



## *Analysis of Public Opinion Sentiment About Covid-19 Vaccination in Indonesia Using Naïve Bayes and Decision Tree*

### **Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decision Tree**

Ahmad Harun<sup>1</sup>, Dea Putri Ananda<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,  
Universitas Islam Negeri Sultan Syarif Kasim, Riau

E-Mail: <sup>1</sup>11850310375@students.uin-suska.ac.id, <sup>2</sup>11850320366@students.uin-suska.ac.id

Received March 01<sup>st</sup> 2021; Revised May 5<sup>st</sup> 2021; Accepted May 19<sup>th</sup> 2021  
Corresponding Author: Dea Putri Ananda

#### **Abstract**

COVID-19 is a new disease reported in Wuhan China in December 2019. Based on [www.covid19.go.id](http://www.covid19.go.id), in Indonesia as of January 20, 2021 there were more than 939 thousand cases and more than 26 thousand cases which resulted in death. The rapid spread of COVID-19 and the dangers it causes, the Indonesian government is taking precautions with vaccination, the information of which has been spread across various social media, including the facebook page owned by the Ministry of Health. Facebook pages that have a comment feature on their posts have not been able to determine the amount of user sentiment about positive or negative comments automatically. Sentiment analysis is part of text mining for grouping text polarity in knowing the polarity of an opinion given is positive or negative using a certain algorithm. This study aims to provide results of public opinion about the sentiment analysis of COVID-19 vaccination using the Naïve Bayes Classifier (NBC) algorithm and the Decision Tree and to compare the accuracy of the two algorithms. The results of the research on the analysis of public opinion about the COVID-19 vaccination that have been carried out, tend to be negative responses with an accuracy value of 100.00% using the NBC algorithm and 50.39% using the Decision Tree algorithm.

Keyword: Covid-19, Decision Tree, Facebook, NBC, Sentiment Analysis.

#### **Abstrak**

COVID-19 adalah penyakit baru yang dilaporkan di Wuhan China pada Desember 2019. Berdasarkan dari [www.covid19.go.id](http://www.covid19.go.id), di Indonesia per tanggal 20 Januari 2021 terdapat 939 ribu lebih kasus dan sebanyak lebih dari 26 ribu kasus yang mengakibatkan kematian. Pesatnya penyebaran COVID-19 dan bahaya yang ditimbulkan, pemerintahan Indonesia melakukan pencegahan dengan vaksinasi yang informasinya sudah tersebar diberbagai media sosial, diantaranya ialah facebook page yang dimiliki oleh Kementerian Kesehatan. Facebook page yang memiliki fitur komentar pada postingannya, belum dapat menentukan besar sentimen pengguna terhadap komentar positif atau negatif secara otomatis. Analisa sentimen merupakan bagian dari teks mining untuk pengelompokan polaritas teks dalam mengetahui polaritas suatu opini yang diberikan bersifat positif atau negatif menggunakan algoritma tertentu. Penelitian ini bertujuan untuk memberikan hasil opini masyarakat tentang analisa sentimen vaksinasi COVID-19 menggunakan algoritma Naïve Bayes Classifier (NBC) dan Decision Tree serta membandingkan tingkat akurasi kedua algoritma tersebut. Hasil penelitian analisa sentimen opini masyarakat tentang vaksinasi COVID-19 yang telah dilakukan, cenderung ke tanggapan negatif dengan nilai akurasi 100.00% menggunakan algoritma NBC dan 50.39% menggunakan algoritma Decision Tree.

Keyword: Analisa Sentimen, Covid-19, Decision Tree, Facebook, NBC.

#### **1. PENDAHULUAN**

Corona Virus Disease 2019 (COVID-19) adalah penyakit baru yang dilaporkan di Wuhan China pada Desember 2019 [1]. Meskipun pusat penyebaran virus tersebut berada di Kota Wuhan, China, kini virus tersebut telah tersebar menjangkiti keseluruhan masyarakat dunia, berdasarkan sumber dari [www.covid19.go.id](http://www.covid19.go.id) di Indonesia sendiri per tanggal 20 Januari 2021 terdapat 939 ribu lebih kasus dan sebanyak lebih dari 26 ribu lebih kasus yang mengakibatkan kematian. Dilihat dari pesatnya penyebaran COVID-19 serta bahaya yang

ditimbulkan bila tidak lekas ditangani, pemerintah Indonesia mengembangkan vaksin sebagai salah satu cara menghambat penyebaran. Vaksin tidak hanya melindungi mereka yang telah divaksinasi namun juga akan melindungi masyarakat luas dengan mengurangi penyebaran penyakit yang terjadi didalam populasi tersebut [2].

