian publik setelah menetapkan logo halal baru pada tanggal 10 Februari 2022 yang disahkan oleh Kepala BPJPH Muhammad Aqil Irham. Logo halal mulai diterapkan mulai tanggal 1 Maret 2022 hingga tahun 2026[2], [3].

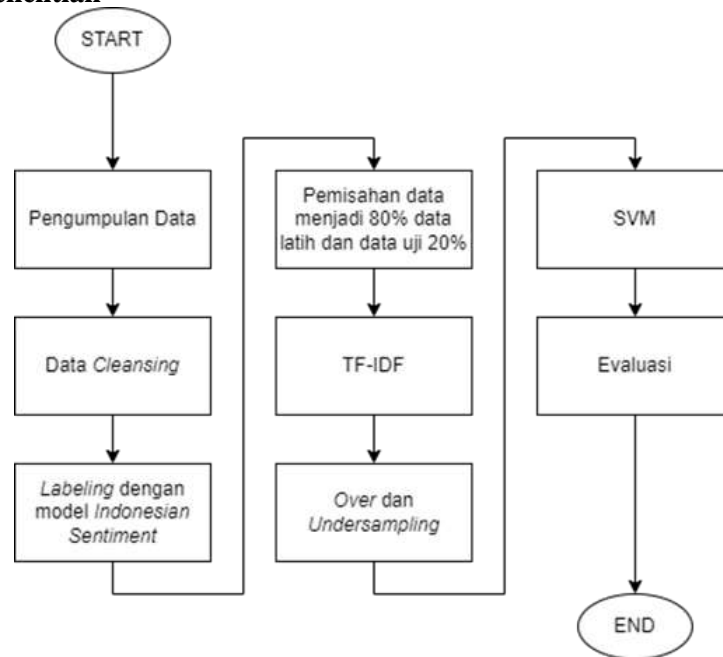
Pergantian logo halal disebabkan adanya pelimpahan otoritas pengeluaran sertifikasi halal yang sebelumnya dikelola oleh Lembaga Pengkajian Pangan, Obat-obatan, dan Kosmetika Majelis Ulama Indonesia (LPPOM MUI) kemudian dilimpahkan ke Badan Penyelenggara Jaminan Produk Halal (BPJPH) Kemenag. Menurut Muhammad Aqil Irham, desain dari logo halal baru memiliki nilai keindonesiaan yang unik dan menggambarkan halal Indonesia[4], [5]. Namun pergantian logo halal yang dilakukan oleh BPJPH mendapatkan pro dan kontra di masyarakat. Masyarakat mengeluhkan desain logo halal baru yang dianggap terlalu Jawa sentris serta tulisan Arab halal yang dianggap sulit untuk dibaca[6]. Hal ini menjadi topik pembahasan masyarakat baik secara langsung atau melalui sosial media. Salah satu sosial media untuk menyampaikan dan menampung opini masyarakat tentang pergantian logo halal adalah Twitter. Tweet opini tentang pergantian logo halal banyak ditemukan di Twitter hingga saat ini yang bisa dicari dengan menggunakan kata kunci “logo halal”. Dengan banyaknya tweet opini tentang pergantian logo halal dapat dimanfaatkan sebagai sumber data untuk mendapatkan informasi tentang opini masyarakat terhadap pergantian logo halal melalui analisis sentimen.

Analisis sentimen saat ini menjadi topik yang ramai dibahas dan dikembangkan. Analisis sentimen adalah proses mengekstraksi, memahami dan memproses kalimat opini masyarakat di sosial media untuk mendapatkan informasi pada kalimat opini. Analisis sentimen dimanfaatkan untuk melakukan evaluasi atau menunjukkan tingkat kepuasan masyarakat terhadap suatu kasus menggunakan data dari sosial media yang diperoleh melalui proses *crawling*. Analisis sentimen dapat menggunakan pendekatan *machine learning*[7], [8]. Ada beberapa model *machine learning* yang digunakan untuk analisis sentimen seperti *K-Nearest Neighbor*, *Support Vector Machine*, *Random Forest*, *Naïve Bayes*, dan algoritma lainnya. Berdasarkan beberapa penelitian yang dilakukan sebelumnya, algoritma *Support Vector Machine* atau sering disebut juga SVM memiliki akurasi yang baik dalam melakukan tugas klasifikasi teks dengan mendapatkan akurasi 75% pada data uji[7]–[10].

