



Implementasi Algoritma Naïve Bayes Dalam Menganalisis Jumlah Live Stream VTuber Skem

Ananda Aulia, Miftah Rusydi Tanjung, Muhammad Iqbal, Rian Farta Wijaya

Fakultas Ilmu Komputer Dan Teknologi Informasi, Universitas Pembangunan Panca Budi Medan, Medan, Indonesia

Email: anandaaulia30@gmail.com, miftahrusdyditj@gmail.com, muhammadiqbal@dosen.pancabudi.ac.id,
rianfartawijaya@dosen.pancabudi.ac.id

Email Penulis Korespondensi: anandaaulia30@gmail.com

Abstrak— *YouTube* merupakan salah satu bukti perkembangan teknologi *digital* terkemuka dalam bidang media dan hiburan[1]. *YouTube* merupakan sebutan bagi mereka yang membuat konten maupun melakukan siaran langsung pada media *YouTube*[2]. Tidak semua *Youtuber* mampu untuk mengekspresikan dirinya secara langsung pada konten yang mereka buat. Untuk itu terciptalah *Virtual YouTuber*[3] yang merupakan bentuk karakter *virtual* 2 dimensi atau 3 dimensi yang dibuat melalui perangkat lunak komputer untuk memudahkan *YouTuber* berinteraksi dengan penonton. Penelitian ini memiliki tujuan untuk menganalisa siaran langsung VTuber HoloH3ro yaitu Kaela Kovalskia[4], Vestia Zeta, dan Kobo Kanaeru yang mengandung skem[5] selama live streaming berlangsung. Dengan jumlah data yang besar untuk setiap waktu live stream yang dilakukan, dibutuhkan algoritma Naïve Bayes melalui penggunaan fitur range dan frekuensi, maka akan diperoleh hasil akurasi yang akan diuji menggunakan Confusion Matri. Dataset yang digunakan diambil dari situs kaggle.com tentang jadwal live stream skem yang dilakukan oleh HoloH3ro.

Kata Kunci : *Youtuber; Virtual Youtuber; Skem; Vtuber; HoloH3ro; Naïve Bayes*

Abstract— *YouTube* is evidence of the development of leading digital technology in the media and entertainment sector. *YouTube* is a term for those who create content or broadcast live on *YouTube* media. Not all *YouTubers* are able to express themselves directly in the content they create. For this reason, *Virtual YouTubers* were created, which are 2-dimensional or 3-dimensional virtual characters created using computer software to make it easier for *YouTubers* to interact with viewers. This research aims to analyze the live broadcasts of VTuber HoloH3ro, namely Kaela Kovalskia, Vestia Zeta, and Kobo Kanaeru, which contain schemes during the live streaming. With a large amount of data for each live stream time, a Naïve Bayes algorithm is needed through the use of range and frequency features, so accuracy results will be obtained which will be tested using the Confusion Matrix. The dataset used was taken from the kaggle.com site regarding the scheme live stream schedule conducted by HoloH3ro.

Keywords: *Youtuber; Virtual Youtuber; Skem; Vtuber; HoloH3ro; Naïve Bayes*

1. PENDAHULUAN

YouTube adalah sebuah situs *web video sharing* (berbagi video) yang populer dimana para pengguna dapat memuat, menonton, dan berbagi klip video secara gratis. Didirikan pada bulan februari 2005 oleh 3 orang mantan karyawan *PayPal*, yaitu Chad Hurley, Steve Chen dan Jawed Karim. Umumnya video-video di *YouTube* adalah video klip film, TV, serta video buatan para penggunanya sendiri. (Tjanatjanita. Widika, 2013)[6]. Dengan hadirnya *YouTube*, hampir semua orang mampu melakukan siaran langsung sendiri. Saat ini, hampir semua orang mampu melakukan siaran langsung melalui media siber seperti *YouTube*, *Twitch*, dan lain-lain. Banyak dari mereka yang bekerja dan menjadikan *livestreaming* atau biasa yang disebut sebagai penyiar siaran langsung sebagai mata pencaharian mereka. Mereka yang menekuni pekerjaan ini biasa disebut dengan *YouTuber*.

