



Analisis Sentimen Masyarakat Terhadap Pemakaian Galon Sekali Pakai

Bafinatul Umami¹, Erwin Agung Nur Rohmat², Lisda Oktaviana³, Rafael Agintha Tarigan⁴, Robert Kurniawan⁵

^{1, 2, 3, 4}Program Studi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

⁵Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

¹212111944@stis.ac.id

²212112026@stis.ac.id

³212112158@stis.ac.id

⁴212112298@stis.ac.id

⁵robertk@stis.ac.id

Corresponding author email: 212112026@stis.ac.id

Abstract: The innovation of using disposable gallons has caused mixed reactions among the public. The use of recycled PET plastic as product packaging is considered to have the potential to worsen the waste situation in Indonesia. Apart from that, disposable gallons which are considered better than refillable gallons, in fact also contain substances that are dangerous to health. Therefore, this research aims to determine the response of the Indonesian people to the innovation of disposable gallons from environmental and health aspects. In this research, sentiment is divided into 3, namely positive, negative and neutral sentiment. However, this study focused on people's acceptance and rejection of disposable gallons, so neutral sentiment was removed. The classification models used are Naive Bayes, SVM, and KNN. Based on the comparison results of each model, Naive Bayes is the model used in the model because it has the highest accuracy, namely 91.19%. Furthermore, Naive Bayes is able to predict positive and negative sentiment well based on precision, recall and F1-Score values that are close to 1.

Keywords: Naive Bayes, waste, BPA, Ethylene glycol, recycling

Abstrak: Adanya inovasi penggunaan galon sekali pakai menimbulkan reaksi yang beragam di kalangan masyarakat. Penggunaan jenis plastik daur ulang PET sebagai kemasan produk dinilai berpotensi memperburuk kondisi sampah di Indonesia. Selain itu, galon sekali pakai yang dinilai lebih baik dibanding galon isi ulang, nyatanya juga memiliki zat yang berbahaya bagi kesehatan. Oleh karena itu, penelitian ini bertujuan untuk mengetahui tanggapan masyarakat Indonesia akan adanya inovasi galon sekali pakai dari aspek lingkungan dan kesehatan. Dalam penelitian ini, sentimen dibagi menjadi 3 menjadi sentimen positif, negatif, dan netral. Akan tetapi penelitian ini berfokus pada penerimaan dan penolakan orang-orang terhadap galon sekali pakai, sehingga sentimen netral dihapuskan. Model klasifikasi yang digunakan adalah Naive Bayes, SVM, dan KNN. Berdasarkan hasil perbandingan setiap model, Naive Bayes menjadi model yang digunakan dalam model karena memiliki akurasi tertinggi yaitu sebesar 91,19%. Lebih lanjut, Naive Bayes mampu memprediksi sentimen positif maupun negatif dengan baik berdasarkan hasil nilai precision, recall, dan F1-Score yang mendekati 1.

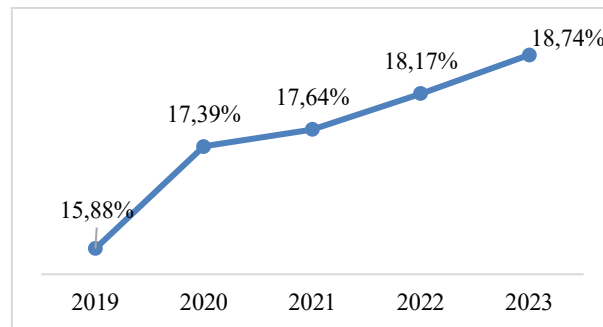
Kata kunci: Naive Bayes, sampah, BPA, Etilen glikol, daur ulang

I. PENDAHULUAN

Plastik menjadi komponen penting yang dibutuhkan manusia dalam kehidupan sehari-hari. Plastik menjadi bahan yang disukai karena memiliki beberapa keunggulan, diantaranya ringan, tahan air, dan tahan lama [1]. Keberadaan plastik ini dapat ditemukan di berbagai fungsi, seperti kantong plastik, kemasan makanan dan minuman, pakaian, peralatan rumah tangga dan perkantoran, serta masih banyak lainnya. Menurut *Making Oceans Pacific Free* (2017), rata-rata penggunaan kantong plastik di Indonesia mencapai 182,7 miliar yang menghasilkan sampah sebesar 1278,9 ton setiap tahunnya. Jika dilihat besarnya kebutuhan pada plastik, dapat dikatakan bahwa plastik tidak dapat dipisahkan dari kehidupan. Penggunaan plastik yang besar ini pada akhirnya akan berdampak pada meningkatnya produksi sampah dari plastik bekas pakai. Menurut Kementerian Lingkungan Hidup dan Kehutanan



(KLHK), plastik merupakan kontributor terbesar kedua dalam produksi sampah di Indonesia pada tahun 2019-2023.



Gambar 1. Kontribusi Plastik terhadap Sampah di Indonesia tahun 2019-2023
Sumber: Kementerian Lingkungan Hidup dan Kehutanan

Gambar 1 menunjukkan kontribusi plastik terhadap sampah di Indonesia terus mengalami peningkatan setiap tahunnya pada tahun 2019 hingga 2023. Pada kurun waktu tersebut, kontribusi sampah plastik tertinggi terjadi pada tahun 2023 dengan nilai sebesar 18,74%, tertinggi setelah jenis sampah sisa makanan. Hal ini menunjukkan bahwa keberadaan sampah plastik terus mengalami peningkatan jika dibandingkan jenis sampah lainnya. Adanya tren yang terus meningkat ini merupakan capaian yang buruk mengingat bahwa plastik merupakan bahan yang sulit untuk terurai, sehingga dapat dipastikan bahwa keberadaan sampah-sampah plastik yang dihasilkan masih ada di lingkungan hingga saat ini.

