

# Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan *Naive Bayes Classifier*

Aziz Perdana<sup>[1]</sup>, Arief Hermawan<sup>[2]</sup>, Donny Avianto<sup>[3]</sup>

Magister Teknologi Informasi<sup>[1],[2],[3]</sup>

Universitas Teknologi Yogyakarta

Sleman, Indonesia

azizperdana@gmail.com<sup>[1]</sup>, ariefdb@uty.ac.id<sup>[2]</sup>, donny@uty.ac.id<sup>[3]</sup>

**Abstract**—The postponement of Indonesian elections has become a hot and controversial issue recently. This issue was first raised by the general chairman of PKB, Muhaimin Iskandar after receiving Micro Small Medium Enterprises actors, entrepreneurs and economic analysts at the Indonesian House of Representatives Delegation Room, Nusantara III, Jakarta on Wednesday, February 23, 2022. The issue of postponing the election caused public sentiment to split into three, namely negative, neutral, and positive sentiments. To find out the largest percentage of sentiment, a sentiment analysis on twitter tweet data is needed for the period January 1, 2022 until the time this research is conducted (May 27, 2022). This research is expected to map sentiment analysis before Muhaimin Iskandar raises issues (31,831 tweets) and after Muhaimin Iskandar raises issues (151,538 tweets) using the Naive Bayes Classifier model. Sentiment analysis in the time duration before Muhaimin Iskandar raises the issue is needed to test whether the claim of postponing the election is the will of the people is true. Meanwhile, sentiment analysis in the duration of time after Muhaimin Iskandar raised the issue is needed to see the polarization on Twitter, how much support or rejection there is for the issue. The results of the training obtained 98% precision, 100% recall, 99% f1-score for negative sentiment, 100% precision, 94% recall, 96.9% f1-score for neutral sentiment, 96.1% precision, 100% recall, 98% f1-score for positive sentiment. The resulting wordcloud shows that people only discussed the issue of postponing elections after being discussed by Muhaimin. Meanwhile, the results of the prediction of sentiment on the issue of postponing the election after being raised by Muhaimin tended to show a positive sentiment of 60,442 unique tweets, with 58,873 unique tweets indicating negative sentiment, while the remaining 32,223 showed neutral sentiment.

**Keywords**—election postponement, sentiment analysis, naive bayes classifier, twitter

**Abstrak**—Penundaan pemilihan umum (pemilu) Indonesia menjadi isu terhangat dan kontroversial baru-baru ini. Isu ini dilontarkan pertama kali oleh ketua umum PKB, Muhaimin Iskandar setelah menerima pelaku UMKM, pengusaha dan para analis ekonomi di ruang Delegasi DPR RI, Nusantara III, Jakarta pada Rabu, 23 Februari 2022. Isu penundaan pemilu ini menyebabkan sentimen publik terpecah menjadi tiga, yaitu sentimen negatif, netral, dan positif. Untuk mengetahui persentase sentimen terbesar, diperlukan suatu analisis sentimen terhadap data cuitan twitter pada periode waktu 1 Januari 2022

sampai dengan saat penelitian ini dilakukan (27 Mei 2022). Penelitian ini diharapkan dapat memetakan analisis sentimen sebelum Muhaimin Iskandar mengutarakan isu (31.831 data cuitan) dan setelah Muhaimin Iskandar mengutarakan isu (151.538 data cuitan) dengan menggunakan model Naive Bayes Classifier. Analisis sentimen di durasi waktu sebelum Muhaimin Iskandar mengutarakan isu diperlukan untuk menguji, apakah klaim penundaan pemilu adalah hal yang benar-benar diinginkan oleh rakyat. Sedangkan analisis sentimen di durasi waktu setelah Muhaimin Iskandar mengutarakan isu perlu dilakukan untuk melihat polarisasi di twitter, seberapa besar dukungan maupun penolakan terhadap isu tersebut. Hasil dari pelatihan diperoleh nilai precision 98%, recall 100%, f1-score 99% untuk sentimen negatif, precision 100%, recall 94%, f1-score 96,9% untuk sentimen netral, precision 96,1%, recall 100%, f1-score 98% untuk sentimen positif. *Wordcloud* yang dihasilkan menunjukkan masyarakat baru membahas isu penundaan pemilu setelah Muhaimin mengutarakan isu tersebut. Sementara hasil prediksi sentimen terhadap isu penundaan pemilu menggunakan data pasca isu dilontarkan oleh Muhaimin cenderung menunjukkan sentimen positif sebesar 60.442 cuitan unik, dengan 58.873 cuitan unik yang menunjukkan sentimen negatif, sedangkan sisanya, yaitu sebesar 32.223 menunjukkan sentimen netral.