Saat ini banyak *YouTuber* yang melakukan *livestreaming*, namun tidak semua *youtuber* ingin menampilkan wajah mereka saat melakukan *livestreaming* atau siaran langsung. Maka dari itu diciptakanlah *Virtual YouTuber* yang merupakan bentuk karakter *virtual* 2 dimensi atau 3 dimensi yang dibuat melalui perangkat lunak komputer untuk memudahkan *YouTuber* berinteraksi dengan penonton. *Virtual Youtuber* (*VTuber*) mulai terkenal pada tahun 2016 di Jepang dan sampai saat ini peminat penonton *vtuber* juga semakin bertambah pesat dan bahkan tidak sedikit orang yang ingin menjadi *vtuber*[7].

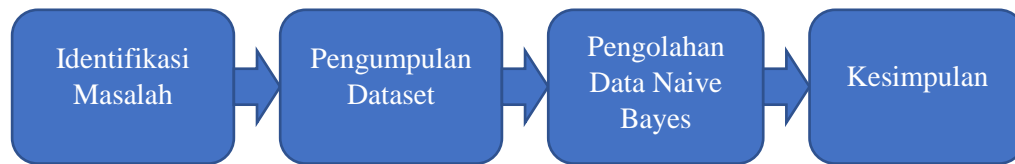
VTuber pertama kali hadir pada tahun 2016 di Jepang dengan mengadopsi karakteristik visual seperti animasi Jepang atau biasa yang disebut anime. Saat ini konten siaran langsung *VTuber* sangat digemari banyak orang, terutama kalangan gen Z. Di Indonesia banyak *VTuber* yang terkenal akan konten yang ditawarkan, seperti Mythia Batford dan Elaine Celestia yang juga dikenal dengan *VTuber Indie* atau *VTuber* yang memulai karirnya sendiri atau mandiri. Selain itu, ada juga *VTuber* agensi yang terkenal di Indonesia bahkan sampai ke dunia Internasional. *VTuber* yang dimaksud berasal dari agensi Hololive Indonesia, Maha Panca, Nijisanji dan masih banyak agensi lainnya. *VTuber*

yang berasal dari agensi biasanya disebut dengan *Talent*. Para *talent* memiliki keunikan masing-masing. Biasanya para *talent* akan menawarkan konten bermain game, menggambar, *talkshow*, bernyanyi, dan lain-lain dalam siaran langsung mereka.

Salah satu dari 3 Group Hololive-ID yaitu HoloH3ro. Ddidalam HoloH3ro terdapat 3 talent yaitu Kaela Kovalskia, Vestia Zeta, dan Kobo Kanaeru. Ketiga talent ini sering melakukan live streaming, namun saat melakukan streaming terdapat perilaku skem yang mereka lakukan. Pada pembahasan kali ini, akan menggunakan metode analisa statistika sederhana yaitu dengan membandingkan jumlah live stream yang dinilai skem berdasarkan batas toleransi waktu tertentu dengan total live stream yang digunakan.

2. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

- Identifikasi Masalah**
Pada tahapan ini dilakukan identifikasi masalah untuk mendefinisikan masalah yang ada juga sebagai acuan data yang akan diolah.
- Pengumpulan Data**
Pengumpulan data dilakukan dengan pencarian dataset yang diambil dari situs kaggle.com
- Pengolahan Data Menggunakan Metode Naïve Bayes**
Pengolahan data yang sudah terkumpul akan diuji dengan metode Naïve Bayes
- Kesimpulan**
Kesimpulan berisi data yang telah selesai diolah dalam penelitian.

2.2 Algoritma

Algoritma adalah urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis". Kata Logis adalah kunci dari algoritma. Langkah-langkah didalam algoritma wajib/harus logis dan dapat ditentukan nilainya benar dan salah. Dalam beberapa konteks algoritma adalah urutan langkah-langkah yang spesifikasi dalam melaksanakan pekerjaan tertentu[8]-[9]-[10]-

Algoritma adalah langkah - langkah dalam menyelesaikan masalah, sedangkan program adalah realisasi dari algoritma dalam bahasa pemrograman. Program ditulis dalam bahasa pemrograman dan kegiatan membuat program disebut pemrograman (programming). Orang/user yang menulis coding/program disebut pemrogram (programmer). Setiap langkah Pada program disebut instruksi. Instruksi yang menjadi penghubung dari kode dan sistem ini yang bertanggung jawab untuk memberikam informasi yang tepat untuk dieksekusi.

