



Implementasi Naïve Bayes Dalam Analisis Sentimen Masyarakat pada Twitter Terhadap Flash Sale Shopee

Doni Winarso^{*1}, Ilham Kurniawan²

Email: ¹doniwinarso@umri.ac.id, ²ilhamkurniawan@gmail.com

^{1,2}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau (10pt)

Diterima: 29 Maret 2023 | Direvisi: 18 April 2023 | Disetujui: 28 April 2023

©2023 Program Studi Teknik Informatika Fakultas Ilmu Komputer,

Universitas Muhammadiyah Riau, Indonesia

Abstrak

Twitter sudah sangat umum digunakan untuk membahas berbagai topik dan isu-isu yang sedang terjadi. Salah satu isu yang terjadi adalah kegiatan flash sale yang diadakan oleh salah satu e-commerce yaitu Shopee. Shopee merupakan sebuah marketplace yang membantu para penjual dan pembeli untuk melakukan transaksi penjualan dan pembelian barang dan jasa. Tersebar isu dikalangan masyarakat bahwa terdapat kecurangan yang terjadi pada proses flash sale shopee. Penelitian ini bertujuan untuk mengimplementasikan algoritma Naïve Bayes dalam analisis sentimen masyarakat pada Twitter terhadap flash sale Shopee. Metode pengumpulan data dilakukan dengan melakukan scraping pada akun Twitter menggunakan Twitter API yang terkait dengan flash sale Shopee. Data twiter yang digunakan pada penelitian ini adalah data tweet mulai dari bulan Januari 2021 – Juni 2021 sejumlah 3,467. Setelah melewati tahapan text preprocessing data tersebut menjadi data bersih yang berjumlah 457. Selanjutnya, data diklasifikasikan ke dalam tiga kategori sentimen yaitu positif, negatif, dan netral menggunakan algoritma Naïve Bayes. Hasil analisis menunjukkan bahwa mayoritas sentimen yang diekspresikan oleh masyarakat pada Twitter terhadap flash sale Shopee adalah positif. Kategori sentimen positif sebesar 19.4%, sentimen negatif sebesar 43.1%, dan sentimen netral sebesar 37.4% dengan tingkat akurasi 98%. Diharapkan hasil penelitian ini dapat memberikan insight bagi Shopee dalam meningkatkan kualitas layanan pada masa flash sale sehingga dapat memberikan pengalaman belanja yang lebih baik bagi pelanggan.

Kata kunci: *Analisis Sentimen, Flash Sale, Naïve Bayes.*

Implementation of Naïve Bayes in the Analysis of Public Sentiment on Twitter for Shopee's Flash Sale

Abstract

Twitter has become very common for discussing various topics and issues that are currently happening. One of the issues is the flash sale event held by one of the e-commerce platforms, Shopee. Shopee is a marketplace that helps sellers and buyers to conduct transactions for goods and services. There is an issue circulating among the public that there is cheating happening during the Shopee flash sale process. This research aims to implement the Naïve Bayes algorithm in analyzing public sentiment on Twitter towards the Shopee flash sale event. The data collection method was done by scraping Twitter accounts related to the Shopee flash sale using the Twitter API. The data used in this study are tweets from January 2021 to June 2021, amounting to 3,467. After going through text preprocessing, the data became clean data consisting of 457 tweets. Next, the data was classified into three sentiment categories: positive, negative, and neutral, using the Naïve Bayes algorithm. The analysis shows that the majority of public sentiment expressed on Twitter towards the Shopee flash sale is positive. The positive sentiment category is 19.4%, the negative sentiment category is 43.1%, and the neutral sentiment category is 37.4%, with an accuracy rate of 98%. It is expected that the results of this research can provide insights for Shopee to improve the quality of its services during the flash sale event, thus providing a better shopping experience for its customers.

