

Analisis Sentimen Tentang Penggunaan Galon Bebas BPA di Indonesia Menggunakan Algoritma Support Vector Machine

Muhammad Iqbal Tri Atmojo, Estu Sinduningrum*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹miqb171@gmail.com, ²*estu.ningrum@uhamka.ac.id

Email Penulis Korespondensi: estu.ningrum@uhamka.ac.id

Submitted: 08/12/2023; Accepted: 28/12/2023; Published: 29/12/2023

Abstrak—Penelitian ini menggunakan algoritma Support Vector Machine untuk mengklasifikasikan sentiment komentar pada platform X, Youtube, dan Tiktok terkait penggunaan galon bebas BPA di Indonesia. Dari 1200 data, setelah proses labeling, diperoleh 772 data dengan 552 data positif dan 220 data negatif. Hasil percobaan menunjukkan bahwa SVM memiliki akurasi sebesar 96.15%, sementara Naïve Bayes memiliki akurasi sebesar 84.55%. Temuan ini mengindikasikan bahwa SVM efektif dalam mengklasifikasikan sentimen dengan tingkat akurasi tinggi, memberikan informasi berharga bagi produsen, pemerintah, dan konsumen terkait penggunaan BPA dalam galon air minum di Indonesia. Penelitian ini berkontribusi pada pemahaman peran sosial media dalam membentuk opini publik dan kebijakan terkait isu-isu lingkungan dan kesehatan.

Kata Kunci: Support Vector Machine; BPA; Naïve Bayes; Analisis Sentimen; Crossvalidation Data

Abstract—This research employs the Support Vector Machine algorithm to classify sentiment in comments on the X, Youtube, and Tiktok platforms regarding the use of BPA-free water gallons in Indonesia. From a total of 1200 data points, post-labeling, 772 data points were obtained, with 552 classified as positive and 220 as negative. The experimental results reveal that SVM achieves an accuracy of 96.15%, while Naïve Bayes achieves an accuracy of 84.55%. These findings indicate that SVM is effective in classifying sentiment with a high accuracy rate, providing valuable insights for manufacturers, government entities, and consumers regarding the use of BPA in water gallons in Indonesia. This study contributes to a better understanding of the role of social media in shaping public opinion and policies related to environmental and health issues.

Keywords: Support Vector Machine; BPA; Naïve Bayes; Sentiment Analysis; Crossvalidation Data

1. PENDAHULUAN

Pertumbuhan teknologi informasi terutama media sosial menyebabkan suatu pergeseran di dalam publik [1]. Beberapa platform media sosial yang sangat populer, seperti X, YouTube, dan Tiktok, berbagai topik hangat sering dibicarakan di sana, seperti ekonomi, politik, budaya, hukum, dan kesehatan. Salah satu perbincangan utama di media sosial adalah tentang penggunaan galon berbahan BPA di Indonesia, yang memicu kontroversi antara pendukung dan penentang, masing-masing memiliki pandangan yang berbeda. Ini menghasilkan perdebatan sengit terutama di platform media sosial, yang mencerminkan perhatian kolektif terhadap masalah penggunaan galon berbahan BPA di Indonesia. Galon berbahan bisphenol-A (BPA) telah menjadi subjek perdebatan global, karena beberapa penelitian ilmiah menghubungkannya dengan potensi dampak negatif pada kesehatan manusia. Pada penelitian yang dilakukan oleh National Toxicology Program U.S. Department of Health and Human Services mengemukakan bahwa paparan bisphenol-a pada dosis yang terpapar ke pada manusia dapat menyebabkan efek yang buruk pada otak, perilaku, dan prostat pada janin, bayi, dan anak-anak[2]. Pada negara Uni Eropa telah memberikan perhatian utama dalam konteks keamanan pangan dan obat-obatan. European Food Safety Authority (EFSA) mengidentifikasi beberapa konsekuensi kesehatan yang dapat terjadi dengan penggunaan bisphenol-a dan mengungkapkan bahwa bisphenol-a dosis tinggi cenderung berdampak buruk pada kesehatan ginjal dan hati. [3]. Dalam konteks di Indonesia, isu ini mendapatkan perhatian serius karena berkaitan dengan keamanan dan kesehatan masyarakat. Masyarakat luas di Indonesia memiliki akses ke berbagai platform media sosial yang memungkinkan mereka untuk berbicara, berbagi, dan mengungkapkan pandangan mereka tentang penggunaan galon berbahan dasar BPA. Penelitian ini bertujuan untuk menggali lebih dalam perasaan dan opini yang tersebar di media sosial Indonesia, khususnya di platform X, Youtube, dan Tiktok, mengenai penggunaan galon berbahan BPA. Analisis sentimen adalah metode yang relevan dalam rangka memahami bagaimana pandangan dan perasaan publik berkembang seiring dengan berjalannya waktu. Terdapat beberapa algoritma yang digunakan untuk melakukan analisis sentimen, seperti Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbors, Decision Trees [4]. Pada penelitian ini, peneliti memilih menggunakan algoritma Support Vector Machine (SVM) karena memiliki beberapa keunggulan yang memudahkan implementasi dan mengurangi kesulitan dalam perhitungan, Support Vector Machine dapat bekerja dengan baik bahkan ketika data terdistribusi dengan tidak merata atau tidak seimbang. Ini membuatnya efektif dalam situasi di mana jumlah sampel dalam kelas tertentu jauh lebih banyak atau lebih sedikit dibandingkan dengan kelas lain. Algoritma support vector machine juga dikenal memiliki kemampuan dalam menangani data besar dengan efisien. support vector machine merupakan salah satu algoritma yang terbukti efektif dalam menganalisis sentimen, SVM hanya bergantung pada sejumlah kecil vektor dukungan dari data pelatihan untuk membuat keputusan klasifikasi. Ini membuatnya lebih efisien dalam penggunaan memori dan penyimpanan model [5]. Dengan menerapkan algoritma ini pada data yang diambil dari platform-platform media sosial yang mencerminkan keragaman pandangan, penelitian ini bertujuan untuk

memberikan wawasan yang lebih mendalam mengenai cara masyarakat Indonesia merespons dan merasakan tentang isu penggunaan galon berbahan BPA. Dengan begitu hasil analisis sentimen ini dapat memberikan informasi berharga kepada pihak berkepentingan, termasuk produsen galon, pemerintah, dan konsumen, untuk membuat keputusan yang lebih baik terkait penggunaan BPA dalam galon air minum. Dengan demikian, penelitian ini akan berkontribusi pada pemahaman yang lebih baik tentang peran sosial media dalam membentuk opini publik dan kebijakan terkait isu-isu lingkungan dan kesehatan.