Ragil Dimas Himawan dan Eliyani melakukan penelitian dengan tujuan untuk membandingkan akurasi pada data tweet tentang pemerintah Provinsi DKI Jakarta di Masa Pandemi dengan algoritma *Random forest*, *naive bayes*, dan SVM. Hasil dari penelitian tersebut didapatkan hasil bahwa algoritma SVM memiliki akurasi lebih tinggi 2%-3% dari algoritma lain[8]. Kwee Felicia Iona dan Indra Budi melakukan penelitian dengan membandingkan beberapa algoritma *machine learning* untuk melakukan klasifikasi sentimen pada data tentang level genangan banjir di Jakarta. Dari algoritma yang digunakan, algoritma SVM mendapatkan akurasi 90%[11]. Penelitian yang dilakukan oleh Abdelrahman I. Saad tentang uji perbandingan akurasi pada data ulasan maskapai penerbangan di US menggunakan beberapa algoritma *machine learning*, SVM mendapatkan akurasi 2-5% lebih tinggi dari algoritma lain. Dari penelitian yang disebutkan di atas, algoritma SVM memiliki akurasi yang sangat baik dalam melakukan klasifikasi sentimen.

Dengan demikian, pada penelitian ini akan dilakukan analisis sentimen pada tweet pergantian logo halal yang disahkan oleh BPJPH dengan data hasil dari Twitter melalui proses *crawling* menggunakan kata kunci “logo halal” dengan algoritma *Support Vector Machine* (SVM).

2. Metode Penelitian



Gambar 1 Diagram Alir Penelitian

2.1 Pengumpulan Data

Pengumpulan data menggunakan *library* twint dari bahasa python untuk melakukan proses *crawling*. Proses *crawling* dilakukan dengan kata kunci “logo” dan “halal” dalam rentang tanggal 10 Februari 2022 sampai dengan disahkannya logo halal yang baru pada tanggal 10 Maret 2022. Proses *crawling* berhasil mendapatkan total 1015 tweet.

Tabel 1 Tweet Hasil *Crawling*

No	Tweet	Bahasa	Username
1	uh... i`m not a Muslim but i just want to state my opinion that i dont like the new halal logo... it looks... not right...?	en	claryoouu
2	@firman_nurhuda Intinya bkn labelnya yg budaya bang, tp bentuk logo sih. Soal selera saja. Selama udh jelas ada tulisan halal tentu bs dipake. Itu yg penting 🙏	in	TimHore1112

2.2 Data Cleansing

Tweet yang menggunakan bahasa asing di *translate* ke bahasa Indonesia dengan menggunakan *library* googletrans. Proses selanjutnya akan dilakukan *case folding* atau mengubah abjad menjadi huruf kecil kecil (*lower case*). Tahap selanjutnya adalah menghilangkan karakter yang tidak diperlukan seperti \n, html code (&), link/url, username twitter, dan spasi double. Data tweet kemudian dibersihkan dari karakter *non alphanumeric* seperti tanda baca dan angka. Tahapan selanjutnya data akan dinormalisasi kata per kata dalam tweet dengan tujuan mentransformasikan tweet menjadi kalimat yang lebih baku. Proses selanjutnya dilakukan penghilangan stopwords untuk menghilangkan kata yang terlalu umum dan dianggap tidak terlalu penting bagi sistem. Terakhir yaitu *stemming* untuk mentransformasikan kata yang memiliki imbuhan menjadi kata dasarnya dengan menghapus imbuhan pada kata tersebut [12], [13].