**Kata Kunci**—penundaan pemilu, analisis sentimen, naive bayes classifier, twitter

## I. PENDAHULUAN

### A. Latar Belakang

Isu penundaan pemilihan umum (pemilu) menjadi hal yang menarik untuk dibahas, sejak pertama kali dilontarkan oleh ketua umum PKB Muhaimin Iskandar [1], klaim dukungan penundaan pemilu menjadi suatu hal yang sepatutnya dapat diukur secara ilmiah.

Platform situs jejaring sosial twitter menjadi target untuk menggali *big data* tentang sentimen dukungan penundaan pemilu. Hal ini dikarenakan dari empat situs jejaring sosial yang ada (facebook, twitter, instagram, snapchat), twitter dipilih karena *Information sharing* berada di peringkat ke 6 dari motivasi pengguna menggunakan twitter, di samping 8 motivasi penggunaan situs jejaring sosial lainnya seperti, (1) *Entertainment*, (2) *Convenience*, (3) *Medium Appeal*, (4) *Passing Time*, (5) *Self-Expression*, (6) *Information Sharing*,

(7) *Social Interaction*, dan (8) *Self-Documentation*. [2].

**B. Penelitian Sebelumnya**

Analisis sentimen situasi politik dari data cuitan twitter pernah dilakukan di India, dengan menilai popularitas dari partai BJP dan partai INC menjelang pemilu 2019 di India. Popularitas ini dinilai dengan cara membedakan sarkasme [3].

Sedangkan di Indonesia, Meylan Wongkar dan Apriandy Angdresey dalam jurnalnya yang berjudul 'Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : Twitter' yang diterbitkan di tahun 2019 pada International Conference on Informatics and Computing (ICIC) melakukan analisis sentimen terhadap dua pasangan calon presiden yaitu Jokowi-Ma'ruf Amin dan Prabowo Subianto-Sandiaga Uno. Analisis sentimen dilakukan terhadap data twitter dari periode Januari sampai dengan Mei 2019. Penelitiannya membandingkan akurasi yang dihasilkan antara metode Naive Bayes, Support Vector Machine, dan K-Nearest Neighbor (K-NN). Meylan Wongkar dan Apriandy Angdresey menggunakan library-library dalam python untuk mengambil data twitter, memproses teks, memproses data pelatihan, dan klasifikasi teks Naive Bayes, sedangkan untuk Support Vector Machine, dan K-Nearest Neighbor (K-NN) menggunakan RapidMiner. Hasilnya metode Naive Bayes menghasilkan level akurasi yang lebih tinggi yaitu 80,90 %, sedangkan SVM sebesar 63,99%, dan K-NN sebesar 75,58% [4].

Analisis sentimen terhadap data twitter juga pernah dilakukan untuk melihat sentimen masyarakat di era COVID 19, di antaranya menilai sentimen masyarakat dengan menggunakan VADER (Valence Aware Dictionary and Sentiment Reasoner) terhadap program vaksinasi [5][6] atau bahkan sentimen masyarakat pada saat wabah COVID 19 sedang meninggi [7].

Penelitian-penelitian di atas (yang menggunakan metode *Naive Bayes Classifier*) tidak menyebutkan jumlah angka riil (paling besar 40.000 data) dari data cuitan twitter yang dilakukan prediksi. Sementara analisis sentimen seharusnya dilakukan terhadap data yang besar dalam waktu yang singkat. Hal ini dilakukan agar momentum trending tidak terlewat dan menjadi isu yang ketinggalan zaman. Oleh karena itu diperlukan pembuktian bahwa analisis sentimen dengan model yang dihasilkan oleh *Naive Bayes Classifier* mampu melakukan prediksi sentimen terhadap data yang cukup besar.

**C. Landasan Teori**

Naive Bayes Classifier adalah algoritma *Machine Learning* yang termasuk dalam *supervised classification* [8]. Naive Bayes Classifier diturunkan dari Teorema Bayes, yang mengasumsikan bahwa atribut data secara statistik tidak berkorelasi seperti yang ditunjukkan pada persamaan (1). Naive Bayes banyak digunakan untuk klasifikasi teks dimana salah satu penggunaannya adalah menyaring spam Email, Analisis Sentimen, Sistem Rekomendasi, dan lain-lain [9].