Informasi vaksinasi serta tata cara pencegahan virus ini telah tersebar diberbagai media sosial [3]. Media sosial ialah salah satu sumber yang sangat umum digunakan untuk berkomunikasi, berbagi dokumen serta data dengan jumlah komunitas yang besar [4]. Salah satu media sosial tersebut yakni Facebook yang di dalamnya ada terdapat informasi yang sangat berharga sebagai alat penentu kebijakan dengan jumlah opini yang besar[5].

Facebook mempunyai banyak fitur yang bisa digunakan oleh para penggunanya, namun yang sangat sesuai untuk lembaga pemerintahan ialah Facebook Page. Facebook Page dapat digunakan untuk menyampaikan informasi kepada para pengikutnya. Pihak instansi pemerintahan seperti Kementerian Kesehatan menggunakan Facebook Page untuk memberikan informasi sehingga pengguna Facebook dapat meng-like dan memberikan komentar pada Facebook Page tersebut. Komentar dari pengikut dapat berupa komentar persetujuan maupun penolakan (pro atau kontra). Namun informasi yang berasal dari komentar-komentar yang ada pada Facebook belum bisa mengidentifikasi seberapa besar sentimen pengguna (pro ataupun kontra) dari komentar-komentar tersebut, baik positif maupun negatif secara otomatis[6].

Teknologi komputer memberikan peluang yang sangat besar untuk memerangi wabah penyakit menular dan memiliki peran yang luar biasa, terutama dalam analisis sentimen untuk media sosial [7]. Analisa sentimen adalah bagian dari teks mining yang dapat mengelompokkan polaritas dari teks. Pengelompokan dilakukan untuk mengetahui bagaimana polaritas dari suatu komentar apakah opini yang diberikan bersifat positif atau negatif [8].

Ada beberapa algoritma yang dapat digunakan dalam melakukan analisa sentimen, diantaranya ialah Naïve Bayes Classifier (NBC) dan Decission Tree. NBC dan Decission Tree digunakan untuk menilai akurasi dalam analisis data. Kelebihan NBC yaitu dalam algoritmanya yang sederhana, NBC hanya membutuhkan data pelatihan yang kecil dan terhadap atribut yang tidak relevan dapat diandalkan [9], sedangkan kelebihan Decission Tree dapat menghasilkan akurasi yang baik dalam penelitian mengklasifikasi dan prediksi terhadap mahasiswa yang lulus tepat waktu [10].

Penelitian ini akan membahas bagaimana respon dan opini masyarakat Indonesia terhadap vaksin COVID-19 menggunakan data yang bersumber dari media sosial Facebook dengan melakukan analisis sentimen dan mengklasifikasikan respon masyarakat tersebut ke dalam sentimen positif dan negatif menggunakan algoritma NBC dan Decission Tree. Tujuan penelitian ini ialah untuk memberikan gambaran kepada masyarakat apakah vaksinasi COVID-19 cenderung ke opini positif atau negatif, selain itu untuk membandingkan tingkat akurasi NBC dengan decision tree.

## 2. METODOLOGI PENELITIAN

Penelitian ini menerapkan algoritma NBC dan algoritma Decision Tree untuk melihat hasil akurasi terbaik analisa sentimen opini publik tentang vaksinasi COVID-19 di Indonesia pada komentar Facebook Page Kementerian Kesehatan. Alur penelitian dapat dilihat pada Gambar 1.

### 1.1 *Pre-Processing*

*Pre-processing* merupakan kegiatan pengumpulan dan penggunaan data historis dalam menemukan kesesuaian pola atau hubungan pada data besar [11]. *Pre-processing* digunakan untuk menghasilkan model yang menjadi kelas penting dari data. Teknik *pre-processing* memiliki pengaruh yang signifikan terhadap performa algoritma pembelajaran mesin. Perancangan database dan analisis yang baik dapat mengurangi masalah kehilangan data melalui *processing* [4].