Dilakukan pengecekan data untuk mencegah data duplikat. Hasil dari proses *cleansing* tersisa total data 1005 tweet dari sebelumnya 1015 tweet

2.3 Labeling

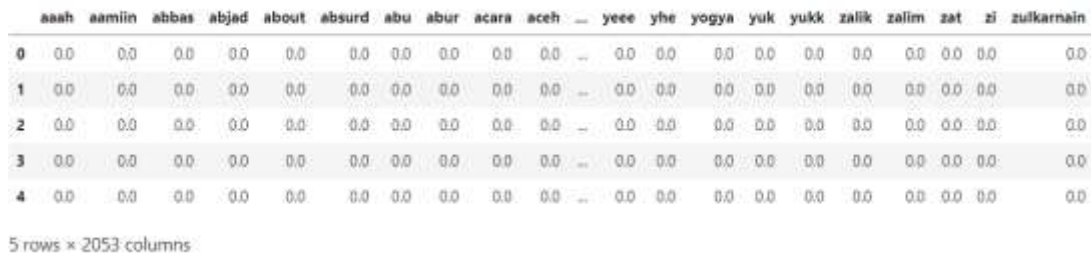
Pelabelan dilakukan dengan menggunakan model *Indonesian Sentiment* yang dibangun oleh Sahri Ramadhan. Pelabelan dilakukan dengan *upload* dataset yang sudah dilakukan normalisasi kata kemudian dataset akan dilabeli dengan tiga label yaitu positif, negatif, dan netral. Dari proses pelabelan menggunakan *Indonesian Sentiment* didapatkan data dengan label negatif sebanyak 590, netral dengan data sebanyak 258, dan positif dengan data sebanyak 157.

2.4 Splitting Data

Data dipisahkan menjadi dua bagian yang proporsinya berbeda, data latih dengan proporsi 80% dari keseluruhan data dan data uji dengan proporsi 20% dari keseluruhan data atau sisa data setelah dikurangi untuk data latih. Data latih dengan jumlah 80% dari total keseluruhan data digunakan untuk proses pelatihan dan sisanya menjadi data uji dengan jumlah 20% dari total keseluruhan data digunakan untuk pengujian performa algoritma yang digunakan.

2.5 TF-IDF

Metode TF-IDF selanjutnya digunakan pada tahap ini untuk mentransformasikan kalimat dalam bentuk teks menjadi bentuk vektor agar dapat diolah algoritma SVM . Vektorisasi dengan TF-IDF akan menghasilkan nilai dalam rentang 0 sampai 1 [14], [15].



Gambar 2 Hasil Vektorisasi dengan TF-IDF

2.6 Oversampling dan Undersampling

Class data memiliki jumlah data yang tidak seimbang antara *class* positif, netral, dan negatif. Digunakan teknik *oversampling* (*SMOTE*) dan *undersampling* untuk menyeimbangkan *class* data dengan memperbanyak data pada *minority class* ataupun mengurangi data pada *majority class*.

2.7 Implementasi SVM

Implementasi algoritma *Support Vector Machine* yang fokus pada penggunaan kernel RBF dengan melakukan *hyperparameter tuning* yang terdapat pada kernel RBF yaitu parameter C dan gamma untuk mencari nilai *accuracy*, *precision*, *recall*, dan *f1-score* tertinggi pada tiga skenario data yaitu data latih asli, data latih *oversampling*, dan data latih *undersampling*. *Hyperparameter tuning* di kernel RBF pada parameter C dan gamma dilakukan dengan menggunakan metode *Grid Search*.

Tabel 2 Parameter untuk Uji Coba

Parameter	Nilai
C	[0.001, 0.01, 0.1, 0.5, 1, 10, 100]
Gamma	['scale', 0.001, 0.01, 0.1, 1, 10, 100]

Skenario pelatihan data latih *oversample* dan *undersample*, pelatihan hanya menggunakan parameter terbaik yang dihasilkan dari *gridsearch* pada data latih asli. Proses mencari parameter terbaik mendapatkan hasil seperti berikut

```

Akurasi yang didapat: 0.6654192546583851
Kombinasi parameter yang didapat:
|  | params | values |
|  | :-----: | :-----: |
| 0 | C | 10 |
| 1 | break_ties | False |
| 2 | cache_size | 200 |
| 3 | class_weight | |
| 4 | coef0 | 0.0 |
| 5 | decision_function_shape | ovr |
| 6 | degree | 3 |
| 7 | gamma | 0.1 |
| 8 | kernel | rbf |
| 9 | max_iter | -1 |
|10 | probability | False |
|11 | random_state | 0 |
|12 | shrinking | True |
|13 | tol | 0.001 |
|14 | verbose | False |

