

## Algoritma Support Vector Machine Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Pandemi Virus Corona Di Media Sosial

Mhd.Furqan<sup>1</sup>, Mhd. Ikhsan<sup>2</sup>, Rafizah Aini<sup>3</sup>

<sup>1,2,3</sup>Prodi Ilmu Komputer Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Indonesia

E-mail: <sup>1</sup>mfurqan@uinsu.ac.id, <sup>2</sup>mhd.ikhsan@uinsu.ac.id,  
<sup>3</sup>rafizahaini9@gmail.com

### Abstract

The corona virus pandemic refers to the spread of coronavirus disease 2019 or what is known as coronavirus disease 2019 in various parts of the world. This outbreak is a change from a new type of coronavirus called SARS-CoV-2. The issue of this pandemic has become a hot topic of discussion, including on social media. The most frequently used platform among the public is Twitter. On social media, the corona virus pandemic has always been a topic of conversation that is often discussed, causing controversy. Controversy occurs because every day the opinions on social media Twitter regarding the corona virus pandemic are always increasing so that, when people read news on social media about the pandemic, it raises concerns because people's opinions are different. From this problem the author will create a system that analyzes opinions from Twitter social media to get opinion sentiment about what is happening in the community regarding the problem of the corona virus pandemic. This study uses the SVM method which is fast and effective for text classification. The results of this study will classify positive, negative and neutral sentences. The accuracy obtained from the model with the SVM algorithm is 98%. Testing is done by calculating precision, recall, F-measure..

**Keyword:** Twitter, opinion, algorithm, classification

### Abstrak

Pandemi virus corona merujuk pada penyebaran penyakit koronavirus 2019 atau yang dikenal dengan sebutan coronavirus disease 2019 di berbagai belahan dunia. Wabah ini merupakan perubahan dari virus jenis koronavirus baru yang dinamai SARS-CoV-2. Isu pandemi ini menjadi perbincangan hangat, termasuk di media sosial. Platform yang paling sering digunakan di antara masyarakat adalah Twitter. Di media sosial twitter pandemi virus corona selalu menjadi topik pembicaraan yang sering di bahas hingga menimbulkan kontroversi. Kontroversi terjadi dikarenakan setiap harinya opini-opini di media sosial twitter terkait pandemi virus corona selalu bertambah sehingga, ketika masyarakat membaca berita di media sosial mengenai pandemi tersebut menimbulkan kekhawatiran dikarenakan opini masyarakat yang berbeda-beda. Dari permasalahan ini penulis akan membuat sebuah sistem yang menganalisis opini dari media sosial twitter untuk mendapatkan sentimen opini apa yang terjadi pada masyarakat mengenai permasalahan pandemi virus corona tersebut. Penelitian ini menggunakan metode SVM yang cepat dan efektif untuk klasifikasi teks. Hasil penelitian ini akan melakukan klasifikasi kalimat positif, negatif dan netral. Akurasi yang diperoleh dari model dengan algoritma SVM yaitu 98%. Pengujian dilakukan dengan perhitungan precision, recall, F-measure.

**Kata kunci:** Twitter, opini, algoritma, klasifikasi

## 1. Pendahuluan

Pandemi virus corona merupakan kejadian penyebaran penyakit Covid-19 di seluruh penjuru dunia. Virus ini dinamakan SARS-CoV-2[1][2]. Virus ini muncul di Kota Wuhan, Provinsi Hubei, Tiongkok pada tahun 2019 di bulan september. Dampak dari penyakit ini telah dilaporkan lebih dari 30 juta kasus pandemi virus corona disemua negara.[3][4]. Pandemi ini tentunya menimbulkan kekhawatiran bagi masyarakat, tak terkecuali masyarakat Indonesia. Pandemi yang terjadi di seluruh dunia menjadi topik pembicaraan, termasuk di media sosial. Twitter salah satu media sosial yang sangat digemari. Di media sosial *twitter* pandemi virus corona selalu menjadi topik pembicaraan yang sering di bahas hingga menimbulkan kontroversi. Kontroversi terjadi dikarenakan setiap harinya opini-opini di media sosial *twitter* terkait pandemi virus corona selalu bertambah sehingga, ketika masyarakat membaca berita di media sosial mengenai pandemi tersebut menimbulkan kekhawatiran dikarenakan opini masyarakat yang berbeda-beda[5][6]. Pendapat masyarakat yang beragam disebabkan oleh adanya peningkatan kasus COVID-19 di Indonesia, yang pada gilirannya menimbulkan kekhawatiran di kalangan masyarakat. Hal ini berdampak pada kenaikan harga masker dan hand sanitizer yang tidak wajar, serta diakibatkan oleh pandangan bahwa kebijakan pemerintah dalam menangani penyebaran virus dianggap kurang serius dan kurang tanggap. Selain itu, transparansi pemberitaan mengenai kasus-kasus tersebut juga dinilai tidak memadai, dan kurangnya ketaatan masyarakat terhadap kebijakan sosial distancing yang telah ditetapkan pemerintah menyebabkan peningkatan jumlah kasus dan menimbulkan kepanikan di kalangan masyarakat [7][8].