Keywords: Sentimen analysis, Flash Sale, Naïve Bayes

1. PENDAHULUAN

Twitter merupakan media komunikasi terpopuler di masyarakat saat ini[1]. Menurut website kominfo data dari lembaga Informasi dan Komunikasi Publik (IKP), Indonesia menempati peringkat ke-5 pengguna Twitter terbesar di dunia. Posisi Indonesia berada setelah USA, Brazil, Jepang dan Inggris [2]. Saat ini twitter sudah sangat umum digunakan untuk membahas berbagai topik dan isu-isu yang sedang terjadi. Salah satu isu yang terjadi adalah kegiatan flash sale yang diadakan oleh salah satu e-commerce yaitu Shopee [3]. Shopee merupakan sebuah marketplace yang merupakan bagian dari e-commerce c2c [4] yang membantu para penjual dan pembeli untuk melakukan transaksi penjualan dan pembelian barang dan jasa.

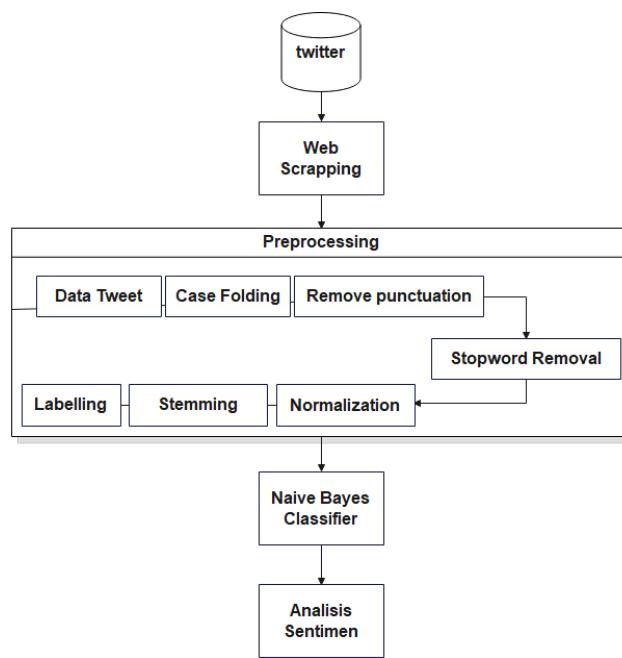
Di kalangan masyarakat tersebut isu bahwa terdapat kecurangan yang terjadi saat kegiatan flash sale itu sedang berlangsung. Hal ini dapat diamati dari beberapa tweet pengguna Twitter, terdapat tweet yang seolah menggambarkan bahwa flash sale Shopee terjadi kecurangan sehingga merugikan banyak pihak. Namun ada beberapa tweet yang mengarah kepada bahwa flash sale yang diadakan oleh pihak Shopee menguntungkan masyarakat. Informasi yang terkandung dalam tweet menjadi penentu untuk mengetahui persentase positif atau negatif kegiatan flash sale itu berjalan dan bisa dilakukan dengan text mining

Opini para pengguna Twitter ini dapat dikenali dengan analisis sentimen. Analisis sentimen atau disebut juga opinion mining yang bertujuan untuk menganalisis, memahami, mengolah, dan menghasilkan data tekstual yang berupa opini terhadap entitas seperti organisasi dan topik tertentu agar dapat mendapatkan suatu informasi [5]. Cara kerja sentimen analisis dilakukan melalui klasifikasi, evaluasi dan visualisasi. Sentimen analisis merupakan topik yang menarik untuk diteliti dan sering dibahas oleh para peneliti [6]. Salah satu metode yang paling sering digunakan untuk klasifikasi pada sentimen analisis adalah naive bayes [7]. Metode ini sering digunakan karena terbukti memiliki tingkat akurasi yang tinggi [8] dan hanya membutuhkan data training yang sedikit [6]. Algoritma ini juga pernah digunakan oleh peneliti [9] untuk melakukan analisis sentimen masyarakat terhadap aplikasi samsat kota Jawa Tengah hasilnya menyatakan bahwa algoritma naïve bayes dapat melakukan klasifikasi dengan baik. Bahkan naïve bayes dapat mengklasifikasi jamur yang beracun atau tidak dengan tingkat akurasi sebesar 100%[10].

Berdasarkan latar belakang tersebut, maka pada penelitian ini akan dilakukan analisis sentimen pada media sosial Twitter terhadap isu Flash Sale Shopee dengan metode Naïve Bayes, yang bertujuan untuk mengelompokkan opini masyarakat ke dalam kategori sentimen positif, negatif, atau netral.