```

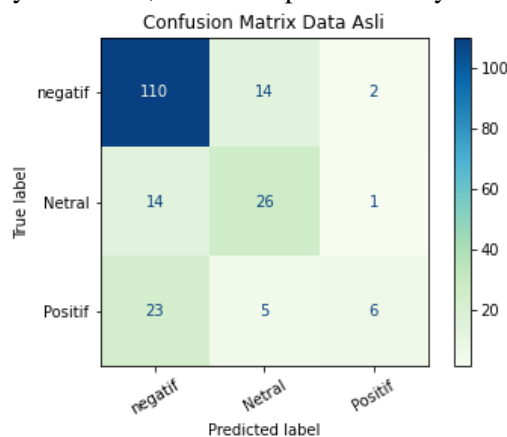
Gambar 3 Hasil Parameter Terbaik dengan *GridSearch*

2.8 Pengujian dan Evaluasi

Hasil pengujian pada masing-masing skenario data akan berbeda antara data latih asli, *oversampling*, dan *undersampling*. Hasil pengujian akan menampilkan *confusion matrix* serta *metrics accuracy*, *precision*, *recall*, dan *f1-score* pada masing-masing skenario data [16], [17]. Proses pengujian menggunakan data uji untuk diprediksi menggunakan model SVM yang sudah melalui proses pelatihan dengan menggunakan data latih asli, *oversample*, dan *undersample*.

3. Results and Analysis

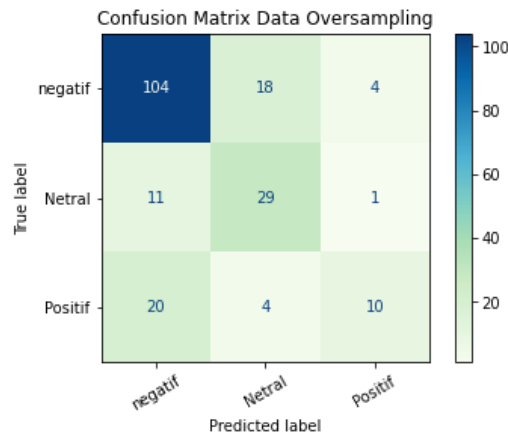
Algoritma SVM pada skenario pelatihan dengan data latih asli memberikan hasil bahwa algoritma SVM mampu melakukan prediksi pada data uji meskipun performanya kurang dalam memprediksi data pada *class* sentimen positif. Dari 126 data dengan *class* negatif, 41 data *class* netral, dan 34 data *class* positif pada data uji, algoritma SVM memprediksi benar negatif sebanyak 110 data, benar netral sebanyak 26 data, dan benar positif sebanyak 6 data.

Gambar 4 *Confusion Matrix* Data Asli

Akurasi dan parameter terbaik yang didapatkan pada skenario pelatihan menggunakan data latih asli adalah 71% dengan parameter C adalah 10 dan gamma 0.1. Parameter tersebut nantinya digunakan untuk pelatihan data latih *oversampling* dan *undersampling* menggunakan algoritma SVM. Hasil *confusion matrix* tersebut juga digunakan untuk menghitung *precision*, *recall*, dan *f1-score* dengan hasil sebesar 66% untuk *precision*, 56% *recall*, dan 56% *f1-score*

Selanjutnya hasil algoritma SVM pada skenario data latih *oversampling* dengan menggunakan parameter terbaik yang dihasilkan pada skenario data latih asli menunjukkan bahwa hasil *matrix* akurasi yang dihasilkan sama besarnya dengan model SVM yang dilatih menggunakan data latih asli. Hal ini ditunjukkan dari 126 data dengan *class* negatif, 41 data *class* netral, dan 34 data *class* negatif pada data uji, algoritma SVM hasil dari pelatihan data *oversampling* dapat