Pada pertengahan tahun 2020, salah satu perusahaan bidang industri air mineral di Indonesia melakukan inovasi dengan perilisan produk air galon sekali pakai. Produk air galon sekali pakai ini menggunakan kemasan plastik sekali pakai, yaitu jenis *Polyethylene Terephthalate* (PET). Jika dilihat dari sudut pandang keramahan lingkungan, penggunaan PET sebagai kemasan dianggap sebagai inovasi yang berkelanjutan. Hal ini dikarenakan PET merupakan jenis *thermoplastic polymer*, yaitu plastik yang dapat dengan mudah didaur ulang pada temperatur yang tinggi [2]. Namun, pada kenyataannya, KLHK menunjukkan bahwa tingkat pengurangan sampah dan penanganan sampah di Indonesia pada tahun 2023 masing-masing hanya mencapai angka 16,4 dan 50,4 persen secara keseluruhan. Angka ini sangat tidak optimal mengingat target pengurangan sampah dan penanganan sampah nasional masing-masing adalah 30 dan 70 persen pada tahun 2025. Performa ini secara tidak langsung menunjukkan bahwa kemampuan mendaur ulang sampah, salah satunya sampah plastik, di Indonesia masih rendah. Hal ini menimbulkan kekhawatiran bahwa adanya produk galon sekali pakai di saat kemampuan pengelolaan sampah di Indonesia masih tergolong rendah hanya akan menimbulkan tingginya tumpukan sampah plastik pada tahun-tahun mendatang. Seperti yang diketahui, sampah merupakan jenis bahan yang sulit terurai. Karena alasan tersebut, plastik biasanya dimusnahkan dengan cara dibakar. Plastik yang dibakar atau dikubur di tanah dapat menyebabkan terjadinya pencemaran tanah dan air akibat zat kimia beracun [3]. Sampah plastik yang menumpuk dikhawatirkan dapat mencemari lingkungan dan mengganggu keberlanjutan sumber daya alam yang diperlukan untuk produksi pangan. Kemudian, plastik dapat mengganggu ekosistem alam, seperti membahayakan manusia dan binatang [4]. Hal ini dikarenakan sampah plastik yang terbuang, baik di tanah atau di laut, mengganggu habitat hidup hewan yang dapat menyebabkan luka hingga kematian pada hewan. Bahkan, sampah plastik menjadi salah satu makanan dari hewan-hewan yang hidup di daerah tersebut. Hal ini membahayakan eksistensi dari hewan itu sendiri, serta manusia yang mengonsumsi bahan pangan atau olahan hewani karena hewan tersebut



terkontaminasi zat berbahaya akibat plastik. Dengan demikian, sampah plastik ini tidak hanya memberikan dampak buruk bagi lingkungan, tetapi juga pada makhluk hidup yang ada di sekitarnya,

Pada aspek kesehatan, penggunaan galon sekali pakai ini dinilai lebih baik dibandingkan galon isi ulang. Hal ini dikarenakan kemasan galon guna ulang diproduksi menggunakan *Bisphenol A* (BPA) yang mungkin akan terlarut bersama air seiring dengan proses pencucian yang dilakukan terus menerus dan berbahaya bagi kesehatan [5]. Namun, pada kenyataannya zat berbahaya tidak hanya terdapat pada kemasan galon guna ulang. Kandungan zat antimon pada PET menjadi salah satu zat yang bersifat karsinogenik pada manusia. Ketika air tersimpan terlalu lama dalam kemasan PET, ditambah dengan adanya paparan suhu yang tinggi, zat antimon yang terkandung dalam PET akan terlepas dan tercampur bersama partikel air [6]. Pelepasan antimon pada air ini dapat menyebabkan penyakit berbahaya bagi manusia, diantaranya gastritis, dermatitis, dan iritasi pada saluran pernapasan manusia. Kemungkinan tersebut membuat keberadaan galon sekali pakai ini berpeluang untuk membahayakan keamanan pangan. Adanya dampak positif dan negatif dalam inovasi penggunaan plastik jenis PET dalam kemasan galon sekali pakai ini memicu tanggapan yang beragam di kalangan masyarakat. Oleh karena itu, perlu dilakukan analisis mengenai pandangan masyarakat Indonesia akan adanya inovasi produk galon sekali pakai.

Penelitian mengenai tanggapan masyarakat tentang permasalahan yang berkaitan dengan kemasan plastik telah dilakukan sebelumnya. Ganczewski & Jemielniak (2022) melakukan penelitian mengenai tagar ‘*zero waste*’ yang berhubungan dengan material pada kemasan, terutama makanan, di Twitter menggunakan metode analisis sentimen [7]. Analisis sentimen dilakukan pada 124077 *tweet* dengan tagar ‘*zero waste*’ dengan menggunakan *TextBlob script*. Analisis sentimen dilakukan terhadap 5 jenis material kemasan, yaitu plastik bioplastik, kertas, kaca, dan logam. Hasil penelitian menunjukkan bahwa dari seluruh *tweet* yang berhubungan dengan plastik, sebesar 87% *tweet* menunjukkan sentimen yang negatif, 7% netral, dan 6% positif. Hasil ini menunjukkan bahwa masyarakat memandang plastik sebagai permasalahan serius bagi lingkungan. Srivastava et al. (2020) melakukan penelitian serupa untuk mengetahui pandangan masyarakat India terhadap kebijakan larangan penggunaan kantong plastik di negara tersebut [8]. Analisis yang dilakukan adalah analisis sentimen terhadap 31634 *tweet* dari Twitter pada periode 15 Juni 2018 hingga 22 Juni 2018 dengan menggunakan metode *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa 66,99% publik memberi tanggapan positif dan 33,01% publik memberi tanggapan negatif. Hal ini berarti bahwa sebagian besar publik di India mendukung adanya pembatasan dalam penggunaan kantong plastik. Otero et al. (2021) melakukan analisis untuk mengetahui ketertarikan publik di media sosial Twitter terhadap dampak dari polusi laut akibat plastik [9]. Analisis sentimen dilakukan terhadap 140 ribu *tweet* pada periode 19 Maret 2020 hingga 16 November 2020. Hasil analisis menunjukkan bahwa publik lebih menunjukkan tanggapan yang positif akan ketertarikan mereka, serta ditemukan bahwa tagar ‘*plastic*’ dan ‘*plastic pollution*’ mendominasi pembahasan terkait topik ini. Dari beberapa penelitian terkait tersebut, belum ditemukan penelitian yang bertujuan untuk mengetahui sentimen publik terhadap plastik di Indonesia. Selain itu, penelitian yang telah dilakukan hanya merujuk pada sentimen masyarakat terhadap limbah plastik atau plastik secara umum, yang digunakan sebagai kemasan produk atau kantong.