Penelitian ini juga telah diterapkan sebelumnya oleh Ferdian Syah [6] dengan menerapkan metode SMOTE, algoritma SVM. Penelitian ini digunakan untuk menganalisis sentimen dari 1000 data tweet yang di crawling menggunakan bantuan aplikasi RapidMiner kemudian data tersebut akan dibagi menjadi 2 kelas yaitu kelas positif dan kelas negatif. Hasil dari analisis menggunakan algoritma SVM mencapai tingkat akurasi yang bagus. Tingkat akurasi yang didapat dari menggunakan algoritma support vector machine ialah akurasi sebesar 80% yang menandakan Algoritma Support vector machine dapat menganalisis sentimen yang ada di sosial media twitter. Penelitian ini memberikan manfaat untuk Indihome, sebagai pihak penyedia jasa dan layanan untuk meningkatkan aspek kepuasan terhadap aspek pengguna jasa dan layanan Indihome.

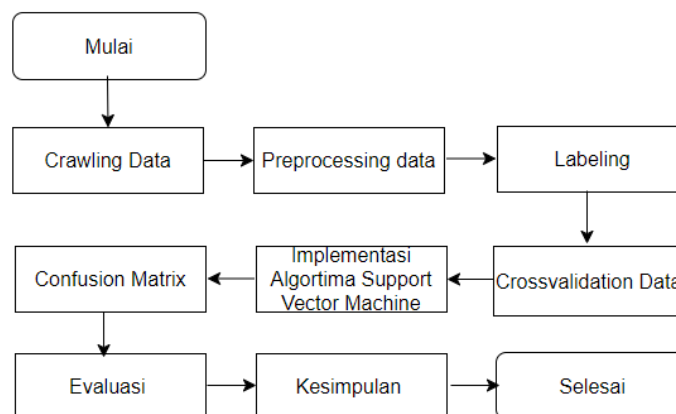
Penelitian ini juga telah diterapkan sebelumnya oleh Riza Fahlapi [7] pada penelitian ini peneliti menggunakan dua algoritma yaitu algoritma Support Vector Machine dan algoritma Naïve Bayes untuk menganalisis sentiment terkait vaksinasi covid-19. Pada tahapan penelitian, data yang didapatkan berjumlah 1013 data komentar dengan diklasifikasikan dua kelas yaitu kelas positif dan kelas negatif. Hasil akurasi yang didapat dari algoritma Support vector machine sebesar 70.51% dan dari algoritma Naïve Bayes sebesar 64.36%. Hasil dari penelitian ini memberikan manfaat untuk menjadi satu alternatif untuk pelaksanaan vaksinasi massal yang dilakukan di Indonesia.

Penelitian ini juga telah diterapkan sebelumnya oleh Muhammad Rafly Al Fattah Zain [8] pada penelitian ini peneliti menggunakan algoritma Naïve Bayes dan menggunakan 2 operator berbeda split data dan crossvalidation data untuk menganalisis sentimen terkait Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana. Pada tahapan penelitian, data yang didapatkan berjumlah 359 data komentar dengan diklasifikasikan dua kelas yaitu kelas positif dan kelas negatif. Hasil akurasi yang didapat dari algoritma Naïve Bayes dan operator splitdata sebesar 86.29%. dan dari algoritma Naïve Bayes dan operator crossvalidation data sebesar 86.23%. Hasil dari penelitian ini memberikan manfaat untuk meningkatkan kualitas produk ubi cilembu Abah Nana.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode yang digunakan dalam penelitian ini meliputi : Crawling Data, Preprocessing Data, Crossvalidation Data, Labeling Data, Implementasi algoritma Support Vector Machine (SVM), serta evaluasi dengan membandingkan hasil antara percobaan pertama dengan menggunakan algoritma Support Vector Machine dan menggunakan proses operator Crossvalidation Data dan percobaan kedua menggunakan algoritma Naïve Bayes dan menggunakan proses operator Crossvalidation Data, serta penyusunan kesimpulan berdasarkan analisis perbandingan kedua percobaan. Secara umum ilustrasi dibawah ini menunjukkan langkah-langkah yang dilakukan oleh peneliti dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.2 Crawling Data

Peneliti menerapkan metode crawling data/web scrapping, Web scrapping merupakan proses mengekstraksi data secara terprogram dari web dan mengubahnya menjadi kumpulan data terstruktur. Hal ini memungkinkan

mengumpulkan secara periode waktu yang lebih singkat, mengumpulkan data secara maksimal dan secara otomatis yang meminimalkan kesalahan [9]. Pada platform X peneliti menggunakan Crawling data x berdasarkan pencarian keyword menggunakan tweet-harvest dengan menggunakan Acces Token unik yang dimana setiap user X berbeda dan data diambil dari platform X. Pada platform Youtube penarikan data menggunakan google spreadsheets dengan bantuan ekstensi apps script, dimana peneliti menggunakan youtube video id untuk pengambilan data secara otomatis. Pada platform Tiktok peneliti dilakukan dengan cara pengambilan komentar secara langsung pada platform tiktok.