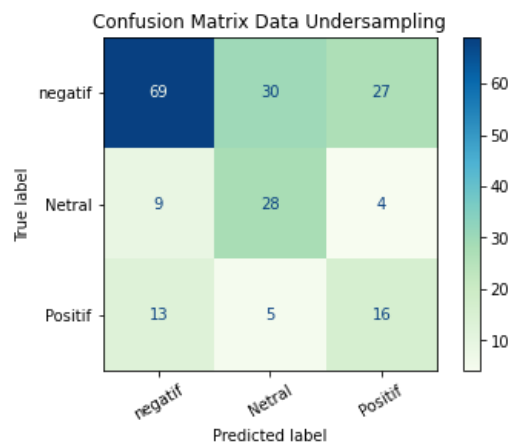
memprediksi benar negative sebanyak 104 data, benar netral sebanyak 29 data, serta benar positif sebanyak 10 data.



Gambar 5 *Confusion Matrix Data Oversampling*

Prediksi benar pada *class* minoritas (*class* netral dan positif) dengan pelatihan menggunakan data *oversampling* mengalami kenaikan 3 sampai 7 data dibandingkan dengan data latih asli. Meskipun begitu hasil akurasi yang didapatkan dari model SVM yang dilatih dengan data latih *oversampling* besarnya sama dengan data latih asli yaitu 71%, 67% untuk *precision*, 61% untuk *recall*, dan 61% *f1-score*.

Terakhir, hasil algoritma SVM pada scenario data latih *undersampling* memberikan hasil akurasi yang kurang baik dibandingkan 2 skenario sebelumnya. Dengan 126 data dengan *class* negatif, 41 data *class* netral, dan 34 data *class* positif pada data uji, algoritma SVM dapat memprediksi benar negatif sebanyak 69 data, benar netral sebanyak 28 data, dan benar positif sebanyak 16 data.



Gambar 6 *Confusion Matrix Data Undersampling*

Akurasi yang didapat lebih kecil dibanding dua skenario data latih sebelumnya yaitu hanya 56%, 51% *precision*, 57% *recall*, dan 52% *f1-score*. Meskipun demikian, teknik *undersampling* memberikan hasil yang berbeda dalam memprediksi benar pada data dengan *class* positif dimana prediksi benar pada *class* minority mengalami peningkatan dibanding dua skenario sebelumnya. Berikut terdapat tabel perbandingan *matrix accuracy*, *precision*, *recall*, dan *f1-score* untuk masing-masing skenario pelatihan.

Tabel 3 Hasil *Classification Report*

Skenario Pelatihan	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Data latih asli	71	66	56	56
Data latih <i>oversampling</i>	71	67	61	61
Data latih <i>undersampling</i>	56	51	57	52

Berdasarkan Tabel 3, dapat dilihat bahwa pelatihan dengan scenario data pelatihan menggunakan data latih *oversampling* memiliki *accuracy*, *precision*, *recall*, dan *f1-score* tertinggi di antara scenario pelatihan lain. SVM dengan data latih *oversampling* juga mengalami peningkatan prediksi benar pada *minority class* yaitu pada *class* netral dan positif.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dengan menggunakan parameter yang sama, model Support Vector Machine (SVM) menggunakan data pelatihan *oversampling* mendapatkan akurasi, recall, precision, dan f1-score tertinggi yaitu 71% accuracy, 67% precision, 61% recall, dan 61% f1-score sedangkan pelatihan dengan menggunakan data pelatihan *undersampling* memiliki performa paling rendah yaitu mendapatkan hasil 56% accuracy, 51% precision, 57% recall, dan 52% f1-score..