Oleh karena itu perlu dibutuhkan sebuah sistem untuk mengetahui dan mengklasifikasikan sentimen opini dari masyarakat mengenai virus tersebut. Dari permasalahan ini penulis akan membuat sebuah sistem yang menganalisis opini dari media sosial *twitter* untuk mendapatkan sentimen opini apa yang terjadi pada masyarakat mengenai permasalahan pandemi virus corona tersebut[9]. Dalam membuat sistem ini penulis menggunakan analisis sentimen dalam menganalisis opini di media sosial. Pada bidang analisis sentimen akan melakukan ekstraksi kata berbentuk teks menjadi sebuah vektor kemudian diberi label untuk digunakan sebagai informasi pada kalimat[10]. Analisis sentimen digunakan untuk mengklasifikasikan polaritas teks, dengan tujuan untuk menentukan apakah opini yang terkandung dalam sebuah tweet bersifat positif atau negatif. Dalam konteks pandemi virus corona, analisis sentimen diterapkan untuk mengekstrak informasi dari opini masyarakat, apakah mengandung sentimen positif atau negatif terhadap situasi tersebut. Penulis akan menerapkan metode SVM dalam mengembangkan sistem ini, karena SVM merupakan metode klasifikasi teks yang cepat dan efektif[11]. Metode Support Vector Machine telah terbukti sangat efektif dalam mengkategorisasi teks secara tradisional dan mengungguli metode lain, termasuk metode Support Vector Machine itu sendiri. Metode Support Vector Machine dipilih untuk penelitian ini karena memiliki beberapa kelebihan, antara lain kemampuannya dalam menghasilkan akurasi yang tinggi meskipun memproses dataset yang besar, kemampuannya bekerja dengan baik dalam berbagai dimensi fitur (baik linier maupun nonlinear), serta memberikan perlindungan dari overfitting, sehingga tidak terlalu tergantung pada jumlah fitur dalam dataset[12].

## 2. Metodologi Penelitian

Metode penelitian berisi rincian mengenai langkah-langkah yang akan diambil dalam menjalankan proses penelitian. Kerangka kerja ini memberikan Gambaran tentang tahapan-tahapan yang akan diikuti untuk menyelesaikan permasalahan yang dibahas. Berikut adalah tahapan-tahapan metode penelitian yang akan digunakan:



**Gambar 1.** Metode penelitian

#### a) Analisis Data

Analisis sentimen adalah proses pemrosesan bahasa alami yang berfungsi untuk menganalisis pendapat dan pandangan seseorang, baik yang disampaikan dalam bentuk tulisan maupun lisan. Proses ini juga melibatkan analisis perilaku dan emosi individu terkait dengan suatu produk, penilaian, evaluasi, organisasi, kejadian, atau topik tertentu. Data yang dihasilkan dari analisis sentimen memberikan informasi berharga yang mendukung pengambilan kesimpulan bagi individu atau organisasi. Dikarenakan volume data yang besar, analisis sentimen tidak dapat dilakukan secara manual oleh manusia. Untuk alasan ini, metode klasifikasi seperti metode naïve bayes digunakan untuk memfasilitasi analisis sentimen. Dalam penelitian ini, dilakukan analisis sentimen terkait pandemi COVID-19 dengan fokus pada pendapat masyarakat tentang konsep "new normal" yang diekspresikan melalui media sosial Twitter. Data dikumpulkan dengan memanfaatkan teknik crawling dengan fitur yang disediakan oleh Twitter. Sebanyak 1000 tweet dengan kata kunci "#NewNormal" diambil sebagai sampel data, dengan syarat bahwa data yang diambil harus berupa teks berbahasa Indonesia dan tidak mengandung Gambar. Contoh data tweet yang berhasil dikumpulkan melalui proses crawling tersebut ditampilkan dalam Tabel 1 dan menjadi bagian dari penelitian ini.