Keterbaruan dari penelitian ini adalah dilakukannya analisis mengenai tanggapan masyarakat terhadap kemasan plastik yang secara spesifik digunakan pada produk galon sekali pakai. Penelitian ini menggunakan data yang diunggah pada kurun waktu 2020 hingga 2023, yaitu saat dimulainya inovasi penggunaan galon sekali pakai di Indonesia. Penelitian ini bertujuan untuk mengetahui tanggapan masyarakat Indonesia akan adanya inovasi produk galon sekali pakai dilihat dari aspek lingkungan dan



kesehatan dengan memanfaatkan data yang diperoleh dari media sosial berupa *tweet*. Kemudian, penelitian ini juga bertujuan untuk mengetahui alasan yang mendasari tanggapan masyarakat tersebut. Selain itu, penelitian ini juga menggunakan beberapa metode untuk memperoleh hasil analisis yang terbaik. Dengan adanya penelitian ini, masyarakat luas diharapkan memiliki gambaran mengenai galon sekali pakai melalui tanggapan dari masyarakat lain yang mungkin sudah mengerti sisi positif dan negatif dari galon sekali pakai. Kemudian, perusahaan yang terkait dengan industri tersebut juga dapat mengetahui tanggapan masyarakat akan inovasi galon sekali pakai, sehingga dapat menjadi referensi mereka dalam membuat inovasi berikutnya. Selain itu, pemerintah juga diharapkan dapat membuat kebijakan yang mampu menjawab tanggapan masyarakat berkaitan dengan galon sekali pakai.

II. METODE PENELITIAN

II.1. Landasan Teori

II.1.1. Text mining

Text mining adalah bagian dari data mining yang mengidentifikasi pola dan informasi yang bermakna dari teks yang tidak terstruktur [10], [11]. Pada umumnya text mining digunakan untuk tujuan eksplorasi, yaitu mendeskripsikan suatu fenomena, serta tujuan eksplanasi dengan melakukan pengujian hipotesis mengenai saling ketergantungan antar objek tertentu. *Text mining* seringkali diterapkan untuk mengetahui pendapat atau tanggapan publik akan suatu topik tertentu di media sosial.

II.1.2. Analisis Sentimen

Analisis sentimen merupakan sebuah metode yang digunakan untuk mengidentifikasi emosi yang terdapat dalam teks yang berkaitan dengan topik tertentu [12]. Teks yang dianalisis akan diklasifikasikan ke dalam emosi positif, negatif, atau netral [13]. Terdapat tiga pendekatan dalam analisis sentimen, yakni pendekatan *machine learning*, pendekatan *dictionary*, dan pendekatan *hybrid*. Dalam pendekatan *machine learning*, terdapat beberapa jenis algoritma yang sering digunakan, seperti *Naïve Bayes*, *Support Vector Machine*, *Maximum Entropy* dan *Word2vec*, dsb. Dalam pendekatan *dictionary* terdapat kamus yang berisi polaritas teks untuk menghitung skor sentimen dan menentukan polaritas dari teks. Lalu, metode *hybrid* merupakan gabungan dari kedua metode tersebut [14].

II.2. Metode

II.2.1. Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini dilakukan melalui *crawling* data X menggunakan bahasa pemrograman Python. Data *tweet* yang telah diambil melalui *crawling* ini, akan disimpan dengan format csv. Proses pengambilan data ini hanya menggunakan kata kunci “galon sekali pakai”, dimana *tweet* yang dikumpulkan adalah *tweet* dari tahun 2020 sampai 2023. *Tweet* ini diambil sebanyak empat kali pengambilan, yaitu diambil per tahunnya. Hasil dari pengambilan data ini diperoleh sebanyak 2270 *tweet*, yang selanjutnya akan dilakukan preprocessing data.

II.2.2. Preprocessing

Berdasarkan data yang diperoleh dari hasil *crawling* pada X, diketahui bahwa data tersebut merupakan data tidak terstruktur sehingga masih terdapat *noise* seperti emoji, tanda baca, *link website*, *username*, kata hubung, dan kata imbuhan. Adanya *noise* tersebut akan mengganggu analisis data serta akan mempengaruhi hasil analisis data. Untuk mengatasi hal tersebut, *noise* akan dihilangkan melalui suatu proses, yaitu *preprocessing*. Namun, sebelum melakukan *preprocessing*, dilakukan penghapusan data yang terduplikasi ataupun *retweet*. *Preprocessing* terbagi menjadi beberapa tahap sebagai berikut.

1. Case Folding



Pada tahap ini, dilakukan perubahan seluruh karakter alfabet (huruf) menjadi huruf kecil (*lowercase*). Tujuan dilakukannya tahap ini yaitu mengurangi adanya variasi kata yang disebabkan adanya perbedaan penggunaan huruf kapital dan kecil.

2. *Cleaning Data*

Pada X, banyak tanda baca yang digunakan dalam membuat *tweet*, misalnya “@” yang menunjukkan username akun X si pembuat *tweet*. Oleh karena itu, tahap ini dilakukan untuk membersihkan karakter-karakter yang tidak digunakan pada analisis data, seperti emoji, *link website*, *username*, tanda pagar, dan tanda baca. Hal ini dilakukan untuk mempermudah analisis text.