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} \tag{1}$$

dimana

- P(A|B) : probabilitas hipotesis A jika diberikan data B
- P(B|A) : probabilitas data B benar jika hipotesis A benar
- P(A) : probabilitas hipotesis A benar (terlepas dari datanya)
- P(B) : probabilitas data (terlepas dari hipotesis)

**II. METODE PENELITIAN**

Metode Penelitian yang dilakukan adalah ekstraksi data cuitan twitter, pembersihan data, pelabelan sentimen, pelatihan analisis sentimen, dan perhitungan polarisasi sentimen.

**A. Ekstraksi Data Cuitan Twitter**

Ada beragam cara untuk melakukan ekstraksi data cuitan twitter, tentunya dengan kelebihan dan kekurangan masing-masing [10]. Dari twitter API [11], tweepy, TWINT, dan snsrape [5], penulis memilih snsrape yang berbasis bahasa python untuk melakukan ekstraksi data cuitan twitter. Penulis melakukan ekstraksi data cuitan twitter berdasarkan dua durasi waktu, yaitu sebelum Muhaimin Iskandar melontarkan isu (1 Januari 2022 sampai dengan 23 Februari 2022) dan setelah Muhaimin Iskandar melontarkan isu (24 Februari 2022 sampai dengan 27 Mei 2022). Pada proses ini (dan proses-proses selanjutnya), data di simpan dalam file berformat.xlsx. Dari dua data ini, penulis membuat *wordcloud* untuk mengetahui kata-kata yang sering muncul dalam cuitan.

**B. Pembersihan Data**

Agar *wordcloud* dan analisis sentimen dapat mencerminkan hasil yang sebenarnya, harus dilakukan pembersihan data terlebih dahulu. Pembersihan data dilakukan untuk membersihkan *url*, *mention*, *hashtag*, *stopwords*, *emoticon*, tanda baca, dan angka, juga mengecilkan huruf. Pembersihan data *stopwords* pada penelitian ini menggunakan library python pySastrawi [12] dan hasilnya dapat dilihat pada Tabel I.

TABEL I. TABEL DATA EKSTRAKSI SEBELUM DAN SESUDAH PEMBERSIHAN

Sebelum	Sesudah
@msaid_didu Menunda pemilu bagi? proyek belum tuntas pak @msaid_didu Para penguasa lagi sibuk memperkaya diri. Mereka lupa bahwa semua harta tidak di bawah mati.	menunda pemilu bagi proyek belum tuntas pak para penguasa lagi sibuk memperkaya diri mereka lupa bahwa semua harta tidak di bawah mati
KPU: Penundaan Pemilu Inkonstitusional Jka Tak Amandemen UUD 1945 <a href="https://t.co/Jh9gV6bfnn">https://t.co/Jh9gV6bfnn</a>	kpu penundaan pemilu inkonstitusional jika tak amandemen uud
Pemilu tahun ini sajalah, Kami Sudah Capek dikhianati #SaatnyaRezimJokowiTurun #SaatnyaRezimJokowiTurun	pemilu tahun ini sajalah kami sudah capek dikhianati saatnyarezimjokowiturun saatnyarezimjokowiturun

https://t.co/jVUtkVk8BS	
Wah, mungkin yg minta injury time itu partai yg ga siap pemilu	wah mungkin yg minta injury time itu partai yg ga siap pemilu
<input type="checkbox"/> https://t.co/9xDOIdbkAp	

C. Pelabelan Sentimen

Dari data ekstraksi cuitan twitter sejak tanggal 1 Januari 2022 hingga 27 Mei 2022, dipilih 300 data cuitan twitter untuk merepresentasikan contoh dari data cuitan twitter dengan sentimen netral (100 data), positif (100 data), dan negatif (100 data). 300 data ini kemudian yang akan menjadi set data pelatihan (yang sudah dilabeli) dan set data tes (data sama yang tidak dilabeli). Untuk pelabelan ini kelas dari setiap data, penulis menggunakan kode angka dengan ketentuan: 1 sebagai sentimen negatif, 2 sebagai sentimen netral, dan 3 sebagai sentimen positif.

Pelabelan positif (kode=3) tidak berarti kalimat tersebut mendukung penundaan pemilu secara eksplisit, tetapi pelabelan positif juga dilakukan untuk cuitan-cuitan yang membela pemerintahan atau dalam hal ini pak Jokowi.