Menurut para ahli definisi algoritma adalah yaitu: Algoritma adalah urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis". Kata logis merupakan kata kunci dalam algoritma. Langkah-langkah dalam algoritma harus logis dan harus dapat ditentukan bernilai salah atau benar. Dalam beberapa konteks, algoritma adalah spesifikasi urutan langkah untuk melakukan pekerjaan tertentu[11]-[12]

2.3 Pengertian Naïve Bayes

Bayes adalah klasifikasi menggunakan metode probabilistik dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes Dengan kata lain, memprediksi masa depan berdasarkan data masa lalu Naive Bayes menghitung, untuk setiap kelas keputusan, probabilitas bahwa kelas keputusan itu benar jika diberikan vektor informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen Probabilitas dalam perkiraan akhir dihitung sebagai jumlah frekuensi dari tabel keputusan "master. [13]-[14]-[15]-[16]-[17]

2.4 Persamaan Metode Naive Bayes

Persamaan dari teorema Bayes adalah : [18]

Di mana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Untuk menjelaskan metode Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode Naive Bayes di atas disesuaikan sebagai berikut:

Di mana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \quad (4) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk ($F_1, F_2 \dots F_n$) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C) \quad (6)$$

Persamaan di atas merupakan model dari teorema Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss :

Di mana :

P : Peluang

μ_i : Atribut ke i μ : Nilai atribut ke i σ : Kelas yang dicari μ_j : Sub kelas Y yang dicari σ_j : mean, menyatakan rata - rata dari seluruh atribut σ^2 : Deviasi standar, menyatakan varian dari seluruh atribut.

3. HASIL DAN PEMBAHASAN

3.1 Data Alternatif

Dalam penelitian ini pengambilan data dilakukan dengan mengambil dataset melalui situs kaggle.com. Data yang diambil berjumlah 1204 data live stream, yang dimana output dari dataset yang diambil terdiri dari : channel_id, channel_name, video_id, title, schedule_at, actual_start, diff_time (dalam satuan detik). Dalam pelabelan data

digunakan toleransi 3 tingkat keterlambatan yaitu 1 menit, 3 menit dan 5 menit. Diketahui Jumlah data live stream dengan kategori skem sebagai berikut:

Table 1 Import dataset setelah text preprocessing

	Kobo Kaneru	Kaela Kovalskia	Vestia Zeta
>1 Menit	169	236	119
>3 Menit	85	83	32
>5 Menit	53	41	18
# live stream	335	470	399

Jika kita hitung dalam bentuk persen dibandingkan dengan total live stream dari masing-masing VTuber adalah sebagai berikut:

Table 2. Klasifikasi data dalam bentuk persen (%)

	Kobo Kaneru	Kaela Kovalskia	Vestia Zeta
>1 Menit	50.45%	50.21%	29.82%
>3 Menit	25.37%	17.66%	8.02%
>5 Menit	15.82%	8.72%	4.51%

3.2 Data Labeling

Setelah mengklasifikasi data, proses selanjutnya adalah mencari tahu apakah video live stream tersebut tergolong skem atau tidak. Cara yang dilakukan dengan mengurangi waktu live streaming dengan waktu yang live streaming dijadwalkan. Maka hasil dari proses ini adalah sebagai berikut:

	video_id	title	scheduled_at	actual_start	time_diff	human_diff	is_late_1
channel_name							
Kobo Kanaeru	YCMaEglwcbY	【Eggy Party】 push rank telor	2023-10-31 13:00:00+00:00	2023-10-31 13:06:19+00:00	379	0:06:19	True
Kobo Kanaeru	u4ST1s8IRnA	【Free Talk】 3D IS DONE!!! now what?	2023-10-29 13:15:00+00:00	2023-10-29 13:17:56+00:00	176	0:02:56	True
Kobo Kanaeru	IxQ48VjoC3M	【Post-3D Talk】 SPILLING ALL THE DETAILS	2023-10-27 16:35:00+00:00	2023-10-27 16:41:37+00:00	397	0:06:37	True
Kobo Kanaeru	8xF1xccpc98	【3D SHOWCASE】 LET THIS RAIN SHAMAN SHINE YOUR D...	2023-10-27 13:00:00+00:00	2023-10-27 12:57:23+00:00	-157	-0:02:37	False
Kobo Kanaeru	YYG8IULF04M	【スイカゲーム】 Watermelon Game & 3D Pre-Talk】 3...	2023-10-26 13:30:00+00:00	2023-10-26 13:39:28+00:00	568	0:09:28	True