2.3 Preprocessing Data

Pada tahap ini data yang sudah dikumpulkan dan disimpan dalam bentuk xlsx selanjutnya akan dilakukan preprocessing, proses ini bertujuan untuk membersihkan teks non-terstruktur sehingga pada proses berikutnya lebih terstruktur [10].

a. **Cleansing**

Pada tahap ini setelah peneliti memperoleh data melalui web scraping, maka data tersebut masih perlu dibersihkan dari simbol, komentar kosong, komentar duplikat, emoji, dan Alamat URL yang perlu dihilangkan karena tidak diperlukan lagi pada saat memproses data [11].

b. **Case Folding**

Setelah terkumpulnya data bersih dari hasil proses sebelumnya, maka perlu ditindak lanjuti dengan penyesuaian huruf kapital menjadi huruf kecil berdasarkan abjad diawali huruf 'a' sampai huruf 'z' [12].

c. **Tokenizing**

Pada tahap ini bertujuan untuk memotong string kalimat berdasarkan tiap kata yang membentuknya. Proses ini juga memecah sekumpulan karakter menjadi satuan kata [13].

d. **Stemming**

Selanjutnya pada tahap Stemming merupakan proses untuk mengembalikan kata menjadi kata dasar yang dihilangkan awalan dan akhirnya, pada tahap ini selama pelatihan dan klasifikasi, stemming dapat secara optimal dapat mengurangi beban memori [14].

e. **Filtering**

Filtering bertujuan untuk membersihkan karakter-karakter yang tidak diinginkan, seperti username, hastag, tautan, maupun tanda baca dan menyaring kata yang diinginkan. Disini peneliti menggunakan Stopword merupakan suatu metode yang bertujuan untuk menghapus kata yang tidak memiliki makna seperti kata penghubung [15].

2.4 Labeling

Data yang telah diolah melalui preprocessing data akan dilabeli secara manual dan akan dilakukan pengklasifikasian ke dalam dua kategori yaitu kategori positif dan negatif berdasarkan sentimen kata dalam komentar [16].

2.5 Crossvalidation Data

Crossvalidation dijelaskan sebagai teknik evaluasi model dengan membagi dataset secara berulang menjadi subset pelatihan dan pengujian. Tujuannya adalah mengukur sejauh mana model dapat menggeneralisasi informasi. Proses melibatkan langkah-langkah pembagian dataset, pelatihan model, dan pengujian. Jenis-jenis crossvalidation seperti k-fold dan leave-one-out perlu dijelaskan, beserta keuntungan seperti menghindari overfitting dan memberikan estimasi performa yang konsisten [17].

2.6 Implementasi Algoritma Support Vector Machine

Support vector machine (SVM) merupakan salah satu dari algoritma yang digunakan untuk klasifikasi dan termasuk dalam kategori supervised learning. Konsep kerja dari Support Vector Machine adalah mencari hyperplane atau garis batas paling optimal yang berfungsi untuk memisahkan kedua kelas [18]. Dalam penerapan algoritma SVM, Terdapat fungsi yang digunakan pada Support Vector Machine yaitu kernel radial basis function (RBF), kernel linier, kernel sigmoid, dan kernel polynomial [19].

2.7 Confusion Matrix

Confusion matrix, yaitu suatu teknik yang umum digunakan untuk melakukan perhitungan akurasi tentang konsep pada data mining. Rumus ini melakukan perhitungan dengan 4 cara, yaitu: recall, precision, accuracy dan error rate [20]. Untuk rumus confusion matrix terdapat pada persamaan rumus 1 [21].

Tabel 1. Confusion Matrix

| Prediction Class | Actual Class | |
|------------------|----------------|----------------|
| | Pred. Positive | Pred. Negative |
| Negative | TP | FN |

| Prediction Class | Actual Class | |
|------------------|----------------|----------------|
| | Pred. Positive | Pred. Negative |
| Positive | FP | TN |

Pengujian algoritma naive bayes dilakukan menggunakan metode confusion matrix berdasarkan pengukuran accuracy, precision dan recall seperti yang diperlihatkan persamaan 2, 3, dan 4.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

Dengan keterangan TP ialah True Positive, TN adalah True Negative, FP adalah False Positive dan FN adalah False Negative [22].

2.8 Evaluasi

Evaluasi adalah langkah untuk memberikan analisis terhadap hasil yang didapatkan dari hasil pengujian yang telah dilakukan. Proses evaluasi ini krusial untuk memahami performa sistem analisis sentimen yang telah dilakukan dalam program [23].

2.9 Kesimpulan

Tahap terakhir yaitu kesimpulan akan menampilkan perbandingan eksperimen yang telah dilakukan dan akan di tampilkan oleh analisis sentimen yang bersifat positif dan negatif dari komentar yang telah dituliskan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Crawling Data

Data yang berhasil diperoleh oleh peneliti sebanyak 34 data dari hasil crawling data X, hasil crawling data di platform Tiktok sebanyak 540 data. Pada platform terakhir yaitu Youtube peneliti mendapatkan sebanyak 626 data, dengan jumlah total hasil crawling data ialah 1200 data. Penelitian ini menggunakan 1 keyword untuk mengcrawling data dari ketiga platform tersebut yaitu “galon BPA”. Hasil crawling data dapat dilihat pada gambar 2 dibawah ini.