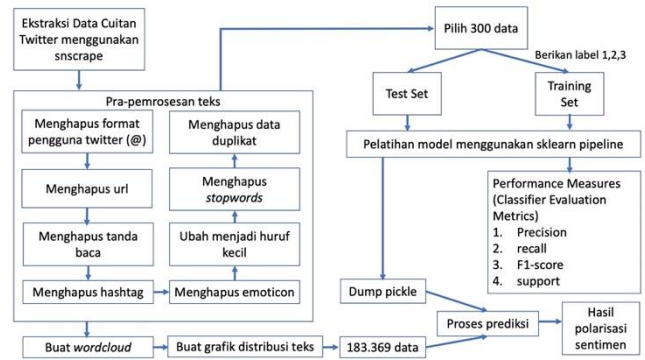
D. Pelatihan Analisis Sentimen

Untuk proses pelatihan, dalam pelaksanaannya menggunakan python dengan scikit-learn Multinomial Naive Bayes [13]. Evaluasi di lakukan dalam bentuk *Classifier Evaluation Matrics* [14]. Model dari pelatihan kemudian disimpan dalam bentuk *pickle* python [15] yang selanjutnya akan digunakan untuk memprediksi polarisasi sentimen dari 151.538 data cuitan unik twitter.

E. Perhitungan Polarisasi Sentimen

Dengan menggunakan model hasil pelatihan yang telah disimpan dalam format *pickle*, dilakukan perhitungan analisis sentimen dari sisa hasil ekstraksi data cuitan twitter yang belum digunakan sebagai set data pelatihan. Data cuitan twitter dengan durasi waktu 24 Februari 2022 hingga 27 Mei 2022 sangat mungkin mencerminkan keberpihakan isu penundaan pemilu (tergantung *wordcloud* yang dihasilkan). Sedangkan data cuitan twitter dengan durasi waktu 1 Januari 2022 sd 23 Februari 2022 mungkin belum mencerminkan keberpihakan isu penundaan pemilu.

Selengkapnya metodologi penelitian dapat digambarkan seperti Gambar 1.



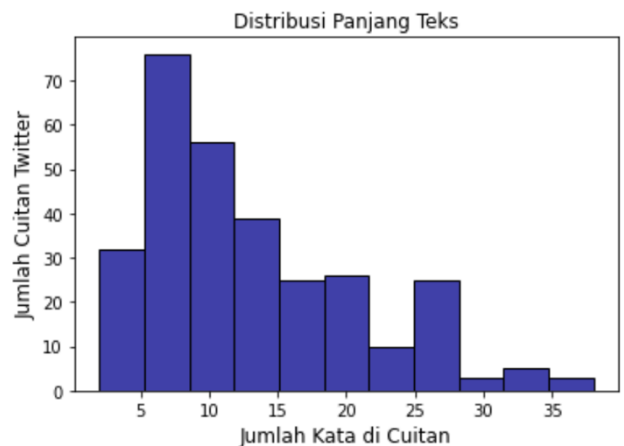
Gambar 1. Tahapan Penelitian

III. HASIL DAN ANALISA

300 data tersebut dapat dilihat karakteristiknya dari *wordcloud* dan grafik distribusi panjang teks seperti pada Gambar 2 dan Gambar 3.



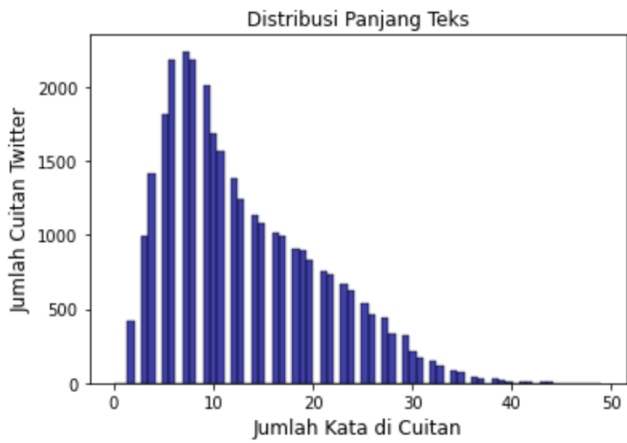
Gambar 2. Wordcloud Data Pelatihan



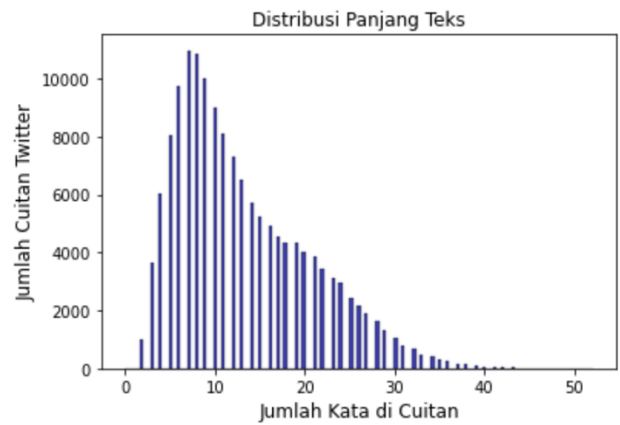
Gambar 3. Distribusi Panjang Teks Data Pelatihan