**Tabel 1.** Data tweet

| No | Data Tweet                                                                                                                                                                                                                                                                                                                                                                                 |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | <a href="https://bit.ly/2L6CcbBA">https://bit.ly/2L6CcbBA</a>   Seharusnya saat ini Pemerintah Lampung bisa memberikan bantuan kepada mahasiswa tersebut. Mengingat begitu derasnya donasi yang masuk ke Pemprov Lampung melalui tim gugus tugas COVID-19. <a href="https://pic.twitter.com/AndU9vXtJG">pic.twitter.com/AndU9vXtJG</a>                                                     |
| 2  | Klo kata gw Pemerintah tuh lagi menerapkan Herd immunity karena kemampuan pemerintah nihil soal pemberantasan covid. Gambling udah siapa kuat dia bertahan hidup yang lemah maka mati. Padahal dinegara lain begitu serius menangani virus tsb.                                                                                                                                            |
| 3  | Saat ini yang bisa saya lakukan hanya menyiapkan makanan untuk berbuka dan sahur untuk 10 kepala keluarga terdampak covid di Tasik dan tidak dpt bantuan dr pemerintah krn tidak masuk di data. Untuk akhir Ramadhan kita sedang rencanakan paket sembako lengkap. Semoga ada rejekinya.                                                                                                   |
| 4  | Satu Warga Positif Corona, Bupati: Kondisi ini Menuntut Kebersamaan Kita: RAKYAT NEWS, BANTAENG æ" Pemerintah Kabupaten Bantaeng melalui tim gugus tugas penanganan Covid-19 di Bantaeng merilis. [æ ] <a href="http://dlvr.it/RWBnbDA">http://dlvr.it/RWBnbDA</a> #rakyadotnews #referensiterkini #rakyatnews <a href="https://pic.twitter.com/jOwYbh3Txw">pic.twitter.com/jOwYbh3Txw</a> |
| 5  | emosi bgt. lg kondisi begini gue disuruh liputan tentang covid. ini dosen gue bercanda ga si? pemerintah aja nyuruh #dirumahaja masa gue disuruh liputan keluar rumahðYz                                                                                                                                                                                                                   |
| 6  | ANTISIPA penyebaran pandemi COVID-19 membutuhkan kerjasama antara pemerintah dan masyarakat. Pandemi ini adalah musibah berkatagori non alam.                                                                                                                                                                                                                                              |
| 7  | berbahaya covid 19 pemerintah mengatasinya                                                                                                                                                                                                                                                                                                                                                 |

|    |                                                                                                                                                                                                                                          |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 8  | ya sepiantanya <u>ciduk bebas sih</u> jrx berpendapat covid konspirasi disebar luaskan pemikiran akal akalanya pemerintah penyakitnya nyata toxic bego                                                                                   |
| 9  | denny galau banget nih disuruh mudik orang tuanya mudik <u>dilarang pemerintah menghentikan penyebaran covid 19</u> nih sih alpa disugesti tukang sembuhan kesurupan yuk ikuti keseruannya <u>sahursegerr15 gooddoctorid baikbersama</u> |
| 10 | salah keluarga menit <u>meninggal gara gara covid 19</u> pemerintah santai                                                                                                                                                               |

#### b) Preprocessing Data

Pada langkah ini akan menggunakan data berupa teks hasil dari proses crawling, dan akan menerapkan teknik preprocessing untuk membersihkan dan membuat data terstruktur, preprocessing teks bertujuan untuk merubah data yang bersifat tidak terstruktur menjadi terstruktur [9]. Berikut adalah tahapan preprocessing data yang akan dilakukan