| | A | B | C |
|----|------------------------------|-------------------|--|
| | NAMA | Date | Sentimen |
| 1 | marquest | 09/03/23 23:54:36 | apapun itu merk-nya tetap yang menang aqua |
| 2 | arwityaksono77 | 10/03/23 00:48:36 | praktis memang beli air free galon dan seketika banyak ku temui sampah galon. |
| 3 | cepu boys | 10/03/23 03:30:55 | tpi tetep nomer 1 aqua |
| 4 | WARRID_00 | 10/03/23 04:00:50 | pangalah banyu godokan lebih sehat |
| 5 | itapurnamasari8534 | 10/03/23 05:01:34 | aqua gak ada sampah sebaliknya |
| 6 | messL_10 | 10/03/23 06:17:17 | apapun merknya orang2 nyebut ada Aqua... |
| 7 | etianisa88 | 15/04/23 13:26:23 | tetap Aqua |
| 8 | Ubi Cilembu | 15/04/23 13:40:52 | punteun saya tim Aqua |
| 9 | yantidw1715inspirasi Batulah | 15/04/23 13:51:48 | g tau knp aku msh blnya aqua |
| 10 | uat utami | 15/04/23 14:01:04 | tp aqua selalu dihati... |
| 11 | HBaCIBOnE) | 15/04/23 14:48:50 | kalau ke warung cari minum pasti yang pertama di tanya ada aqua |
| 12 | Danjar M | 15/04/23 14:59:48 | asal tau cara daur ulang galon le mineral ntr belnya itu terus |
| 13 | diach02 | 15/04/23 23:35:16 | team aqua sih aq... |
| 14 | Gembul | 16/04/23 02:02:20 | dan aqua tertawa ketika dimana" melihat sampah galon saingannya menumpuk di dalam rumah... |
| 15 | aisah cantik | 16/04/23 03:42:55 | aqua dunk |
| 16 | djatsudrajat | 16/04/23 06:26:49 | tadinya aku langganan Aqua,tp setelah kejadian,didalam galon ada tutup galon isi ulang...mana udah berlumut lagi,aku g pernah bli lagi,kapok |
| 17 | DEN ZAIDAN | 16/04/23 11:12:37 | mantap emang lemineral |
| 18 | Gisela Nathania Asteria | 16/04/23 15:34:14 | Beli galon le minerale gak perlu khaw atir adanya sampah yg menumpuk...krn bekas galon seing diminta ibu2 tetangga utk dibuat pot tanaman. |
| 19 | poput.store | 17/04/23 15:43:50 | Tim lee mineral...aqua kok agak pahit2 ya skrg |

Gambar 2. Hasil Crawling Data

3.2 Preprocessing

Setelah dilakukan crawling data maka tahap selanjutnya dalam penelitian ini ialah preprocessing data , Langkah ini guna untuk mengurangi dan menghapus kata-kata yang tidak relevan dalam data, meningkatkan mutu data, dan menghasilkan sentimen yang lebih efisien dalam dataset. Proses ini mencakup beberapa tahap diantaranya, yaitu : cleansing, case folding, tokenizing, stemming dan filtering. Langkah-langkah ini melibatkan beberapa tahap yaitu:

a. Cleansing

Setelah dilakukan proses crawling data peneliti lanjut memasuki tahap Cleansing data, Proses Cleansing data bertujuan untuk menghilangkan data duplikat, data yang mengandung emoji, dan juga data yang menggunakan simbol atau tanda baca. Pada tahap ini peneliti menggunakan operator replace untuk membersihkan data-data tersebut seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Hasil Sampel Cleansing Data

| Input | Output |
|--|---|
| Padahal keluargaku pengonsumsi air galon | Padahal keluargaku pengonsumsi air galon sedari |

| Input | Output |
|--|--|
| <p>sedari dulu 😞😞,,tapi baru tau skrg efek negatif BPA ini. Next harus lebih hati-hati pilih galonnya..</p> <p>Dr air minum yg tercemar BPA bisa menimbulkan kanker, diabetes sampai penyakit Jantung ini g bisa dianggap sepele sih !!!</p> <p>Galon yg bermerek ada BPA? Kalo uda tahu mengandung BPA kenapa dapat izin beredar. Aktif d produksi Dan d Jual beli n d pakai bebas d pasaran □□</p> <p>wahh ternyata galon yang di pakai selama ini engga baik buat tubuh kita loh bahaya banget untuk masyarakat nih 😞😞😞</p> <p>Ayo ...Kementerian Kesehatan dan BPOM rilis pengumuman galon dan botol plastik perusahaan mana yg msh mengandung BPA 🚫🚫🚫</p> <p>praktis memang beli air yang galonnya bpafree dan seketika banyak ku temui sampah galon dmn mn!!</p> | <p>dulu tapi baru tau skrg efek negative BPA ini Next harus lebih hatihati pilih galonnya</p> <p>Dr air minum yg tercemar BPA bisa menimbulkan kanker diabetes sampai penyakit Jantung ini g bisa dianggap sepele sih</p> <p>Galon yg bermerek ada BPA Kalo uda tahu mengandung BPA kenapa dapat izin beredar Aktif d produksi Dan d Jual beli n d pakai bebas d pasaran</p> <p>wahh ternyata galon yang di pakai selama ini engga baik buat tubuh kita loh bahaya banget untuk masyarakat nih</p> <p>Ayo Kementrian Kesehatan dan BPOM rilis pengumuman galon dan botol plastic Perusahaan mana yg msh mengandung BPA</p> <p>Praktis memang beli air yang galonnya bpafree dan seketika banyak ku temui sampah galon dmn mn</p> |

b. Case Folding

Pada tahap selanjutnya yaitu Case Folding bertujuan untuk merubah suatu kalimat atau kata dengan mengubah setiap kata atau kalimat yang bermula dari huruf kapital menjadi huruf kecil. Hasil dari tahapan Case Folding dapat ditunjukkan pada Tabel 3.