3. *Normalized*

Tahap ini merupakan proses mengubah kata tidak baku atau kata gaul pada *tweet* yang ada menjadi kata baku. Pada penelitian ini, pengubahan kata tersebut dilakukan dengan membuat list kata tidak baku dan kata bakunya terlebih dahulu yang kemudian diubah secara serentak melalui *syntax* Python. Selain itu juga dilakukan perbaikan kata-kata yang disingkat melalui excel secara manual.

4. *Stopword*

Penghapusan/penyaringan data (*stopword*) yang tidak berguna merupakan bagian utama dari *preprocessing* data [15]. Pada analisis sentimen, tidak semua kata pada *tweet* digunakan. Oleh karena itu, pada tahap ini dilakukan penghapusan terhadap kata-kata yang dianggap tidak penting atau tidak terlalu berpengaruh.

5. *Tokenized*

Tokenized merupakan salah satu langkah penting dalam melakukan analisis teks. Pada tahap ini, dilakukan pemecahan *tweet* menjadi kata-kata penyusunnya. Pemisahan kata ini didasarkan pada aturan tertentu, seperti spasi, koma, ataupun tanda baca tertentu. Jadi, setiap kata yang ada pada *tweet* akan terpisah ketika terdapat spasi pada antar katanya.

6. *Stemming*

Pada dasarnya, *stemming* adalah suatu operasi untuk mengubah kata sasaran menjadi bentuk dasarnya [15]. Tahapan ini dilakukan untuk mengubah kata pada potongan-potongan *tweet* menjadi kata dasar. Proses pengubahan itu dilakukan menggunakan kamus sastrawi.

Setelah dilakukan *preprocessing*, diperoleh data *clean* (data yang siap untuk dianalisis) sebesar 2105 *tweet*.

II.2.3. Pelabelan Manual

Pelabelan manual merupakan salah satu proses sebelum melakukan analisis data. Pada tahap ini, peneliti melakukan pelabelan pada setiap *tweet* secara manual untuk semua data *tweet* yang tersedia, dimana *labelling* ini terbagi menjadi 3 kategori. Kategori 1 merupakan *tweet* yang berisi sentimen positif. Lalu, Kategori 2 merupakan *tweet* yang berisi sentimen negatif. Terakhir, *tweet* yang tidak berhubungan dengan topik atau dapat dikatakan bahwa *tweet* bersifat netral terhadap pemakaian galon sekali pakai. Dari seluruh data yang digunakan, data akan dibagi menjadi 2 kelompok yaitu data *training* dengan proporsi 80 persen dan data *testing* dengan proporsi 20 persen.

II.2.4. Metode Analisis Data

Terdapat tiga metode Analisis data yang digunakan pada penelitian ini, yaitu *K-Nearest Neighbor* (KNN), *Naïve Bayes*, dan *Support Vector Machine* (SVM). Lalu, akan dilakukan perbandingan hasil



analisis data dari ketiga metode tersebut. Selanjutnya, data yang digunakan pada analisis ini hanya tanggapan yang positif dan negatif, sehingga untuk tanggapan netral dihapus dalam proses analisis data.

1. *K-Nearest Neighbor* (KNN)

Pengklasifikasi *nearest neighbor* didasarkan pada perbandingan *test tuple* dengan *training tuples* yang serupa [16]. Metode ini mengklasifikasikan data berdasarkan tetangga terdekatnya. Prinsip dari KNN dengan menghitung jarak terdekat antara data *testing* dan data *training* [17]. Metode ini sangat simpel, mudah untuk direpresentasikan, mempunyai kemampuan untuk melatih data yang memiliki banyak *noise*, dan efektif untuk proses pengelompokan data.

2. *Naïve Bayes*

Naïve Bayes merupakan salah satu metode klasifikasi yang digunakan dalam analisis sentimen. *Naïve Bayes* adalah kumpulan algoritma klasifikasi yang didasarkan pada Teorema Bayes [15]. Klasifikasi *Naïve Bayes* berdasar pada teorema Bayes dengan asumsi tiap *features* adalah independen [18]. Beberapa kelebihan dari *Naïve Bayes* adalah implementasi yang mudah, *learning process* yang cepat, dan memberikan hasil yang baik [18]. Tetapi, *Naïve Bayes* memiliki kelemahan karena di dunia nyata *features* mungkin dependen. Pengklasifikasian *Naïve Bayes* digunakan sebagai pengklasifikasi probabilistik [15]. *Probability* dari *Naïve Bayes* didefinisikan sebagai berikut:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Berdasarkan persamaan 1, dimana $P(A)$ adalah *prior probability*, $P(B)$ adalah *marginal probability*, $P(A|B)$ merupakan *posterior probability*, dan $P(B|A)$ adalah *likelihood probability* [19].

3. *Support Vector Machine* (SVM)

Support Vector Machine merupakan suatu metode baru yang dapat digunakan untuk klasifikasi data linier ataupun nonlinier [16]. SVM adalah *supervised learning* yang menggunakan *hyperplane* untuk membagi data dengan memaksimalkan *marginal hyperplane* [20]. Pada SVM, dokumen berupa teks tidak cocok untuk analisis model ini [15]. Oleh karena itu, perlu dilakukan perubahan format data menjadi format terstruktur agar format data yang digunakan cocok dengan algoritma SVM.

II.2.5. Evaluasi Model

Untuk mengukur seberapa baik model yang digunakan, perlu dilakukan evaluasi model. Evaluasi ini dilakukan melalui beberapa cara, seperti *confusion matrix*, nilai *accuracy*, *precision*, *recall*, dan *F1 Score*.