### 1.2 *Tokenize*

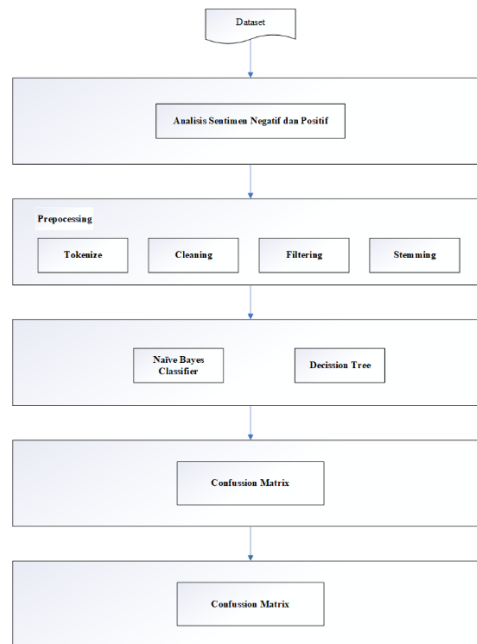
Menurut Santoso, *et al.*, “*tokenize* merupakan proses memisahkan kata-kata pada setiap kalimat menjadi kata tersendiri, sedangkan untuk tanda baca akan dihilangkan”.

### 1.3 *Stemming*

Akar adalah sekelompok kata alami dengan arti yang sama (atau sangat mirip). Metode ini menjelaskan dasar dari kata tertentu. *Stemming infleksi* dan *derivasional* adalah dua jenis metode. Salah satu algoritma yang populer untuk *stemming* adalah algoritma Porter. misalnya jika sebuah dokumen berkaitan dengan kata seperti pengunduran diri, mengundurkan diri, mengundurkan diri maka itu akan dianggap mengundurkan diri setelah menerapkan metode *stemming* [12].

## 1.4 Klasifikasi

Klasifikasi dikategorikan kedalam metode *supervised learning* yang menemukan relasi atribut *input* dan antribut target yang bertujuan untuk meningkatkan keunggulan hasil yang didapat dari data.



**Gambar 1.** Metodologi Penelitian

## 2.5. Analisis Sentimen

Analisis sentimen adalah cabang text mining yang bertujuan untuk memperjelas ulasan ke dalam kelas tertentu. Review bisa diklasifikasikan menjadi kelas positif atau negatif. Dalam skala perusahaan, analisis sentimen tinjauan penjualan dapat digunakan oleh dewan manajemen perusahaan sebagai dasar berbagai proses pengambilan keputusan di perusahaan [13].

## 2.6. Feature Selection

*Feature selection* merupakan salah satu optimasi sebelum digunakannya algoritma sebagai model dalam analisis sentiment [14].

## 2.7. Data Mining

Yuli Mardi mengemukakan data mining ialah proses yang memanfaatkan teknik atau metode tertentu dalam pencarian informasi dari kumpulan data yang besar [15]. Data mining memanfaatkan banyak disiplin ilmu yang menggabungkan teknik *mechine learning*, statistik, pengenalan pola, database, dan visualisasi dalam penyelesaian masalah pengumpulan informasi dari database [16].

## 2.8. Naïve Bayes Classifier (NBC)

NBC adalah algoritma klasifikasi berdasarkan teorema Bayes dengan asumsi independen yang digunakan untuk memprediksi data seakurat mungkin [9]. Dalam menerapkan metode pengklasifikasi *Naive Bayes Classifier*, digunakan persamaan berikut:

$$p(W_i|C_j) = \frac{N_{CW} + 1}{N_C + V} \quad (1)$$

## 2.9. Decission Tree

*Decission tree* merupakan *flowchart* yang berstruktur pohon (*tree*), dimana setiap *node* (simpul internal) mempresentasikan atribut dan cabangnya menandakan hasil tes atau nilai dari atribut, sedangkan daunnya digunakan untuk mempresentasikan kelas atau distibusi dari kelas [17].