2. METODE PENELITIAN

Tahapan yang dilakukan pada analisis sentimen masyarakat terhadap flash shopee ini adalah dengan cara mengumpulkan data dari tweter untuk kemudian dilakukan text preprocessing sebelum dianalisis menggunakan metode naïve bayes. Tahapan pada penelitian ini dapat dilihat melalui gambar berikut ini.



Gambar 1. Metodologi Penelitian

Langkah yang dilakukan pertama kali pada penelitian ini yaitu identifikasi masalah bagaimana melakukan analisis sentimen pada media sosial Twitter dengan menggunakan metode Naïve Bayes. Pengguna Twitter dapat menuliskan pesan, komentar atau opini apapun di Twitter. Namun cakupan penulisan Twitter sangat terbatas yaitu hanya 280 karakter, sehingga membuat pengguna Twitter menuliskan pesan berupa singkatan. Hal ini tentu menjadi permasalahan sendiri dalam menemukan sentimen pada data Twitter. Untuk itu diperlukan sebuah algoritma yang dapat menyeleksi kata salah dan kemudian merubahnya menjadi kata yang baku sehingga dapat digunakan untuk melakukan analisis sentimen.

2.1 Studi Literatur

Studi Literatur merupakan proses mencari, mempelajari, dan menggunakan berbagai macam literatur berupa buku, jurnal, e-book, atau literatur lain yang berkaitan tentang metode Naïve Bayes khususnya yang digunakan untuk analisis sentimen. Sehingga dapat menjadi acuan dalam melakukan penyusunan pada penelitian ini.

Pengumpulan Data

Tahap ini merupakan tahap pengumpulan dan pencarian dataset twitter Bahasa Indonesia. Data yang dikumpulkan adalah data tweet masyarakat dari Twitter yang merupakan pesan, komentar atau opini dengan topik Flash Shale pada Shopee. Data yang diambil adalah data tweet dari bulan Januari sampai Juni 2021. Data diambil menggunakan proses crawling melalui Twitter API menggunakan library Tweepy dan menggunakan bahasa pemrograman Phyton.

Preprocessing

Data tweet yang telah dikumpulkan dari twitter, banyak memiliki kesalahan sehingga perlu adanya tahapan proses menghilangkan kesalahan tersebut agar proses analisis sentimen menjadi lebih akurat dan dapat digunakan secara general.

Naïve Bayes Classifier

Setelah melakukan preprocessing data, tahap selanjutnya yang dilakukan yaitu klasifikasi menggunakan metode Naïve Bayes. Proses klasifikasi ini dilakukan dengan cara pengecekan kata yang ada pada dataset. Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class [11]. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Naïve Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [10].

Algoritma naïve bayes classifier, pada tiap-tiap kalimat keluhan masyarakat di representasikan dengan pasangan atribut “x1, x2, x3,...xn” dimana x1 adalah kata pertama, x2 adalah kata kedua dan seterusnya, serta V adalah himpunan kategori [12]. Pada saat proses klasifikasi algoritma akan mencari nilai probabilitas yang tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaannya adalah sebagai berikut:

$$C_{MAP} = \operatorname{argmax} p(cj) \cdot \prod_{i=1}^n p(wj|cj) \quad (1)$$

Keterangan:

Cj = Kategori komentar masyarakat

P(wj|cj) = Probabilitas xi pada kategori cj

P(cj) = Probabilitas dari cj

Untuk P(cj) dan P(wi|cj) dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut:

$$P(cj) = \frac{|docsj|}{|contoh|} \quad (2)$$

$$P(wj|cj) = \frac{1+n}{|C|+n(kosakata)} \quad (3)$$

Keterangan:

p(wi |cj) : Probabilitas kata wi pada kategori

cj |docs j|: Jumlah dokumen pada kategori

j |contoh|: Jumlah seluruh dokumen sampel yang digunakan dalam proses training

ni : Frekuensi kemunculan kata wi pada kategori

cj | C |: Jumlah semua kata pada kategori

cj n(kosakata): Jumlah kata yang unik pada semua data training.