Tabel 3. Hasil Sampel Case Folding

| Input | Output |
|---|---|
| <p>Padahal keluargaku pengonsumsi air galon sedari dulu tapi baru tau skrg efek negative BPA ini Next harus lebih hatihati pilih galonnya</p> <p>Dr air minum yg tercemar BPA bisa menimbulkan kanker diabetes sampai penyakit Jantung ini g bisa dianggap sepele sih</p> <p>Galon yg bermerek ada BPA Kalo uda tahu mengandung BPA kenapa dapat izin beredar Aktif d produksi Dan d Jual beli n d pakai bebas d pasaran wahh ternyata galon yang di pakai selama ini engga baik buat tubuh kita loh bahaya banget untuk masyarakat nih</p> <p>Ayo Kementrian Kesehatan dan BPOM rilis pengumuman galon dan botol plastic Perusahaan mana yg msh mengandung BPA</p> <p>Praktis memang beli air yang galonnya bpafree dan seketika banyak ku temui sampah galon dmn mn</p> | <p>padahal keluargaku pengonsumsi air galon sedari dulu tapi baru tau skrg efek negative bpa ini next harus lebih hatihati pilih galonnya</p> <p>dr air minum yg tercemar bpa bisa menimbulkan kanker diabetes sampai penyakit jantung ini g bisa dianggap sepele sih</p> <p>galon yg bermerek ada bpa kalo uda tahu mengandung bpa kenapa dapat izin beredar aktif d produksi dan d jual beli n d pakai bebas d pasaran wahh ternyata galon yang di pakai selama ini engga baik buat tubuh kita loh bahaya banget untuk masyarakat nih</p> <p>ayo kementrian kesehatan dan bpom rilis pengumuman galon dan botol plastic perusahaan mana yg msh mengandung bpa</p> <p>praktis memang beli air yang galonnya bpafree dan seketika banyak ku temui sampah galon dmn mn</p> |

c. Tokenizing

Tahap Tokenizing bertujuan untuk memecah setiap kata dalam sebuah kalimat.operator yang digunakan adalah Tokenizing dan menggunakan rumus yang terdapat pada web.kaggle.com[24]. Hasil dari tahapan Tokenizing dapat ditunjukkan pada Tabel 4.

Tabel 4. Hasil Sampel Tokenizing

| Input | Output |
|--|---|
| <p>padahal keluargaku pengonsumsi air galon sedari dulu tapi baru tau skrg efek negative bpa ini next harus lebih hatihati pilih galonnya</p> <p>dr air minum yg tercemar bpa bisa menimbulkan kanker diabetes sampai penyakit jantung ini g bisa dianggap sepele sih</p> <p>galon yg bermerek ada bpa kalo uda tahu mengandung bpa kenapa dapat izin beredar aktif d produksi dan d jual beli n d pakai bebas d pasaran</p> | <p>padahal, keluarga, ku pengonsumsi, air, galon, sedari, dulu, tapi, baru, tau, skrg, efek, negative, bpa, ini, next, harus, lebih, hati, hati, pilih, galonnya</p> <p>dr, air, minum, yg, tercemar, bpa, bisa, menimbulkan, kanker, diabetes, sampai, penyakit, jantung, ini, g, bisa, dianggap, sepele, sih</p> <p>galon, yg, bermerek, ada, bpa, kalo, uda, tahu, mengandung, bpa, kenapa, dapat, izin, beredar, aktif, dproduksi, dan, djua,l, beli, n, dpakai, bebas,</p> |

| Input | Output |
|---|---|
| wahh ternyata galon yang di pakai selama ini engga baik buat tubuh kita loh bahaya banget untuk masyarakat nih ayo kementrian kesehatan dan bpom rilis pengumuman galon dan botol plastic perusahaan mana yg msh mengandung bpa praktis memang beli air yang galonnya bpa free dan seketika banyak ku temui sampah galon dmn mn | dpasaran wahh, ternyata, galon, yang, di, pakai, selama, ini, engga, baik, buat, tubuh, kita, loh, bahaya, banget, untuk, masyarakat ayo, kementrian, kesehatan, dan, bpom, rilis, pengumuman, galon, dan, botol, plastic, Perusahaan, mana, yg, msh, mengandung, bpa, praktis, memang, beli, air, yang, galonnya, bpa, free, dan, seketika, banyak, ku, temui, sampah, galon, dmn, mn |

d. Stemming

Proses Stemming adalah untuk pemrosesan bahasa alami yang bertujuan untuk mengubah kata-kata ke bentuk dasarnya dengan menghapus imbuhan, akhiran, dan afiks lainnya. Tujuan utamanya adalah untuk menyederhanakan kata-kata sehingga variasi bentuk yang mungkin dari suatu kata dapat direpresentasikan sebagai bentuk dasar yang sama, seperti yang diilustrasikan dalam Tabel 5.

Tabel 5. Sampel kata untuk Stemming

| Input | Output |
|------------------|----------|
| skrng, skrg | sekarang |
| yg | yang |
| g,ga, gak, nggak | engga |
| d | di |
| dr | dari |
| d,dn | dan |
| udh, ud | udah |
| knp | kenapa |
| msh | masih |
| gnti | ganti |
| aj | aja |
| jg | juga |
| kl, klau | kalua |

e. Filtering

Filtering bertujuan untuk proses penyaringan atau pemilihan berdasarkan karakter-karakter yang tidak diinginkan, seperti username, hastag, tautan, maupun tanda. Filtering diilustrasikan pada Tabel 6.