1. *Accuracy*

Keakuratan dari pengklasifikasian merupakan persentase tuple yang diklasifikasi dengan benar oleh pengklasifikasi [16]. Akurasi dapat dihitung dengan cara sebagai berikut:

$$\frac{TP+TN}{TP+FP+TN+FN} \quad [21] \quad (2)$$

2. *Precision*

Precision menunjukkan ukuran ketepatan model dalam melakukan prediksi. Nilai ini digunakan untuk melihat persentase *tuple* yang diprediksi bernilai positif, lalu nilai aktualnya juga positif. Rumus dari ukuran ini adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

3. *Recall*



Recall merepresentasikan jumlah item yang diklasifikasikan dengan benar dibagi dengan total positif. *Recall* dapat dihitung dengan rumus berikut:

$$\frac{TP}{TP+FN} \quad [22] \quad (4)$$

4. *F1 Score*

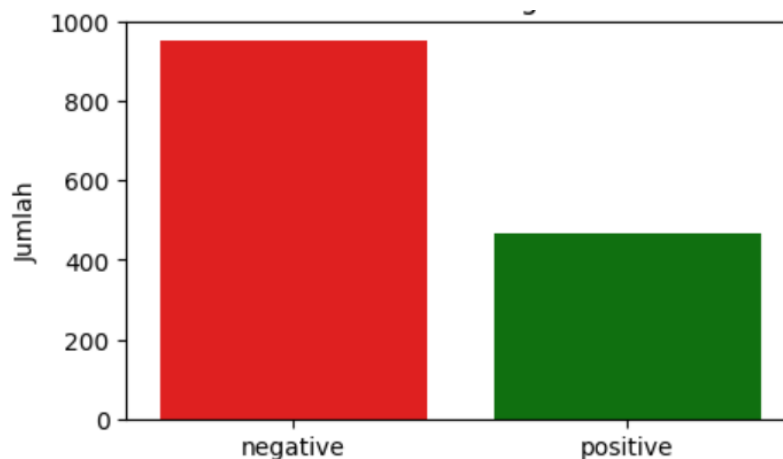
F1 Score merupakan suatu ukuran untuk menilai performa suatu model. Nilai ini diperoleh melalui penghitungan rata-rata harmonik dari *precision* dan *recall*. Nilai ini dapat dihitung melalui rumus berikut:

$$F1 \text{ Score} = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (5)$$

III. HASIL DAN PEMBAHASAN

III.1. Pengumpulan dan Pelabelan Data

Pada penelitian ini menggunakan data teks yang diambil melalui situs X (<https://twitter.com>) yang berisi reaksi publik terhadap adanya galon sekali pakai dengan kata kunci “galon sekali pakai”. Data ini diambil pada rentang waktu 2020 sampai 2023. Pertimbangan rentang waktu yang digunakan untuk mengambil data adalah sejak galon sekali pakai mulai resmi dipakai sebagai wadah untuk air minum. Dari hasil pengumpulan data diperoleh sebanyak 2270 data. Kemudian untuk hasil analisis yang lebih baik dilakukan pembersihan pada data dan mengkategorisasikan sentimen data menjadi sentimen positif, negatif, dan netral. Pembersihan dilakukan sehingga sebanyak 2105 data yang sudah bersih akan dipakai dalam analisis sentimen dengan rincian 954 data dengan sentimen negatif, 466 data dengan sentimen positif, dan sisanya data dengan sentimen netral yang tidak dipakai dalam analisis karena penelitian fokus pada reaksi dan alasan masyarakat dalam menerima atau menolak galon sekali pakai.



Gambar 2. Perbandingan Jumlah *Tweet* Positif dan Negatif

Berdasarkan gambar 2, diketahui bahwa jumlah *tweet* negatif lebih banyak dibandingkan *tweet* positif. *Tweet* positif dan negatif memiliki perbedaan jumlah yang cukup besar, dimana jumlah *tweet* negatif lebih dari dua kali lipat jumlah *tweet* positif. Dengan adanya ketimpangan jumlah data ini, akan mempengaruhi hasil analisis data sehingga harus dilakukan SMOTE.

III.2. Preprocessing

Melakukan *case folding*, *cleaning*, menghilangkan tanda baca dan spasi yang berlebihan, serta tahapan-tahapan *preprocessing* yang telah dijelaskan sebelumnya. Hasil *preprocessing* dapat dilihat pada tabel 1.



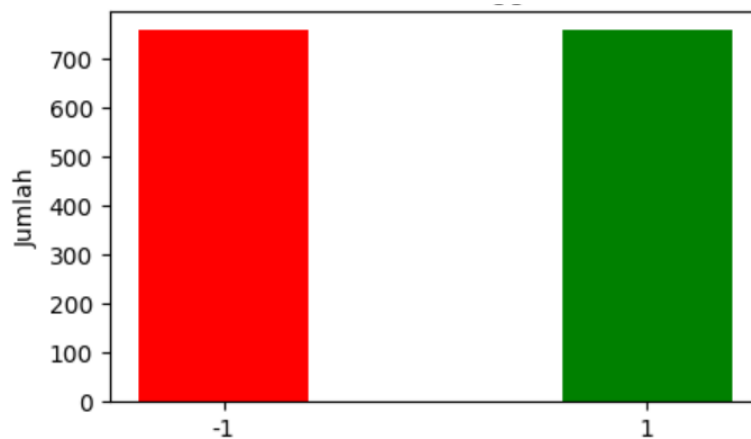
Berdasarkan contoh *tweet* pada tabel 1, tahap awal *preprocessing data* yaitu mengubah seluruh karakter huruf menjadi *lowercase*. Lalu, menghilangkan username (@morgan_indra) beserta karakter-karakter tanda baca lainnya. Selanjutnya, pada tahap *normalized* tidak ada yang berubah dikarenakan seluruh kata yang digunakan sudah termasuk sebagai kata baku. Pada tahap *stopword*, terdapat beberapa kata yang dihapus guna mendukung analisis data, seperti “mereka”, “dan”, dan “yang”. Kemudian, *Tweet* tersebut dipisah menjadi per kata yang selanjutnya diubah menjadi kata dasar, seperti kata “mengklaim” menjadi “klaim”, “menambah” menjadi “tambah”, dan “kesehatan” menjadi “sehat”.