1. Tahap Cleaning: Pada langkah ini, data tweet akan dibersihkan atau karakter-karakter yang tidak diperlukan akan dihilangkan, seperti tanda baca, angka, URL, nama pengguna. Karakter-karakter tersebut dapat dihapus atau diganti dengan karakter tertentu. Contohnya adalah ("~&?!><#%{}([0-9]+;:')[1122].
2. Tahap Case Folding: Tahap ini bertujuan untuk melakukan perubahan seluruh huruf besar dalam data menjadi huruf kecil, sehingga tidak ada perbedaan antara kata yang ditulis dengan huruf kapital atau huruf kecil.
3. Tahap Normalisasi: Proses normalisasi dilakukan untuk mengubah kata-kata yang disingkat menjadi kata yang memiliki arti yang sama berdasarkan Kamus Besar Bahasa Indonesia (KBBI), agar data menjadi lebih mudah untuk diproses. Sebagai contoh, "utk" diubah menjadi "untuk" dan "yg" diubah menjadi "yang".
4. Tahap Stemming: Pada tahap stemming, data yang mengandung sifat imbuhan akan dihapus, termasuk awalan, akhiran, dan sisipan. Tujuannya adalah untuk mengembalikan kata ke bentuk dasarnya.
5. Tahap Filtering: Proses filtering dilakukan untuk menghapus kata-kata yang tidak memiliki makna dalam data. Pada tahap ini, digunakan algoritma stoplist atau stopword yang berisi daftar kata-kata umum yang dianggap tidak relevan dalam analisis,

### 3. Hasil dan Pembahasan

Penelitian ini akan memanfaatkan metode SVM untuk mencapai tingkat akurasi tertinggi dalam mengklasifikasikan data berdasarkan sentimen terhadap data virus Covid-19 di media sosial Twitter.

#### a) Preprocessing Data

Proses pra-pemrosesan data terdiri dari beberapa langkah, termasuk pembersihan (cleaning), (case folding), normalisasi, penghapusan awalan dan akhiran kata (stemming), penyaringan (filtering), dan pemisahan teks menjadi kata-kata (tokenizing). Contoh data hasil pra-pemrosesan yang ditampilkan dalam Tabel 2.

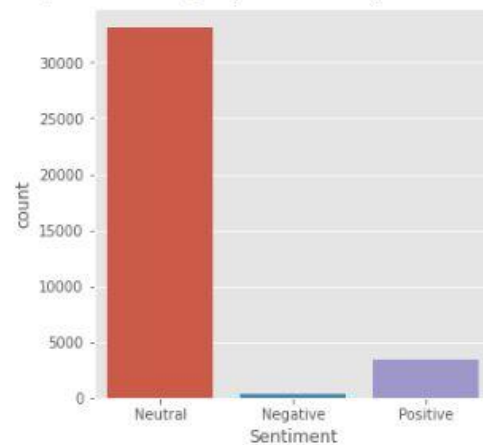
**Tabel 2.** Hasil preprocessing data

| No | data tweet                                                                                                                                                     |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | pemerintah lampung bantuan mahasiswa derasnya donasi masuk pemerintah provinsi lampung tim gugus tugas covid 19                                                |
| 2  | pemerintah menerapkan herd immunity kemampuan pemerintah nihil pemberantasan covid gambling kuat bertahan hidup lemah mati negara serius menangani virus       |
| 3  | lakukan makanan berbuka sahur 10 kepala keluarga terdampak covid tasik bantuan pemerintah masuk data ramadan rencanakan paket sembako lengkap semoga rejekinya |

|    |                                                                                                                                                                                                            |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4. | warga positif corona bupati kondisi menuntut kebersamaan rakyat news bantaeng pemerintah kabupaten bantaeng tim gugus tugas penanganan covid 19 bantaeng menulis rakyatdotnews referensiterkini rakyatnews |
| 5. | emosi banget kondisi disuruh liputan covid dosen bercanda sih pemerintah menyuruh dirumahaja disuruh liputan rumah                                                                                         |