## 2.10. Confusion Matrix

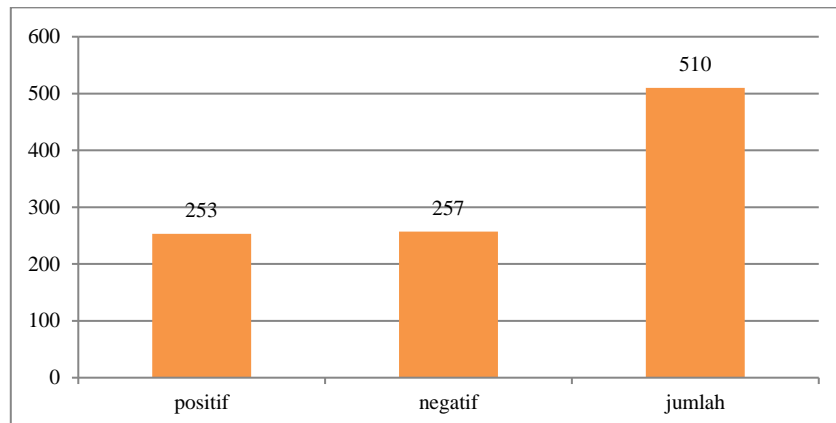
*Confusion matrix* merupakan salah satu alat yang digunakan untuk mengukur kinerja suatu metode klasifikasi dalam mengestimasi objek yang bernilai *true* atau *false*. Pada dasarnya *confusion matrix*

membandingkan kelas input awal yang memuat informasi nilai aktual dan prediksi pada hasil klasifikasi. *Confusion matrix* juga merupakan model pengukuran kinerja klasifikasi berdasarkan pengujian data dan setiap hasil prediksi data dengan proporsi yang tepat [11]. Rumus yang digunakan untuk menghitung akurasi.

$$\text{Akurasi} = \frac{\text{Jumlah data yang benar}}{\text{Total data}} \times 100\% \quad (2)$$

### 3. HASIL DAN ANALISIS

Data training yang diujikan diambil pada Facebook Page Kementerian Kesehatan Republik Indonesia dengan meng-scrape komentar dan didapatkan data sebanyak 510 data diantaranya 253 data ber opini positif dan 257 data ber opini negatif. Selanjutnya data tersebut diolah, di testing dan di training sehingga didapatkan nilai Accuracy, recall dan precision. Jumlah populasi data yang digunakan dapat dilihat pada Gambar 2.



**Gambar 2.** Jumlah Populasi Data

#### 3.1. Analisis Sentimen Positif dan Negatif

Dataset yang digunakan sebagai analisis sentimen positif dan negatif yang diperoleh dapat dilihat pada Tabel 1:

**Tabel 1.** Analisa Sentimen Positif dan Negatif

No	Text	Label
1.	alhamdulillah moga proses vaksinasi laksana disiplin jaga protokol sehat vaksinasi dilaksanakan tuntas	positif
2.	semangat pikir positif doa moga pandemi cepat akhir utama medis korban jasa garda moga allah lindung amin	positif
3.	mantapp jabat lindung ayo jabat monggo pake vaksin duluan rakyat kalah	negatif
4.	ragu vaksin sinovac china padahal china vaksin maka china beli vaksin pfizer biontech dari negara lain kenapa indonesia mau nya alat coba vaksin sinovac jelas udah uji	negatif
5.	dgn varian virus yg muncul apakah vaksin yg beli tangkal atau konsumen vaksin yg lanjut varian virus baru moga tyme mudah musnah virus bumi ini amiin	negatif
6.	perintah usaha maksimal putus rantai virus geram tangan pandemi ini tinggal kita masyarakat mari bantu perintah kita salah jaga sehat	negatif
7.	harap kl vaksin fokus dl presiden jajar biar sehat krn tugas2 yg selesai	positif
8.	alhamdulillah moga bentar covid19 pergi negara cinta aamiin	positif
9.	semangat dokter awat moga sllu dlm lindung allah amiin tip pakai masker stop salin salah tetap semangat berdo a moga demi cepat amiin	positif
...	...	...
10	vaksin blum ijin bpom hrs teliti sah bpom baru bs masyarakat km sdh teliti dokter byk vaksin yg efek nya tdk bgs utk tubuh utk jangka nya sdh byk saksi warga amerik	negatif