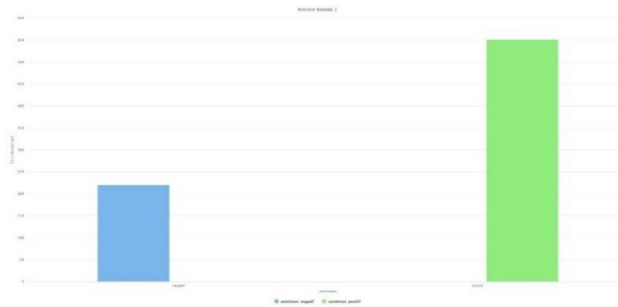
Tabel 6. Hasil Sampel Filtering

| Input | Output |
|---|---|
| padahal, keluarga, ku pengonsumsi, air, galon, sedari, dulu, tapi, baru, tau, skrg, efek, negative, bpa, ini, next, harus, lebih, hati, hati, pilih, galonnya dr, air, minum, yg, tercemar, bpa, bisa, menimbulkan, kanker, diabetes, sampai, penyakit, jantung, ini, g, bisa, dianggap, sepele, sih galon, yg, bermerek, ada, bpa, kalo, uda, tahu, mengandung, bpa, kenapa, dapat, izin, beredar, aktif, d, produksi, dan, d, jua,l beli, n, dpakai, bebas, d, pasaran wahh, ternyata, galon, yang, di, pakai, selama, ini, engga, baik, buat, tubuh, kita, loh, bahaya, banget, untuk, masyarakat, nih ayo, kementrian, kesehatan, dan, bpom, rilis, pengumuman, galon, dan, botol, plastic, perusahaan, mana, yg, msh, mengandung, bpa praktis, memang, beli, air, yang, galonnya, bpa, free, dan, seketika, banyak, ku, temui, sampah, galon, dmn, mn | padahal, keluarga, ku pengonsumsi, air, galon, sedari, dulu, tapi, baru, tau, sekarang, efek, negative, bpa, ini, next, harus, lebih, hati, pilih, galonnya dari, air, minum, yang, tercemar, bpa, bisa, menimbulkan, kanker, diabetes, sampai, penyakit, jantung, ini, engga, bisa, dianggap, sepele, sih galon, yang, bermerek, ada, bpa, kalo, udah, tahu, mengandung, bpa, kenapa, dapat, izin, beredar, aktif, diproduksi, dan, dijual, beli, dan, dipakai, bebas, dipasaran wahh, ternyata, galon, yang, dipakai, selama, ini, engga, baik, buat, tubuh, kita, loh, bahaya, banget, untuk, masyarakat, nih ayo, kementrian, kesehatan, dan, bpom, rilis, pengumuman, galon, dan, botol, plastic, perusahaan, mana, yang, masih, mengandung, bpa praktis, memang, beli, air, yang, galonnya, bpa, free, dan, seketika, banyak, ku, temui, sampah, galon, dimana, mana |

3.3 Labeling

Pada fase ini terdapat 4 tahapan proses klasifikasi data. Dengan labeling data untuk memberi label pada data yang akan di analisis dengan dua kategori yaitu “positif” dan “negatif”, kemudian hasil dari labeling data dilanjutkan keproses crossvalidation data untuk mengatasi masalah variabilitas yang mungkin muncul ketika membagi dataset menjadi data pelatihan (training) dan data pengujian (testing), Penerapan algoritma menggunakan operator Support Vector Machine (SVM) dan tahap akhir Evaluasi dengan menambahkan histogram dan wordcloud sebagai hasil akhir pengujian data. Pada tahap evaluasi peneliti menggunakan dua eksperimen untuk dianalisis dengan menerapkan operator crossvalidation data dan menggunakan dua algoritma berbeda yaitu algoritma Support Vector Machine (SVM) dan algoritma Naïve Bayes.

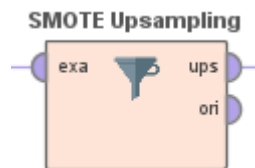
- a. Gunalan huruf kecil dan abjed untuk penomor list. Peneliti menggunakan sampel labeling data, gambar 3 merupakan grafik dari hasil Labaling yang telah dilakukan secara manual dan mengklasifikasikan dalam bentuk kelas postif dan negatif, Pada grafik berwarna biru menunjukkan jumlah dari kelas negatif sedangkan dari grafik hijau menampilkan grafik dari kelas positif. Grafik dapat dilihat pada gambar 3.



Gambar 3. Grafik Labeling

Didapatkan dari jumlah data sebanyak 772 data, dan data yang terdapat pada kelas positif berjumlah 552 data dan data yang terdapat pada kelas negatif berjumlah 220 data.

- b. Proses SMOTE bertujuan untuk memodifikasi kelas-kelas yang tidak seimbang untuk menyeimbangkannya guna mendapatkan lapisan sintetik baru dengan kinerja klasifikasi yang lebih baik dari ini sebelum memasuki proses cross-validation data. Untuk menghindari overfitting atau underfitting, peneliti menggunakan operator SMOTE yang berguna dalam menyeimbangkan dataset. Proses dari SMOTE dapat dilihat pada gambar 4.



Gambar 4. Proses SMOTE

3.4 Wordcloud

Wordcloud adalah representasi visual dari sekumpulan kata, yang mana kata-kata yang lebih sering muncul ditampilkan dalam ukuran lebih besar dan kata-kata yang lebih jarang muncul ditampilkan dalam ukuran lebih kecil. Ini memberikan representasi visual tentang seberapa sering kata muncul dalam teks atau kumpulan data.

- a. Visualisasi wordcloud kelas positif

Visualisasi wordcloud dilakukan guna untuk melihat kata apa yang sering muncul pada sentimen positif yang mendukung penggunaan galon nonBPA, ditunjukkan pada gambar 5. Contoh dari kata yang sering muncul dari kalimat positif ialah:

‘bpafree’, ‘ganti’, ‘pakai’, ‘bpa’, ‘galon’, ‘pet’, ‘mineral’, ‘beralih’, ‘pilih’, ‘anak’, ‘pindah’, ‘mending’, ‘tubuh’, ‘anak’, ‘bpom’, ‘alhamdulillah’



Gambar 5. Tampilan wordcloud sentimen positif

b. Visualisasi wordcloud kelas positif

Visualisasi wordcloud dilakukan guna untuk melihat kata apa yang sering muncul pada ulasan negatif yang menolak penggunaan galon nonBPA, ditunjukkan pada gambar 6. Contoh dari kata yang sering muncul dari kalimat negatif ialah :

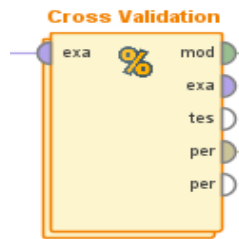
'aqua', 'minum', 'galon', 'air', 'sehat', 'aman', 'aja', 'bahaya', 'freebpa', 'kandungan', 'bpa', 'kemasan', 'mending', 'ulang', 'bpom'