### b) Distribusi Kata

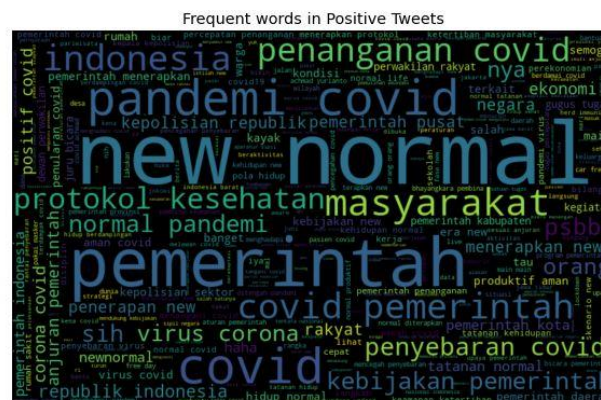
Distribusi kata adalah kumpulan kata dengan frekuensi tertinggi yang ditemukan dalam data set yang diperoleh dari proses crawling. Pada Gambar 2 akan disajikan distribusi kata



**Gambar 2.** Distribusi Kata

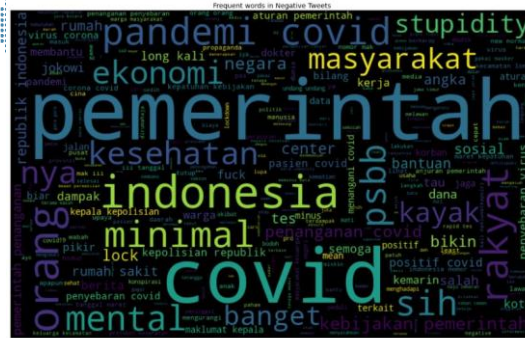
c) Word Cloud

Word Cloud adalah salah satu teknik untuk memvisualisasikan data teks secara grafis. Metode ini sering digunakan dalam analisis teks karena kemudahannya dalam pemahaman. Berikut adalah word cloud yang akan menampilkan hasil terhadap kata yang mengandung makna positif, netral, dan negatif terhadap dataset yang terdapat pada Gambar 3



**Gambar 3.** *Word Cloud* Kata Tweet dengan label positif

Keterangan Gambar 4 menjelaskan *word cloud* memiliki fungsi untuk memunculkan citraan visual dari sebuah data tertulis kemudian dapat dijadikan sebagai sebuah alat bantu dalam melakukan sentimen analisis.



Gambar 4. Word Cloud Kata Tweet dengan label negatif



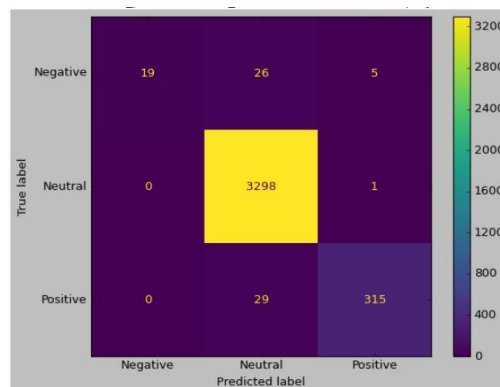
Gambar 5. Word Cloud Kata Tweet dengan label Netral

#### d) Confussion Matrix

Merupakan sebuah proses dalam melihat performa yang digunakan untuk evaluasi terhadap proses klasifikasi. Masalah klasifikasi ini melibatkan keluaran yang dapat berada dalam dua kelas atau lebih. Pada Gambar 6 hasil dari coonfusion matriks

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Negative     | 1.00      | 0.38   | 0.55     | 50      |
| Neutral      | 0.98      | 1.00   | 0.99     | 3299    |
| Positive     | 0.98      | 0.92   | 0.95     | 344     |
| accuracy     |           |        | 0.98     | 3693    |
| macro avg    | 0.99      | 0.77   | 0.83     | 3693    |
| weighted avg | 0.98      | 0.98   | 0.98     | 3693    |

Gambar 6. Hasil Precision, Recall, dan F1-score



Gambar 7. Confussion Matrix

Pada Gambar 7 di tampilkan bahwa dalam model klasifikasi terdapat hasil evaluasi dengan confusion matriks yang terdapat hasil dari nilai precision, recall dan f1 score sebagai berikut

$$\frac{TP + TNe + TN}{TP + TNe + TN + FP + FNe + FN} \times 100 = \frac{19 + 3298 + 315}{19 + 26 + 5 + 0 + 3298 + 1 + 0 + 29 + 315} \times 100 = \frac{3632}{3693} \times 100 = 98\%$$