#### 3.2. Pre-Processing Data

Pada tahap ini dilakukan pembersihan data, diantaranya ialah dengan cara:

a. *Tokenize*

Yaitu proses pemotongan teks menjadi kata, simbol, karakter atau tanda baca, sehingga menjadi token yang dapat di analisa.

b. *Cleaning*

Tahap untuk mengganti atau membersihkan data maupun karakter yang tidak baku seperti penggunaan tag @ yang dapat mengganggu pengolahan data.

c. *Filtering*

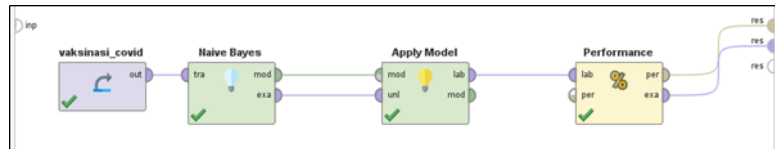
Tahap ini merupakan proses menghapus kata yang tidak mempunyai makna seperti angka, kata ke, yang, di, atau disebut *stop words*.

d. *Stemming*

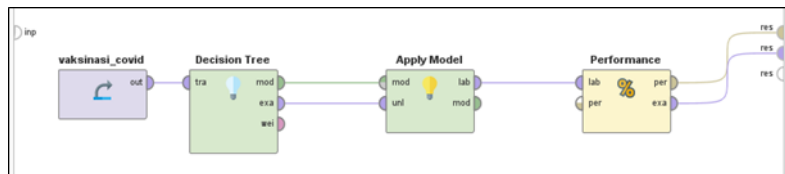
*Stemming* merupakan proses pengecekan ejaan untuk menghitung frekuensi data secara akurat, pemeriksaan ini sangat penting dilakukan untuk mengetahui kalimat yang didapat terdiri dari kata baku atau kata yang tidak keliru dalam pengejaan.

**3.3. Perbandingan Algoritma *Naïve Bayes Classifier* dan *Decision Tree***

Setelah semua tahap pembersihan data selesai, langkah selanjutnya mengolah data menggunakan algoritma *Naïve Bayes Classifier* dan *Decision Tree*. Penulis menggunakan aplikasi Rapid Miner versi 9.8, pada tahap ini penulis menggambarkan rancangan pengolahan data sebagai berikut:



**Gambar 3.** Rancangan Pengolahan Data Menggunakan *Naïve Bayes Classifier*



**Gambar 4.** Rancangan Pengolahan Data Menggunakan *Decision Tree*

Dari rancangan pengolahan data diatas, jika di jalankan maka akan menghasilkan data sebagai berikut:

a. *Naïve Bayes Classifier*

*Confusion matrix* yang didapatkan dari hasil pemodelan pengolahan data menggunakan *Naïve Bayes Classifier* dengan hasil akurasi sebesar 100.00% dapat dilihat pada Tabel 1.

**Tabel 2.** Hasil *Confusion Matrix Naïve Bayes Classifier*

	<i>True Positif</i>	<i>True Negatif</i>	<i>Class Precision</i>
<i>Pred. Positif</i>	253	0	100.00%
<i>Pred. Negatif</i>	0	257	100.00%
<i>Class Recall</i>	100.00%	100.00%	

b. *Decision Tree*

*Confusion Matrix* yang didapatkan dari hasil pemodelan pengolahan data menggunakan *Decision Tree* dengan hasil akurasi sebesar 50.39% dapat dilihat pada Tabel 2.

**Tabel 3.** Hasil *Confusion Matrix Decision Tree*

	<i>True Positif</i>	<i>True Negatif</i>	<i>Class Precision</i>
<i>Pred. Positif</i>	0	0	0.00%
<i>Pred. Negatif</i>	253	257	50.39%
<i>Class Recall</i>	0.00%	100.00%	