Gambar 6. Tampilan wordcloud sentimen negatif

3.5 Crossvalidation Data

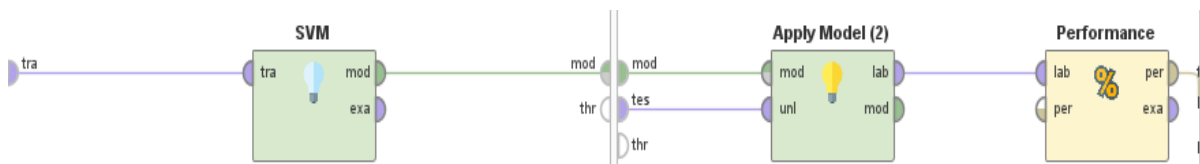
Langkah selanjutnya adalah Crossvalidation Data, Dengan membagi dataset menjadi subset pelatihan dan pengujian secara bergantian. Proses pengaplikasian operator Crossvalidation data dapat dilihat pada gambar 7.



Gambar 7. Proses Crossvalidation

3.6 Implementasi Algoritma Support Vector Machine

Langkah berikutnya melibatkan penerapan algoritma Support Vector Machine dalam tahap klasifikasi. Pada tahap ini, data latih akan menjalani analisis dan diklasifikasi menggunakan algoritma Support Vector Machine. Data uji kemudian akan diaplikasikan dan digunakan kedalam proses pemodelan. Proses pengaplikasian algoritma Support Vector Machine dapat dilihat pada gambar 8.



Gambar 8. Proses Implementasi Algoritma SVM

3.7 Evaluasi

Langkah terakhir dari penelitian ini adalah evaluasi, dimana peneliti membandingkan dua percobaan dengan hasil analisis klasifikasi yang berbeda. Percobaan pertama dilakukan dengan menggunakan algoritma Support Vector Machine dengan operator Crossvalidation data dan Percobaan kedua dilakukan dengan menggunakan algoritma Naïve Bayes dengan operator Crossvalidation data. Berikut hasil dari masing-masing percobaan dengan hasil yang dapat ditampilkan dengan menggunakan tabel Confusion Matrix.

1. Percobaan 1

Percobaan 1 merupakan hasil yang dihasilkan dari analisis dengan menggunakan algoritma Support Vector Machine dan operator Crossvalidation Data. Tabel hasil dari percobaan 1 dapat diperhatikan pada Tabel 7.

Tabel 7. Hasil Percobaan 1

| | True Positive | True Negative | Class Precision |
|----------------|----------------------|----------------------|------------------------|
| Pred. Positive | 436 | 28 | 93.97% |

| | True Positive | True Negative | Class Precision |
|----------------|---------------|---------------|-----------------|
| Pred. Negative | 6 | 414 | 98.57% |
| Class recall | 98.64% | 93.67% | |
| Accuracy | | | 96.15% |

Didapatkan hasil analisis dari percobaan pertama menggunakan operator Crossvalidation Data. Dari percobaan pertama didapatkan accuracy sebesar 96.15%. Untuk kelas positif hasil precision sebesar 93.97%, recall sebesar 93.67% dan dari kelas negatif hasil precision sebesar 98.57%, recall sebesar 93.67%.

2. Percobaan 2

Percobaan 2 merupakan hasil yang dihasilkan dari analisis dengan menggunakan algoritma Naïve Bayes dan operator Crossvalidation Data. Tabel hasil dari percobaan 2 dapat diperhatikan pada Tabel 8.

Tabel 8. Hasil Percobaan 2

| | True Positive | True Negative | Class Precision |
|----------------|---------------|---------------|-----------------|
| Pred. Positive | 423 | 12 | 97.24% |
| Pred. Negative | 129 | 540 | 80.72% |
| Class recall | 76.63% | 97.83% | |
| Accuracy | | | 84.55% |

Didapatkan hasil analisis dari percobaan kedua menggunakan operator Crossvalidation Data. Dari percobaan kedua didapatkan accuracy sebesar 84.55%. Untuk kelas positif hasil precision sebesar 97.24%, recall sebesar 76.63% dan dari kelas negatif hasil precision sebesar 80.72%, recall sebesar 97.83%.

4. KESIMPULAN

Berdasarkan hasil percobaan pada penelitian ini, algoritma Support Vector Machine (SVM) menunjukkan kinerja yang sangat baik dan cocok digunakan dalam analisis sentimen terkait topik penggunaan galon berbahan BPA. Pada percobaan pertama dengan menggunakan operator Crossvalidation Data, SVM mencapai hasil yang sangat memuaskan dengan tingkat akurasi sebesar 96.15%. Selain itu, untuk kelas positif, presisi (precision) mencapai 93.97%, recall sebesar 93.67%, sedangkan untuk kelas negatif, presisi mencapai 98.57%, dan recall sebesar 93.67%. Di sisi lain, pada percobaan kedua dengan menggunakan algoritma Naïve Bayes dan operator Crossvalidation Data, didapatkan hasil akurasi sebesar 84.55%. Untuk kelas positif, presisi mencapai 97.24%, recall sebesar 76.63%, dan untuk kelas negatif, presisi sebesar 80.72% dengan recall mencapai 97.83%. Namun, hasil penelitian secara konsisten menunjukkan bahwa Percobaan 1 memberikan hasil yang lebih optimal dibandingkan dengan Percobaan 2. Temuan ini menegaskan keunggulan SVM dalam mengklasifikasikan sentimen pada data media sosial terkait isu penggunaan galon berbahan BPA. Hasil yang lebih tinggi pada akurasi, presisi, dan recall memberikan indikasi bahwa SVM dapat memberikan kontribusi yang signifikan dalam menganalisis pandangan masyarakat terhadap masalah tersebut. Implikasinya, SVM dapat menjadi pilihan yang lebih unggul dalam pengembangan model analisis sentimen untuk topik sejenis di masa depan. Temuan ini memberikan implikasi signifikan, di mana keunggulan SVM dapat dimanfaatkan oleh pihak-pihak terkait, seperti produsen galon, pemerintah, dan konsumen, dalam menginterpretasikan dan merespons opini publik secara lebih efektif. Dengan kontribusi yang signifikan dalam menganalisis kompleksitas pandangan masyarakat, SVM dapat dianggap sebagai pilihan yang lebih unggul dalam mengembangkan model analisis sentimen untuk topik sejenis di masa depan. Kesimpulan ini memberikan dasar yang kuat bagi pengembangan lebih lanjut dalam penggunaan SVM untuk mendapatkan wawasan yang lebih mendalam dari respons masyarakat terhadap isu-isu kontroversial.