Tabel 1. Contoh Hasil Tahap *Preprocessing*

Proses	Hasil
Data	@morgan_indra lucunya mereka mengklaim galon sekali pakai itu higienis. yang ada galon sekali pakai Le Minerale menambah sampah plastik dan bahaya juga buat kesehatan (Mikro plastik etilen glikol).
Case folding	@morgan_indra lucunya mereka mengklaim galon sekali pakai itu higienis. yang ada galon sekali pakai le minerale menambah sampah plastik dan bahaya juga buat kesehatan (mikro plastik etilen glikol).
Cleaning Data	lucunya mereka mengklaim galon sekali pakai itu higienis yang ada galon sekali pakai le minerale menambah sampah plastik dan bahaya juga buat kesehatan mikro plastik etilen glikol
Normalized	lucunya mereka mengklaim galon sekali pakai itu higienis yang ada galon sekali pakai le minerale menambah sampah plastik dan bahaya juga buat kesehatan mikro plastik etilen glikol
Stopword	lucunya mengklaim galon pakai higienis ada galon pakai le minerale menambah sampah plastik bahaya buat kesehatan mikro plastik etilen glikol
Tokenized	[lucunya, mengklaim, galon, pakai, higienis, ada, galon, pakai, le, minerale, menambah, sampah, plastik, bahaya, buat, kesehatan, mikro, plastik, etilen, glikol]
Stemming	lucu klaim galon pakai higienis ada galon pakai le minerale tambah sampah plastik bahaya buat sehat mikro plastik etilen glikol

III.3. Perbandingan Model

Pada bagian ini, dilakukan perbandingan ketiga model yang digunakan untuk melihat model mana yang terbaik. Namun, sebelum itu, dilakukan *SMOTE* untuk menyeimbangkan jumlah *tweet* pada setiap kategorinya. Hasil penyeimbangan data melalui *SMOTE* dapat dilihat pada gambar 3.



Gambar 3. Jumlah *Tweet* Setelah *SMOTE*



Setelah dilakukan *SMOTE*, *tweet* positif (berkode 1) memiliki jumlah yang sama seperti *tweet* negatif (berkode -1), dimana setiap kategorinya terdiri dari 760 *tweet*. Selanjutnya dilakukan perbandingan pada model *Naïve Bayes*, *K Nearest Neighbor (KNN)*, dan *Support Vector Machine kernel Linear (SVM Linear)*.

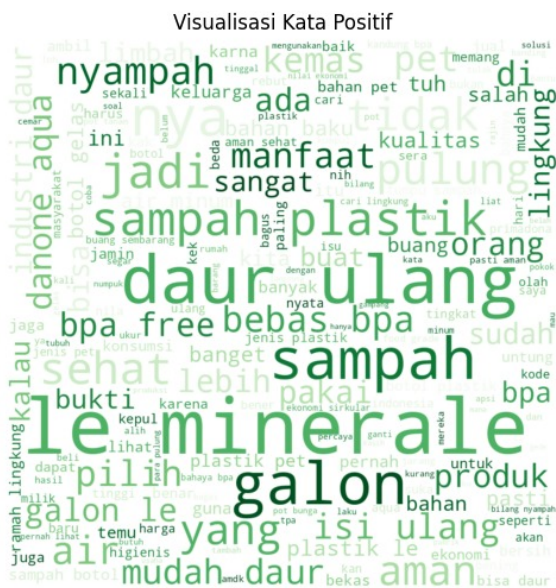
Tabel 2. Perbandingan Akurasi Model

Kriteria	<i>K Nearest Neighbor (KNN)</i>	<i>Naïve Bayes</i>	<i>Support Vector Machine (SVM)</i>
<i>Accuracy</i>	57,39%	91,19%	86,97%

Berdasarkan tabel 2, dapat dibuktikan bahwa model *Naïve Bayes* lebih baik dibandingkan model-model lainnya. Hal ini dilihat dari nilai akurasi model *Naïve bayes* sebesar 91,19%, dimana nilai tersebut lebih besar dibandingkan model KNN dan SVM Linear.

III.4. Analisis Deskriptif

Analisis deskriptif dengan penggunaan visualisasi dalam bentuk *wordcloud* atau kumpulan kata digunakan untuk memunculkan kata-kata yang banyak digunakan pada data *tweet*. *Wordcloud* untuk analisis deskriptif memanfaatkan *package* Python.



Gambar 4a. Visualisasi *Wordcloud* dengan Sentimen Positif



Gambar 4b. Visualisasi *Wordcloud* dengan Sentimen Negatif

Kata-kata pada gambar 4a merupakan kata-kata dengan sentimen positif. Pada gambar 4a, kata-kata yang mendominasi adalah kata “daur”, “ulang”, “bpa”, “free”, “aman”, “sehat”, “pulung”, “plastik”, “pet”. Dari kumpulan kata tersebut dapat disimpulkan bahwa berdasarkan aspek lingkungan dan kesehatan, galon sekali pakai mendapatkan sentimen positif karena galon sekali pakai terbuat dari *Polyethylene terephthalate (PET)* yang merupakan jenis plastik kualitas tinggi yang pembuatannya diklaim bebas dari *Bisphenol A (BPA)* yang sudah sering digunakan untuk pembuatan produk kemasan pangan. Kemudian dalam aspek ekonomi, galon sekali pakai mendapatkan sentimen positif karena galon sekali pakai mudah untuk didaur ulang dibandingkan dengan galon isi ulang, misalnya beberapa orang menggunakan galon sekali pakai untuk membuat pot. Di sisi lain, ternyata galon sekali pakai menjadi salah satu barang yang dicari oleh industri daur ulang dan pemulung sehingga memiliki nilai jual yang lebih tinggi dibandingkan plastik botol ataupun gelas. Oleh karena itu, beberapa orang



menyetujui adanya galon sekali pakai karena setelah dipakai galonnya dapat dijual kepada pemulung atau industri daur ulang atau didaur ulang menjadi barang-barang tertentu.