**4. KESIMPULAN**

Dari analisis dan hasil yang telah dilakukan dapat disimpulkan bahwa Analisis sentimen opini masyarakat tentang vaksinasi COVID-19 di Indonesia cenderung sedikit lebih banyak tanggapan negatif nya dibanding tanggapan positif. Dari sisi penggunaan algoritma machine learning dapat disimpulkan bahwa *Naïve Bayes Classifier* dan *Decision Tree* dapat digunakan untuk menganalisis sentimen dari data kolom komentar Facebook namun dengan perbandingan akurasi yang jauh berbeda. Analisis sentimen opini publik terhadap vaksinasi COVID-19 di Indonesia yang dilakukan dengan menggunakan *Naïve Bayes Classifier* dengan nilai Akurasi 100.00% dan *Decision Tree* dengan nilai akurasi 50.39%. Pada paper ini dapat diimplementasikan *Naive Bayes Classifier* sebagai algoritma rekomendasi terkait dengan sentimen masyarakat di Indonesia pada Fanspages Kementerian Kesehatan Republik Indonesia.

## REFERENSI

- [1] S. S. Aljameel *et al.*, “A Sentiment Analysis Approach to Predict an Individual’s Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, p. 218, 2020, doi: 10.3390/ijerph18010218.
- [2] F. F. Rachman and S. Pramana, “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter,” *Indones. Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020.
- [3] Ratino, N. Hafidz, S. Anggraeni, and W. Gata, “Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes,” *J. JUPITER*, vol. 12, no. 2, pp. 1–11, 2019.
- [4] Mustakim *et al.*, “DBSCAN algorithm: Twitter text clustering of trend topic pilkada pekanbaru,” *J. Phys. Conf. Ser.*, vol. 1363, no. 1, 2019, doi: 10.1088/1742-6596/1363/1/012001.
- [5] A. Salam, J. Zeniarja, and R. S. U. Khasanah, “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia),” *Pros. SINTAK*, pp. 480–486, 2018.
- [6] A. Rachmat C and Y. Lukito, “Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes,” *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 02, no. 02, pp. 26–34, 2016.
- [7] A. A. Z. A.H. Alamoodi a, B.B. Zaidan, R. Q. M. , O.S. Albahri, K.I. Mohammed, and M. A. E.M. Almahdi, M.A. Chyad, Z. Tareq, A.S. Albahri, Hamsa Hameed, “Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review,” *Expert Syst. Appl.*, vol. 14(4), no. January, pp. 337–339, 2020.
- [8] R. Kurniawan and A. Apriliani, “Analisis Sentimen Masyarakat Terhadap Visur Corona Berdasarkan Opini dari Twitter Berbasis Web Scraper,” *Inform. Sains dan Teknol.*, vol. 5, no. 1, 2020.
- [9] Mustakim *et al.*, “Data Sharing Technique Modeling for Naive Bayes Classifier for Eligibility Classification of Recipient Students in the Smart Indonesia Program,” 2019, doi: 10.1088/1742-6596/1424/1/012009.
- [10] F. Satria, Z. Zamhariri, and M. A. Syaripudin, “Prediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma C4.5 Pada Fakultas Dakwah Dan Ilmu Komunikasi UIN Raden Intan Lampung,” *J. Ilm. Matrik*, vol. 22, no. 1, pp. 28–35, 2020, doi: 10.33557/jurnalmatrik.v22i1.836.
- [11] Mustakim, R. Hastarimasuci, P. Papilo, Zarkasih, Zaitun, and A. Nazir, “Variable Selection to Determine Majors of Student using K-Nearest Neighbor and Naïve Bayes Classifier Algorithm,” *J. Phys. Conf. Ser.*, vol. 1363, no. 1, 2019, doi: 10.1088/1742-6596/1363/1/012057.
- [12] P. S. M. Inzalkar, J. Sharma, S. M. Inzalkar, and J. D. I. E. T. Yavatmal, “A Survey on Text Mining-techniques and application,” *Int. J. Res. Sci. Engineering*.
- [13] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, “Corrigendum: Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 2, p. 169, 2018, doi: 10.20473/jisebi.4.2.169.
- [14] I. Santoso, Windu Gata, and Atik Budi Paryanti, “Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 364–370, 2019, doi: 10.29207/resti.v3i3.1084.
- [15] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, 2017.
- [16] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. 2005.
- [17] Wahyudin, “Metode Iterative Dichotomizer 3 (ID3) untuk Penyeleksian Penerimaan Mahasiswa Baru,” *J. Pendidik. Teknol. Inf. Dan Komun.*, 2009.