Visualisasi

Setelah tahapan klasifikasi data menggunakan metode Naïve Bayes selesai, selanjutnya dilakukan tahap memvisualisasikan hasil dari klasifikasi analisis sentimen dalam bentuk diagram batang maupun wordcloud.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan software Jupyter notebook. Jupyter notebook merupakan lingkungan komputasi interaktif berbasis web untuk membuat dokumen notebook Jupyter. komputasi eksplorasi yang interaktif. Pengumpulan data dilakukan dengan proses scrapping dan didapat data sebanyak 3467 data tweet mengenai flash sale Shopee.

Gambar 2 merupakan sebagian dari data tweet yang diambil menggunakan proses scrapping. Data diambil dari Twitter menggunakan bahasa pemrograman Phyton dan memanfaatkan library Tweepy. Data field yang diambil yaitu username, handle, timestamp, text, emojis, comments, likes, dan retweets.

UserName	Handle	Timestamp	Text	Emojis	Comments	Likes	Retweets
arielvalentino.	@arielvalentino_	20/01/2022 21:56:00	gle manusia shopee kalo flash sale lebih ganas dari ma...		1	null	null
Cikol Goreng	@itzmhyungg	29/01/2022 21:37:00	Auto Glowing 😊 Makasih Flash Sale nya Shopee, sering...			null	null
ibai	@tawakkal	29/01/2022 20:26:00	flash sale shopee iphone se jadi jd1 doang tapi aku tidak...		3	null	null
Anak KetigII	@cheesecake_ku	29/01/2022 17:26:00	Dehlah polonya gaeskan nyesel beli produknya ebony tu...		1	null	null
Yan?	@dianAy_US	29/01/2022 11:56:00	Eiger nanti ikut event shopee 2.2 ya super flashsale 22 r...			null	null
Puyuul	@sumaritas	28/01/2022 08:30:00	Seller shopee sekarang kerbanyaan drama. Ngapain co...		1	null	null
za dm ke @andukijo_ atau WA	@andukijo_	27/01/2022 23:55:00	Mau ngejar flash sale shopee ah wkwk			null	null
enishi	@narenaren	27/01/2022 18:01:00	Klo mau beli item Flashsale shopee, harus di cek lagi, bi...			null	null
Syahrul	@syahrul_jombang	27/01/2022 17:53:00	akun saya lok gak bisa check shopee pay? Berutan saya...		1	null	null
Muhammad s.a.k Mamat	@braddamamat	26/01/2022 19:49:00	Sayang flash sale POCO M3 gak di Shopee, kalo di Shop...			null	null
S.H.	@vivinalvin	26/01/2022 28:59:00	Awas nya gabut aja lalu flash sale shopee . Eh, malah...			null	null
queen of tyo	@sentenjengecil	26/01/2022 28:50:00	flash sale shopee sangat meresahkan			null	null
Oktay Adityanata 5	@oktyedit	26/01/2022 26:05:00	Kegoblokian Hari Ini, ikut Flashsale. Klik ikut cashback ko...		1	null	null
irin DISC 35% BASE -IRENBH- -sulam-	@blusikycora	26/01/2022 10:09:00	Gilee ikut flash sale xomi m3 dapat Pdti kmrn yg poco...		1	null	null
-sulam-	@jenengku_aslam	26/01/2022 09:46:00	Sekian lama ga belanja di Lazada. Kalo buat flashsale rib...			null	null
born2tobe yours	@born2tobe yours8	25/01/2022 20:02:00	Jangan tanya dong kalo shopee udah ada Flash Sale. Pasti...			null	null
Randa	@Randa56063205	25/01/2022 20:00:00	Bersyukur banget aku bisa merasakan Flash Sale nya th...		2	2	
Ayam geprek cabe cabean	@AyamCabean	25/01/2022 20:00:00	Yaampuni Suka bgt dgn kalo shopee lagi ada Flash Sale...		2	null	
Ayam geprek level tabok	@AyamLevel	25/01/2022 20:00:00	Tau banget shopee ini kalo gw udah nungguin lama ban...			null	null
Bakmi Jowo Blasteran	@BlasteranJowo	25/01/2022 20:00:00	Fercaya gue maha kalo shopee utah Flash Sale pasti ram...			null	null
Bilal	@Bilal53494793	25/01/2022 20:00:00	Beberapa banget ada Flash Sale dari Shopee. Makasih sh...			2	
seviant	@seviant	25/01/2022 19:22:00	Semengak jadi buk bukti liat indomie a'famart 2.650 d...			null	null
seviant	@seviant	25/01/2022 19:19:00	Capek deh nunggu brom mobile bener, mau cekout shop...		1	null	null
Astri Hani Shabrina	@astrihabrina	25/01/2022 18:51:00	Buka shopee belanja cemilan, buka tokoed eh ikut netu...			null	null
hahihuhuhoho	@yuhyuhuhoho	25/01/2022 10:36:00	Kris serasa Shopee Flash Sale			3	