Dari pelatihan menggunakan 300 data yang sudah dilabeli, didapatkan hasil *Classifier Evaluation Matrics*, yaitu hasil *Precision*, *Recall*, dan *F-measures* seperti pada Gambar 4. Untuk sentimen negatif mendapatkan hasil *precision* 98%, *recall* 100%, *f1-score* 99%. Untuk sentimen netral

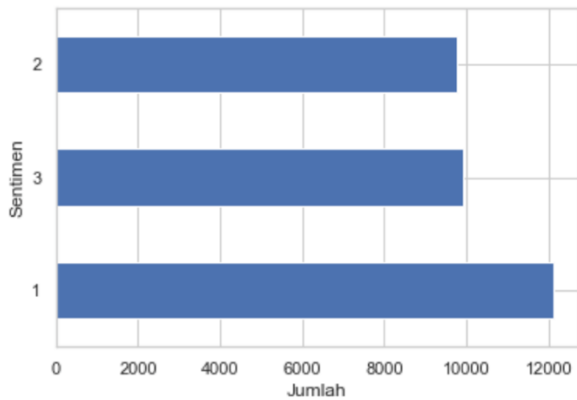




Gambar 6. Distribusi Panjang Cuitan twitter sebelum Muhaimin Melontarkan Isu

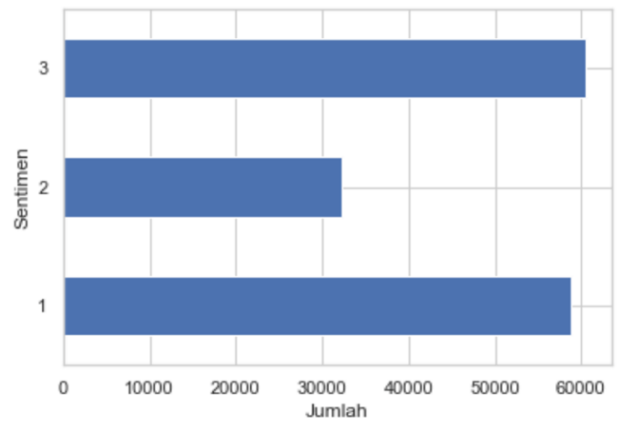


Gambar 9. Distribusi Panjang Cuitan twitter setelah Muhaimin Melontarkan Isu



Gambar 7. Polarisasi Sentimen Sebelum Muhaimin Melontarkan Isu

Adapun hasil analisis sentimen, didapatkan data seperti Gambar 10, yaitu sentimen negatif sebanyak 58.873, sentimen netral sebanyak 32.223, dan sentimen positif sebanyak 60.442



Gambar 10. Polarisasi Sentimen Penundaan Pemilu Setelah Muhaimin Melontarkan Isu

**B. Data Cuitan Twitter Setelah Muhaimin Melontarkan Isu**

Data cuitan twitter periode setelah Muhaimin melontarkan isu adalah sebanyak 151.538. *Wordcloud* dan grafik distribusi panjang cuitan yang dihasilkan dari data ini dapat dilihat pada Gambar 8 dan Gambar 9.



Gambar 8. Wordcloud twitter setelah Muhaimin Melontarkan Isu

Dapat dilihat dari *wordcloud*, kata penundaan, tunda sudah muncul dan cukup besar jumlahnya.