Akurasi yang dihasilkan oleh model klasifikasi

$$\text{Akurasi Skor} = 0,98 \times 100 = 98\%$$

#### 4. Kesimpulan

Penelitian menghasilkan model yang dapat melakukan klasifikasi sentimen dengan algoritma SVM, pada penelitian ini terdapat kesimpulan bahwa algoritma SVM dapat digunakan untuk melakukan analisis sentimen terhadap topik penelitian mengenai pembelajaran covid-19. Sistem saat ini belum dapat melakukan filtrasi kata-kata asing pada tweet yang diambil, dimana tweet tersebut merupakan campuran bahasa Indonesia dan asing. Hasil uji akurasi menunjukkan angka tertinggi, yakni mencapai 98%. Penentuan label untuk kelas positif, negatif, dan netral bergantung pada perhitungan polaritas

#### Daftar Pustaka

- [1] A. R. Lubis, S. Prayudani, M. Lubis, and O. Nugroho, "Sentiment Analysis on Online Learning During the Covid-19 Pandemic Based on Opinions on Twitter using KNN Method," in *2022 1st International Conference on Information System & Information Technology (ICISIT)*, 2022, pp. 106–111.
- [2] A. R. Lubis, S. Prayudani, M. Lubis, and Al-Khowarizmi, "The Effect of E-Commerce Towards Sales Growth on Social Media among Students in Indonesia," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2021-Octob, no. October, pp. 102–106, 2021, doi: 10.23919/EECSI53397.2021.9624290.
- [3] A. Susilo *et al.*, "Coronavirus Disease 2019: Tinjauan Literatur Terkini," *J. Penyakit Dalam Indones.*, vol. 7, no. 1, p. 45, 2020, doi: 10.7454/jpdi.v7i1.415.
- [4] Y. Niar, K. Komariah, A. Surip, R. Saputra, and I. Ali, "Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Rotan," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 1, pp. 28–34, 2022, doi: 10.32485/kopertip.v4i1.112.
- [5] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 2, pp. 1–9, 2021.
- [6] P. Sv *et al.*, "Twitter-Based Sentiment Analysis and Topic Modeling of Social Media Posts Using Natural Language Processing, to Understand People's Perspectives Regarding COVID-19 Booster Vaccine Shots in India: Crucial to Expanding Vaccination Coverage," *Vaccines*, vol. 10, no. 11, p. 1929, 2022.
- [7] J. Schüttler, R. Schlickeiser, F. Schlickeiser, and M. Kröger, "Covid-19 Predictions Using a Gauss Model, Based on Data from April 2," *Physics (College. Park. Md.)*, vol. 2, no. 2, pp. 197–212, 2020, doi: 10.3390/physics2020013.
- [8] J. M. Banda *et al.*, "A large-scale COVID-19 Twitter chatter dataset for open scientific research—an international collaboration," *Epidemiologia*, vol. 2, no. 3, pp. 315–324, 2021.
- [9] Y. Drias and H. Drias, "Sentiment Evolution Analysis and Association Rule Mining for COVID-19 Tweets," *J. Digit. Art Humanit.*, pp. 3–21, 2021.
- [10] F. Gao, G. Huang, S. Li, Z. Huang, and L. Chai, "Integrating the

- Eigendecomposition Approach and k-Means Clustering for Inferring Building Functions with Location-Based Social Media Data,” *ISPRS Int. J. Geo-Information*, vol. 10, no. 12, p. 834, 2021.
- [11] D. Y. Liliana, A. Hardianto, and M. Ridok, “Indonesian news classification using support vector machine,” *World Acad. Sci. Eng. Technol.*, vol. 81, no. 9, pp. 767–770, 2011, doi: 10.5281/zenodo.1074438.
- [12] K. V. Veena and D. Mathew, “Speaker identification and verification of noisy speech using multitaper MFCC and Gaussian Mixture models,” *Proc. 2015 IEEE Int. Conf. Power, Instrumentation, Control Comput. PICC 2015*, no. Lc, 2016, doi: 10.1109/PICC.2015.7455806.
- [13] B. Alexandru-costin, “Comparison of Deep Learning Models for Automatic Detection of Sarcasm Context on the MUSTARD Dataset,” 2023.
- [14] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, “Effective text data preprocessing technique for sentiment analysis in social media data,” *Proc. 2019 11th Int. Conf. Knowl. Syst. Eng. KSE 2019*, pp. 1–8, 2019, doi: 10.1109/KSE.2019.8919368.