Gambar 2. Hasil Scrapping Data

3.2 Tahapan Text Preprocessing

Tahapan text preprocessing dilakukan karena hasil dari scrapping data twitter mengandung banyak noise, seperti duplikasi data dikarenakan data di retweet atau di share kembali oleh pengguna twitter, ataupun mengandung informasi yang tidak berguna untuk proses klasifikasi sentimen seperti hashtag (#), mention (@), link, tanda baca, angka, atau penggunaan bahasa singkatan. Tahapan text preprocessing terdiri dari beberapa tahapan, yaitu:

1. Case Folding

Merupakan tahapan awal dalam melakukan proses preprocessing. Case Folding berfungsi untuk mengubah semua huruf yang ada pada dokumen kedalam bentuk yang sama menjadi lower case atau huruf kecil.

2. Remove Punctuation

Tahap ini merupakan proses menghapus tanda baca dan angka dalam proses analisis sentimen, seperti koma (,), titik (.), 1, 2, dan tanda baca lainnya. Penghapusan tanda baca ini bertujuan untuk mengurangi noise pada data.

3. Normalisasi

Pada tahap ini berfungsi untuk mengubah teks kalimat menjadi teks yang secara lengkap memperlihatkan cara pengucapannya. Meliputi pengubahan singkatan, akronim, angka, tanggal, waktu, karakter khusus, dan simbol-simbol yang menyebabkan ambiguitas.

4. Stopword Removal

Stopword merupakan daftar kata umum yang sering keluar dan di anggap tidak memiliki makna penting seperti waktu, kata penghubung, dan lain sebagainya [5]. Pada tahap ini kata-kata tersebut akan dihapus berdasarkan kata yang telah ditetapkan pada kamus stopword.

5. Stemming

Pada tahap ini akan mengubah kata kedalam bentuk stem (kata dasar) dari kata hasil stopword removal (filtering). Proses stemming yang diterapkan pada penelitian ini menggunakan algoritma Sastrawi. Stemming dapat meningkatkan recall namun dapat menurunkan precision karena berkurangnya informasi dari kata yang telah distem[1].

6. Pelabelan

Tahap ini akan melakukan pelabelan kata dalam kalimat berdasarkan data yang digunakan, sehingga dapat dilakukan pembobotan pada kata. Adapun hasil pada pelabelan ini berupa data sentimen positif, sentimen negatif, ataupun sentimen netral.

3.3 Klasifikasi Naïve Bayes

Pada tahap klasifikasi naïve bayes ini menggunakan metode yang dibagi menjadi dua proses, yaitu proses training dan proses testing. Pada proses TF-IDF sudah dilakukan proses training sebagai pelatihan, dan sudah dilakukan proses testing dengan mengacu probabilitas dari dataset training. Tahap selanjutnya akan memprediksi sentimen ke data frame guna membandingkan data aktual dan data prediksi.

Tahap terakhir adalah membuat confusion matrix prediksi yang dapat dilihat pada gambar 3. Tujuannya untuk mengukur kinerja dari model yang dihasilkan atau untuk melakukan perhitungan akurasi.