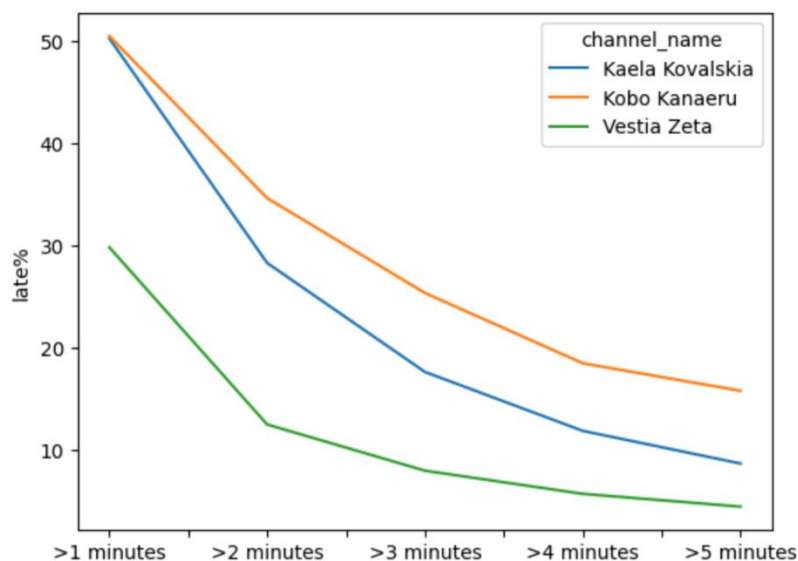
Gambar 2. Perolehan status skem berdasarkan lama waktu live streaming dari yang dijadwalkan dengan mulai stream

Jika mengacu pada dataset yang diolah, perolehan waktu berdasarkan delay live streaming yang didapat sebagai berikut:

channel_name	video_id	title	scheduled_at	actual_start	time_diff	human_diff
Kobo Kanaeru	1LWLFloaec	【Dono Reading + Chit Chat】 hey..	2022-08-14 02:00:00+00:00	2022-08-14 03:13:30+00:00	4410	1:13:30
Vestia Zeta	-uiCXXysa7g	【Baldur's Gate 3】 l'm not a thief	2023-09-21 15:00:00+00:00	2023-09-21 15:50:04+00:00	3004	0:50:04
Kobo Kanaeru	ml5tEnTdIBA	【VALORANT】 mau tamatin battle pass 【Kobo Kanaeru...	2022-06-17 15:10:00+00:00	2022-06-17 15:55:22+00:00	2722	0:45:22
Kobo Kanaeru	PtxeBdd4SBM	【GTA V】 grand theft kobo 【Kobo Kanaeru / Hololiv...	2022-06-18 12:00:00+00:00	2022-06-18 12:35:08+00:00	2108	0:35:08
Kobo Kanaeru	tGnTiolWuvc	【VALORANT】 no screaming 【Kobo Kanaeru / Hololive...	2022-06-03 02:55:02+00:00	2022-06-03 03:29:57+00:00	2095	0:34:55

Gambar 3.Perolehan lama waktu live streaming dari yang dijadwalkan dengan mulai stream

Jika mengacu pada dataset yang diolah, perolehan waktu berdasarkan delay live streaming dalam bentuk grafik yang didapat sebagai berikut:



Gambar 4. Chart perolehan lama waktu live streaming dari yang dijadwalkan dengan mulai stream

4. KESIMPULAN

Berdasarkan penelitian yang disusun dalam jurnal ini, dapat diambil kesimpulan bahwasannya Kobo Kanaeru merupakan VTuber dengan Skem terlama yaitu 1 Jam 13 Menit dan 30 Detik dan mengklaim 4 dari 5 posisi teratas VTuber Skem dalam genmate HoloH3ro lainnya. Dari data yang telah diurai diatas, hal ini sekaligus mengalahkan genmate nya dengan perolehan VTuber HoloH3ro dengan tingkat skem 52.37%, (dengan tingkat toleransi keterlambatan yaitu 3 menit dari jadwal yang ditetapkan), mengalahkan Kaela Kovalskia (17.66%) dan Vestia Zeta (8.02%).