**IV. KESIMPULAN**

Sebelum 23 Februari 2022, publik masih belum sadar akan isu penundaan pemilu, hal itu terlihat dari *wordcloud* yang dihasilkan dari ekstraksi data twitter yang diambil dari 1 Januari 2022 hingga 23 Februari 2022, tidak ada kata penundaan ataupun tunda. Pengguna twitter baru mulai sadar dan mulai membahas penundaan pemilu di periode 24 Februari 2022 sampai 27 Mei 2022. Hal ini terbukti dari munculnya kata penundaan ataupun tunda pada *wordcloud* di dalam periode waktu tersebut. Selama periode setelah Muhaimin melontarkan isu, pengguna twitter condong untuk memberikan sentimen positif dengan jumlah 60.442, dibandingkan dengan 32.223 sentimen netral, maupun 58.873 sentimen negatif. Sentimen positif dalam arti mendukung penundaan pemilu secara eksplisit jarang ditemukan. Sentimen positif hanya menunjukkan pembelaan masyarakat kepada pemerintah (dalam hal ini presiden Jokowi) terhadap munculnya isu tersebut. Penelitian ini juga menunjukkan



analisis sentimen menggunakan *Naive Bayes Classifier* dimungkinkan untuk dilakukan menggunakan data yang cukup besar (151.538 data). Berdasarkan hasil percobaan yang dilakukan, ke depannya perlu dilakukan perbandingan kecepatan prediksi sentimen dengan *Naive Bayes Classifier* antara memakai file berformat *xlsx* dengan basisdata. Kemudian perlu dilakukan perbandingan kecepatan prediksi antara *Naive Bayes Classifier* dengan metode lainnya (VADER, dll).

#### DAFTAR PUSTAKA

- [1] "Gus Muhaimin Usul Pemilu Ditunda atas Masukan Pengusaha dan Pelaku UMKM." Tempo. <https://nasional.tempo.co/read/1564013/gus-muhaimin-usul-pemilu-ditunda-atas-masukan-pengusaha-dan-pelaku-umkm> (accessed 30 Mei 2022).
- [2] Saleem Alhabash and Mengyan Ma, "A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students," *SAGE journals*, p. 13, Februari 2017.
- [3] Arpit Khare, Amisha Gangwar, Sudhakar Singh, Shiv Prakash "Sentiment Analysis and Sarcasm Detection of Indian General Election Tweets," *arXiv*, 2022.
- [4] A. A. Meylan Wongkar, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : Twitter," *International Conference on Informatics and Computing (ICIC)*, pp. 1-5, 2019.
- [5] R. D. Ranganathan Chandrasekaran, Harsh Shah, Vivek Kumar, Evangelos Moustakas, "Examining Public Sentiments and Attitudes Toward COVID-19 Vaccination: Inveillance Study Using Twitter Posts," *JMIR INFODEMOLOGY*, vol. 2, no. 1, 2021.
- [6] R. D. Samira Yousefinaghania, Samira Mubarekab, Andrew Papadopoulos, Shayan Sharif, "An analysis of COVID-19 vaccine sentiments and opinions on Twitter," *International Journal of Infectious Diseases*, vol. 108, pp. 256-262, 2021
- [7] C. A. H. Khairiyah Mohamed Ridhwan, "Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore," *International Journal of Information Management Data Insights*, vol. 1, no. 2, 2021.
- [8] A. H. K. Mohammed H. Alsharif, Khalid Yahya, and Shehzad Ashraf Chaudhry, "Machine Learning Algorithms for Smart Data Analysis in Internet of Things Environment: Taxonomies and Research Trends," *Symmetry*, 2020.
- [9] S. Sawla, "Introduction to Naive Bayes for Classification." [Online]. Available: <https://medium.com/@srishtisawla/introduction-to-naive-bayes-for-classification-baefeb43a2d>
- [10] A. Troya, "Aspect-based Sentiment Analysis of Social Media Data with Pre-trained Language Models," MASTER OF SCIENCE, FACULTY OF SCIENCE, UNIVERSITY OF AMSTERDAM, 2021.
- [11] R. H. P. Akilandeswari, Sumanth.KO.M, "Sentiment Analysis using Machine Learning through Twitter Streaming API," *International Journal of Engineering & Technology*, vol. 7, pp. 1168-1174, 2018.
- [12] A. W. Ariel Elbert Budiman, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, 2020.
- [13] S. Mukhopadhyay, *Advanced Data Analytics Using Python With Machine Learning, Deep Learning and NLP Examples*. Apress, 2018.
- [14] F. K. A. b. Amira M. Idrees, "A COLLABORATIVE EVALUATION METRICS APPROACH FOR CLASSIFICATION ALGORITHMS," *JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY*, vol. 55, no. 1, 2020.
- [15] D. P. R. Willy, Samsuryadi, "Perbandingan Algoritma Random Forest Classifier, Support Vector Machine dan Logistic Regression Clasifier Pada Masalah High Dimension (Studi Kasus: Klasifikasi Fake News)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 4, pp. 1720-1728, 2021.