REFERENCES

- [1] R. Budi Syahputra Siregar, L. Rohani, and R. Devianty, "Analisis Penggunaan Media Sosial Instagram Terhadap Komunikasi Pembangunan Di Kota Medan," *SIBATIK J. J. Ilm. Bid. Sos. Ekon. Budaya, Teknol. dan Pendidik.*, vol. 2, no. 3, pp. 1047-1054, 2023, doi: 10.54443/sibatik.v2i3.720.
- [2] E. H. Sciences and NIH, "National Toxicology Program, US Department of Health and Human Services," *Natl. Toxicol. Progr.*, no. 08, pp. 11-13, 2015, [Online]. Available: <http://ntp.niehs.nih.gov/about/0Ahttps://ntp.niehs.nih.gov/ntp/ohat/bisphenol/bisphenol.pdf>
- [3] R. T. Zoeller et al., "European Medicines Agency Conflicts With the European Food Safety Authority (EFSA) on Bisphenol A Regulation," *J. Endocr. Soc.*, vol. 7, no. 9, pp. 1-4, 2023, doi: 10.1210/jendso/bvad107.
- [4] A. P. Ayudhitama and U. Pujiyanto, "Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer," *J. Inform. Polinema*, vol. 6, no. 2, pp. 1-9, 2020, doi: 10.33795/jip.v6i2.274.
- [5] J. A. Saputra and S. A. Aklani, "Analisis Komparasi Algoritma K-Nearest Neighbor Dan Support Vector Machine Dengan Pendekatan Multi Dataset," *J. Ilm. Betrik*, pp. 415-421, 2022, [Online]. Available: <https://ejournal.pppmitpa.or.id/index.php/betrik/article/view/50%0Ahttps://ejournal.pppmitpa.or.id/index.php/betrik/article/download/50/32>
- [6] F. Syah, H. Fajrin, A. N. Afif, M. R. Saeputra, D. Mirrantly, and D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm

- Optimization,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 53–58, 2023, doi: 10.35870/jtik.v7i1.686.
- [7] R. Fahlahi et al., “Analisa Sentimen Vaksinasi Covid-19 Dengan Metode Support Vector Machine Dan Naïve Bayes Berbasis Teknik Smote,” *J. Inform. Kaputama*, vol. 6, no. 1, pp. 57–63, 2022, doi: 10.59697/jik.v6i1.136.
- [8] M. Rafly, A. Fattah, and M. Kamayani, “Analisis Sentimen Ulasan Pelanggan Online Ubi Madu Cilembu Abah Nana Menggunakan Algoritma Naïve Bayes,” *J. Sist. Komput. dan Inform.*, vol. 5, no. September, pp. 11–21, 2023, doi: 10.30865/json.v5i1.6646.
- [9] M. Dogucu and M. Çetinkaya-Rundel, “Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities,” *J. Stat. Educ.*, vol. 0, no. 0, pp. 1–24, 2020, doi: 10.1080/10691898.2020.1787116.
- [10] D. T. Hermanto, A. Setyanto, and E. T. Luthfi, “Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online,” *Creat. Inf. Technol. J.*, vol. 8, no. 1, p. 64, 2021, doi: 10.24076/citec.2021v8i1.264.
- [11] P. Dengan, M. Saw, M. I. Thohir, S. I. Mulyana, and F. Sembiring, “Analisis Sentimen Aplikasi Dompot Digital Pada Google,” vol. 5, no. 2, pp. 202–213, 2023.
- [12] Jimmy, E. H. Hermaliani, and L. Kurniawati, “Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Penundaan Pemilu Presiden Tahun 2024,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 2, pp. 570–579, 2023, doi: 10.35870/jimik.v4i2.243.
- [13] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, “Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.
- [14] A. Andreyestha and Q. N. Azizah, “Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE,” *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 108–116, 2022, doi: 10.29408/jit.v5i1.4581.
- [15] M. I. Amal, E. S. Rahmasita, E. Suryaputra, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter,” *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 645–660, 2022, doi: 10.28932/jutisi.v8i3.5483.
- [16] R. S. Amardita, A. Adiwijaya, and M. D. Purbolaksono, “Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 1, p. 62, 2022, doi: 10.30865/jurikom.v9i1.3793.
- [17] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, “Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [18] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [19] M. Kantardzic, *DATA MINING Concepts, Models, Methods, and Algorithms*, 3rd ed. 2020.
- [20] M. Fansyuri, “Analisa algoritma klasifikasi k-nearest neighbor dalam menentukan nilai akurasi terhadap kepuasan pelanggan (study kasus pt. Trigatra komunikatama),” *Humanika J. Ilmu Sos. Pendidikan, dan Hum.*, vol. 3, no. 1, pp. 29–33, 2020.
- [21] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining Practical Machine Learning Tools and Techniques.*, 4th ed. 2016.
- [22] N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.
- [23] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [24] I. F. PUTRA, “Preprocessing the Indonesian Hate & Abusive Text,” Kaggle, 2019. kaggle.com/code/ilhamfp31/preprocessing-the-indonesian-hate-abusive-text