References

- [1] S. Handayani, "Peran Logo Dan Sertifikasi Halal Terhadap Niat Beli Konsumen Pada Restoran Cheese Chicken," UIN Syarif Hidayatullah, Jakarta, 2019. Accessed: May 20, 2022. [Online]. Available: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/45975/1/SRI_HANDAYANI-FEB.pdf
- [2] BPJPH, "Badan Penyelenggara Jaminan Produk Halal - BPJPH," <http://halal.go.id/>, 2022. <http://halal.go.id/> (accessed Apr. 24, 2022).
- [3] Kontributor, "Ditetapkan, Label Halal Indonesia Berlaku Nasional." <https://www.kemenag.go.id/read/ditetapkan-label-halal-indonesia-berlaku-nasional-8nja7> (accessed Apr. 11, 2022).
- [4] detikcom, "Kenapa Logo Halal Diganti dan Hal-hal yang Perlu Diketahui soal Label Baru." <https://news.detik.com/berita/d-5983823/kenapa-logo-halal-diganti-dan-hal-hal-yang-perlu-diketahui-soal-label-baru> (accessed Apr. 12, 2022).
- [5] A. R. Kamalina, "Ada Logo Halal Baru, Ekonom: Standardisasi Sertifikasi Juga Harus Berubah - Ekonomi Bisnis.com." <https://ekonomi.bisnis.com/read/20220314/12/1510636/ada-logo-halal-baru-ekonom-standardisasi-sertifikasi-juga-harus-berubah> (accessed Apr. 12, 2022).
- [6] Kompas TV, "Polemik Perubahan Logo Halal Baru, Logo Lama Masih Berlaku Hingga 2026." <https://www.kompas.tv/article/270627/polemik-perubahan-logo-halal-baru-logo-lama-masih-berlaku-hingga-2026> (accessed Apr. 12, 2022).
- [7] B. Pratama *et al.*, "Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods," *Journal of Physics: Conference Series*, vol. 1201, no. 1. 2019. doi: 10.1088/1742-6596/1201/1/012038.
- [8] R. D. Himawan and E. Eliyani, "Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, p. 58, Apr. 2021, doi: 10.26418/jp.v7i1.41728.
- [9] R. H. Muhammadiyah, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," *Khazanah Inform. J. Ilmu Komput. dan Inform. Vol. 8 No. 1 April 2022* DO - 10.23917/khif.v8i1.15213 , Mar. 2022, [Online]. Available: <https://journals.ums.ac.id/index.php/khif/article/view/15213>
- [10] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, "Sentiment Analysis of Tweets Using Machine Learning Approach," *2018 11th International Conference on Contemporary Computing, IC3 2018*. 2018. doi: 10.1109/IC3.2018.8530517.
- [11] K. Felicia Ilona and I. Budi, "Classification of Inundation Level using Tweets in Indonesian Language," *ACM International Conference Proceeding Series*. pp. 137–143, 2021. doi: 10.1145/3457784.3457806.
- [12] K. S. Madhu, B. C. Reddy, C. H. Damarukanadhan, M. Polireddy, and N. Ravinder, "Real

- Time Sentimental Analysis on Twitter,” *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*. pp. 1030–1034, 2021. doi: 10.1109/ICICT50816.2021.9358772.
- [13] A. A. Mir, S. Rathinam, and S. Gul, “Public perception of COVID-19 vaccines from the digital footprints left on Twitter: analyzing positive, neutral and negative sentiments of Twitterati,” *Libr. Hi Tech*, 2021, doi: 10.1108/LHT-08-2021-0261.
- [14] J. Rejito, A. Atthariq, and A. S. Abdullah, “Application of text mining employing k-means algorithms for clustering tweets of Tokopedia,” *Journal of Physics: Conference Series*, vol. 1722, no. 1. 2021. doi: 10.1088/1742-6596/1722/1/012019.
- [15] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, “The effect of the TF-IDF algorithm in times series in forecasting word on social media,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 2, pp. 368–376, 2020, doi: 10.11591/ijeecs.v22.i2.pp368-376.
- [16] A. Borg and M. Boldt, “Using VADER sentiment and SVM for predicting customer response sentiment,” *Expert Syst. Appl.*, vol. 162, 2020, doi: 10.1016/j.eswa.2020.113746.
- [17] S. Radha and C. Nelson Kennedy Babu, “Enterprise big data analysis using SVM classifier and lexicon dictionary,” *Int. J. Enterp. Netw. Manag.*, vol. 11, no. 1, pp. 65–75, 2020, doi: 10.1504/IJENM.2020.103913.