Kata-kata pada gambar 4b merupakan kata-kata dengan sentimen negatif. Pada gambar 4b kata-kata yang mendominasi adalah kata “sampah plastik, etilen glikol, numpuk sampah, limbah plastik”. Dari kumpulan kata tersebut dapat disimpulkan bahwa disamping galon sekali pakai mendatangkan reaksi positif, galon sekali juga mendatangkan reaksi negatif, terkhususnya dalam hal lingkungan dan kesehatan. Banyak orang mengeluhkan dengan penggunaan galon sekali pakai akan menambah sampah plastik yang sudah menumpuk karena sampah plastik tidak bisa terurai serta beberapa pengguna galon sekali pakai sudah mencoba untuk melakukan daur ulang (misalnya dengan membuat pot) tetapi lama kelamaan barang daur ulangnya menjadi menumpuk. Di sisi lain, banyak orang menyinggung tentang etilen glikol pada galon sekali pakai yang membahayakan kesehatan sehingga mereka tidak setuju dengan adanya galon sekali pakai, bahkan beberapa orang mendesak Badan Pengawas Obat dan Makanan (BPOM) agar mengusut tuntas tentang bahaya dari galon sekali pakai sehingga penggunaan galon sekali pakai tidak lagi diizinkan.

III.5. Analisis Sentimen

Hasil dari perbandingan ketiga model menunjukkan bahwa model *Naïve Bayes* merupakan model terbaiknya, dengan nilai akurasi tertinggi. Tabel 3 merupakan *confusion matrix* yang dihasilkan dari model *Naïve Bayes*.

Tabel 3. *Confusion Matrix*

<i>Actual Value</i>	<i>Predicted Value</i>	
	Positif	Negatif
Positif	183	11
Negatif	14	76

Setelah diklasifikasikan menggunakan model *Naïve Bayes*, diperoleh beberapa hasil diantaranya nilai *True Positive* (TP) sebanyak 183 data. Artinya, terdapat 183 *tweet* yang aktualnya tergolong dalam sentimen positif lalu diprediksi sebagai sentimen positif. Nilai *True Negative* (TN) sebanyak 76 data yang artinya terdapat 76 *tweet* yang diprediksi sentimen negatif, dimana prediksi tersebut sama seperti aktualnya yang merupakan sentimen negatif. Nilai *False Positive* (FP) sebanyak 14 data yang artinya terdapat 14 *tweet* yang diprediksi sentimen positif, namun aktualnya merupakan sentimen negatif. Nilai *False Negative* (FN) sebanyak 11 data yang artinya terdapat 11 *tweet* yang diprediksi sentimen negatif, namun aktualnya merupakan sentimen positif. Hasil dari pengklasifikasian ini akan digunakan sebagai acuan dalam mengevaluasi hasil analisis melalui beberapa ukuran evaluasi.

Tabel 4. Evaluasi Model *Naive Bayes*

Kategori	<i>Naive Bayes</i>		
	<i>Recall</i>	<i>Precision</i>	<i>F1-Score</i>
Negative	0,94	0,93	0,94
Positive	0,84	0,87	0,86
Akurasi	0,91197		

Tabel 4 merupakan hasil evaluasi model yang terdiri dari *recall*, *precision*, *F-1 Score*, dan akurasi dari metode analisis sentimen *Naive Bayes*, baik sentimen positif ataupun sentimen negatif. Nilai *precision* menunjukkan persentase rata-rata makro dari model dalam menduga label kelas pada data dan memperoleh hasil yang sesuai. Nilai akurasi yang semakin mendekati 1 menunjukkan keberhasilan model dalam mengkategorisasikan data dengan benar. Kemudian nilai *recall* yang semakin mendekati 1 menunjukkan keberhasilan model *Naive Bayes* dalam memprediksi label kelas pada data dengan



benar. Terakhir, nilai *F1-Score* yang semakin mendekati 1 menunjukkan model memiliki kinerja yang baik dengan keseimbangan yang baik antara *precision* dan *recall*.

Melalui model *Naive Bayes* diperoleh nilai akurasi, *recall*, *precision*, dan *F1-Score* yang mendekati 1, baik dalam sentimen positif dan negatif. Sehingga dapat disimpulkan dalam model *Naive Bayes* yang diperoleh bahwa model sudah berhasil dalam mengkategorisasikan data dengan benar, memprediksi label kelas pada data dengan benar, dan model memiliki kinerja yang baik dengan keseimbangan yang baik antara *precision* dan *recall*.

Nilai evaluasi yang mendekati 1 berhubungan dengan sudah dilakukannya *SMOTE* pada data karena memiliki ketimpangan pada banyaknya data yang memiliki sentimen negatif. Masalah ketimpangan pada sentimen dapat diatasi dengan melakukan *SMOTE* sehingga data yang digunakan sudah memiliki jumlah yang setara sehingga tidak menimbulkan bias dalam analisisnya.

IV. KESIMPULAN

Analisis sentimen pada pemakaian galon sekali pakai memiliki sentimen negatif yang lebih banyak dibandingkan sentimen positif, sehingga terjadi ketimpangan antara sentimen negatif dan positif. Oleh karena itu, perlu dilakukan *SMOTE* untuk menyeimbangkan jumlah sentimen sehingga tidak menimbulkan bias jika dilakukan analisis lebih lanjut. Analisis sentimen menggunakan 3 model klasifikasi, yaitu KNN, *Naive Bayes*, dan SVM. Kemudian, dihasilkan model terbaiknya *Naive Bayes* dengan tingkat akurasi sebesar 91,16%. Lalu, pada sentimen positif, kata yang memiliki frekuensi terbanyak, yaitu “daur ulang” yang berhubungan dengan galon sekali pakai yang lebih mudah didaur ulang menjadi barang baru dibandingkan dengan galon isi ulang. Sedangkan, pada sentimen negatif, kata yang memiliki frekuensi kemunculan terbanyak, yaitu “sampah plastik” yang berhubungan dengan pemakaian galon sekali pakai yang tidak dapat dipakai berulang kali sehingga hanya akan menimbun sampah plastik yang sudah semakin banyak dan akan membahayakan lingkungan karena plastik tidak akan bisa terurai secara alami.