REFERENCES

- [1] Hadeel A. Saed., Ahmad S. Haider., [...], Riyad F. Hussein. "The use of YouTube in developing the speaking skills of Jordanian EFL university students". Heliyon (2021)
- [2] Herryan R. Pratama., Yohanis Franz La Kahija. "APA ARTINYA MENJADI YOUTUBER? ANALISIS FENOMENOLOGIS INTERPRATATIF TENTANG PENGALAMAN MENJADI YOUTUBER". Jurnal Empati (2022)
- [3] Normah., Rifai. B., [...], M. Rifki. "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE". Jurnal Teknik Komputer AMIK BSI (2022)
- [4] A. F. Akhmad., S. S. Sarpudin. (2022). "Menjelajahi Tren Vtuber melalui Case Study Kobo Kanaeru dalam Media Live-Streaming Indonesia = Exploring the Vtuber Rabbit Hole: Looking at the Rise of Kobo Kaaeru in Indonesian Live Streaming Media".
- [5] Zain. I, Zakariya. Z. "Analisis Regresi Logistik Ordinal pada Prestasi Belajar Lulusan Mahasiswa di ITS berbasis SKEM". Jurnal Sains dan Seni ITS (2015)
- [6] Tjanatjantia. Widika. (2013). "Sejarah Berdirinya Youtube _ Sejarah Dunia". Retrieved from <https://canacantya.wordpress.com/sejarah/sejarah-berdirinya-Youtube>
- [7] Faber T. "VTubers and the woman behind the masks", FT.com (2021)
- [8] M. Bobbi, K. Nasution, A. Karim, and S. Esabella, "Sistem Pendukung Keputusan Penilaian Kinerja Ketua Program Studi Menerapkan Metode WASPAS dengan Pembobotan ROC," vol. 4, no. 1, pp. 130–136, 2022, doi: 10.47065/bits.v4i1.1619.
- [9] A. Karim, "Penerapan Algoritma Entropy dan Aras Menentukan Desa Terbaik Di Pemerintah Kabupaten Labuhanbatu," vol. 3, no. 1, pp. 33–43, 2022.
- [10] A. Karim, S. Esabella, T. Andriani, and M. Hidayatullah, "Penerapan Metode Multi-Objective Optimization on the Basis of Simple Ratio Analysis (MOOSRA) dalam Penentuan Lulusan Mahasiswa Terbaik," vol. 4, no. 1, pp. 162–168, 2022, doi: 10.47065/bits.v4i1.1630.
- [11] G. Ginting, S. Alvita, A. Karim, and M. Syahrizal, "Penerapan Complex Proportional Assessment (COPRAS) Dalam Penentuan Kepolisian Sektor Terbaik," vol. 4, no. 2, pp. 616–631, 2020.
- [12] A. Ernawati, "Penerapan Algoritma Entropy Dan Aras Menentukan Penerima Beasiswa Mahasiswa Berprestasi Di Pemerintah Kabupaten Labuhanbatu," vol. 3, no. 2, pp. 74–84, 2022.
- [13] K. Sari and R. P. Cahyono, "ANALISIS MOOD TWITTER MELALUI LAYANAN J & T EXPRESS SELAMA HARBOLNAS 12 / 12 (DENGAN METODE NAIVE BAYES)," vol. 2, no. 11, pp. 1–16, 2022.
- [14] V. No, A. A. Sari, M. K. Anam, and M. Jamaris, "Edumatic : Jurnal Pendidikan Informatika Sistem Prediksi Keuntungan Influencer Pengguna E-Commerce Shopee Affiliates menggunakan Metode Naïve Bayes," vol. 6, no. 2, pp. 394–403, 2022, doi: 10.29408/edumatic.v6i2.6787.
- [15] T. H. Pudjiantoro, F. R. Umbara, and B. Trihatmoko, "Analisis Sentimen Terhadap E-commerce Pada Media Sosial Twitter Menggunakan Metode Naïve bayes," pp. 1–7, 2021.
- [16] J. J. A. Limbong, I. Sembiring, K. D. Hartomo, U. Kristen, S. Wacana, and P. Korespondensi, "ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS," vol. 9, no. 2, pp. 347–356, 2022, doi: 10.25126/jtiik.202294960.
- [17] A. Kepuasan *et al.*, "Analisis kepuasan penggunaan aplikasi shopee menggunakan algoritma naïve bayes," pp. 3–6.
- [18] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," vol. 2, no. 3, pp. 207–217, 2015.