		Prediksi			precision	recall	f1-score	support	
		Negatif	Netral	Positif					
Aktual	Negatif	38	0	1	accuracy	0.97	0.97	39	
	Netral	1	43	0					
	Positif	0	0	31					
		negatif	0.97	0.97	0.97	39			
		netral	1.00	0.98	0.99	44			
		positif	0.97	1.00	0.98	31			
		accuracy			0.98	114			
		macro avg	0.98	0.98	0.98	114			
		weighted avg	0.98	0.98	0.98	114			

Gambar 3. Confusion Matrix

Berdasarkan hasil dari perbandungan Vmap dari ketiga sentimen, nilai Vmap sentimen positif lebih besar. Maka data uji terklasifikasi sebagai sentimen positif.

3.4 Visualisasi

Visualisasi dalam bentuk wordcloud menunjukkan kata-kata yang sering muncul dalam topik flash sale shopee. Pada kelas sentimen positif ditunjukkan pada gambar 4, pada kelas sentimen negatif ditunjukkan pada gambar 5, dan pada kelas sentimen netral ditunjukkan pada gambar 6.



Gambar 4. WordCloud Sentimen Positif



Gambar 5. WordCloud Sentimen Negatif



Gambar 6. WordCloud Sentimen Netral

4. KESIMPULAN

Berdasarkan hasil yang telah dilakukan pada penelitian analisis sentimen masyarakat terhadap flash sale shopee di Indonesia menggunakan metode Naïve Bayes, dari 457 data twitter menunjukkan hasil opini masyarakat dengan sentimen kategori positif sebesar 19,4%, opini sentimen netral sebesar 37,4%, dan opini sentimen negatif sebesar 43,1%. Sehingga dapat disimpulkan dari hasil klasifikasi sentimen menunjukkan bahwa masyarakat pada twitter yang merasa kegiatan flash sale terdapat kecurangan adalah nyata masih cukup tinggi dibandingkan dengan masyarakat yang merasa diuntungkan oleh kegiatan flash sale tersebut. Pada penelitian ini tentunya masih terdapat kekurangan. Adapun saran yang dapat diberikan adalah melakukan preprocessing dengan lebih baik lagi seperti menangani permasalahan sarkasme. Memperbanyak kata pada kamus yang digunakan pada saat proses normalisasi. Hal ini bertujuan agar semakin banyak kata yang berhasil dideteksi saat melakukan perhitungan skor sentimen (TF-IDF). Karena semakin banyak kata yang terdapat dalam kamus maka akan semakin bagus ketika melakukan klasifikasi sentimen.

REFERENSI

- [1] N. Ramdhani and R. H. Al-Fadillah, "Analisis Sentimen Pengguna Twitter Terhadap Belajar Daring Selama Pandemi Covid-19 Dengan Deep Learning," *J. Siliwangi*, vol. 7, no. 2, p. 2021, 2021.

[2] Kominfo, "Kominfo : Pengguna Internet di Indonesia 63 Juta Orang," www.kominfo.go.id, 2013.

[3] Unknown, "best-hashtags flashsaleshopee," 2021. <http://best-hashtags.com/hashtag/flashsaleshopee/>

[4] R. Nabilla and C. Dr. Drs. Zaki Baridwan , Ak., M.Si., CA., CPA., "Analisis Kesesuaian Siklus Pendapatan E-Commerce," vol. 1, pp. 1–13.

[5] B. Liu, "Sentiment analysis and subjectivity," *Handb. Nat. Lang. Process. Second Ed.*, pp. 627–666, 2010.

[6] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.

- [7] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Syst.*, vol. 192, no. xxxx, p. 105361, 2020, doi: 10.1016/j.knosys.2019.105361.
- [8] V. A. Permadi, "Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura," *J. Buana Inform.*, vol. 11, no. 2, p. 140, 2020, doi: 10.24002/jbi.v1i2.3769.
- [9] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 234–241, 2022, doi: 10.37859/coscitech.v3i3.4343.
- [10] R. Hayami, Soni, and I. Gunawan, "Klasifikasi Jamur Menggunakan Algoritma Naïve Bayes," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 1, pp. 28–33, 2022, doi: 10.37859/coscitech.v3i1.3685.
- [11] S. Syarli and A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2016.
- [12] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014. doi: 10.1109/ICMIRA.2013.45.