Dengan mengetahui tanggapan masyarakat, diharapkan pemerintah dapat membuat kebijakan yang mampu menjawab keresahan masyarakat akan adanya inovasi galon sekali pakai. Selain itu, perusahaan yang terkait dengan industri tersebut juga dapat membuat keputusan atau inovasi yang lebih berkelanjutan. Kemudian, saran bagi penelitian selanjutnya dapat mempertimbangkan kemungkinan adanya *buzzer* sehingga diperoleh data yang lebih mencerminkan pendapat masyarakat yang sebenarnya.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada seluruh pihak yang telah membantu dalam penulisan paper ini.

REFERENSI

1. O. Smith and A. Brisman, “Plastic Waste and the Environmental Crisis Industry,” *Crit. Criminol.*, vol. 29, no. 2, pp. 289–309, 2021, doi: 10.1007/s10612-021-09562-4.
2. R. Nisticò, “Polyethylene terephthalate (PET) in the packaging industry,” *Polym. Test.*, vol. 90, no. July, 2020, doi: 10.1016/j.polymertesting.2020.106707.
3. F. Khan, W. Ahmed, and A. Najmi, “Understanding consumers’ behavior intentions towards dealing with the plastic waste: Perspective of a developing country,” *Resour. Conserv. Recycl.*, vol. 142, no. September 2018, pp. 49–58, 2019, doi: 10.1016/j.resconrec.2018.11.020.
4. M. Mukheed and A. Khan, “Plastic Pollution in Pakistan: Environmental and Health Implications,” *J Pollut Eff Cont*, vol. 8, no. 4, p. 251, 2020, doi: 10.35248/2375-4397.20.8.251.
5. A. Okunola A, O. Kehinde I, A. Oluwaseun, and A. Olufiropo E, “Public and Environmental Health Effects of Plastic Wastes Disposal: A Review,” *J. Toxicol. Risk Assess.*, vol. 5, no. 2, 2019, doi:



- 10.23937/2572-4061.1510021.
6. V. Dhaka *et al.*, “Occurrence, toxicity and remediation of polyethylene terephthalate plastics. A review,” *Environ. Chem. Lett.*, vol. 20, no. 3, pp. 1777–1800, 2022, doi: 10.1007/s10311-021-01384-8.
7. G. Ganczewski and D. Jemielniak, “Twitter is garbage: A Thick Big Data exploration of #zerowaste hashtag on Twitter in relation to packaging and food packaging materials,” *Packag. Technol. Sci.*, vol. 35, no. 12, pp. 893–902, 2022, doi: 10.1002/pts.2685.
8. S. Srivastava, J. P. Singh, and D. Mangal, “Time and Domain Specific Twitter Data Opinion,” no. Icimia, pp. 755–761, 2020.
9. P. Otero, J. Gago, and P. Quintas, “Twitter data analysis to assess the interest of citizens on the impact of marine plastic pollution,” *Mar. Pollut. Bull.*, vol. 170, no. July, p. 112620, 2021, doi: 10.1016/j.marpolbul.2021.112620.
10. V. B. Kobayashi, S. T. Mol, H. A. Berkers, G. Kismihók, and D. N. Den Hartog, *Text Mining in Organizational Research*, vol. 21, no. 3. 2018. doi: 10.1177/1094428117722619.
11. J. G. Shim, K. H. Ryu, S. H. Lee, E. A. Cho, Y. J. Lee, and J. H. Ahn, “Text mining approaches to analyze public sentiment changes regarding covid-19 vaccines on social media in korea,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 12, 2021, doi: 10.3390/ijerph18126549.
12. A. Reyes-Menendez, J. R. Saura, and C. Alvarez-Alonso, “Understanding #worldenvironmentday user opinions in twitter: A topic-based sentiment analysis approach,” *Int. J. Environ. Res. Public Health*, vol. 15, no. 11, 2018, doi: 10.3390/ijerph15112537.
13. X. Cao, P. Macnaughton, Z. Deng, J. Yin, X. Zhang, and J. G. Allen, “Using twitter to better understand the spatiotemporal patterns of public sentiment: A case study in Massachusetts, USA,” *Int. J. Environ. Res. Public Health*, vol. 15, no. 2, 2018, doi: 10.3390/ijerph15020250.
14. Z. Li, R. Li, and G. Jin, “Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary,” *IEEE Access*, vol. 8, pp. 75073–75084, 2020, doi: 10.1109/ACCESS.2020.2986582.
15. A. M. Rahat, A. Kahir, and A. K. M. Masum, “Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset,” *Proc. 2019 8th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2019*, pp. 266–270, 2020, doi: 10.1109/SMART46866.2019.9117512.
16. J. Han and M. Kamber, “Classification and Prediction,” in *Data Mining: Concepts and Techniques*, Second Edi. San Fransisco: Morgan Kaufmann, 2006, pp. 337–344.
17. A. Lubis, I. Iskandar, and M. L. W Panjaitan, “Implementation of KNN Methods And GLCM Extraction For Classification Of Road Damage Level,” *IAIC Trans. Sustain. Digit. Innov.*, vol. 4, no. 1, pp. 1–7, 2022, doi: 10.34306/itsdi.v4i1.564.
18. V. Malik and A. Kumar, “Sentiment Analysis of Twitter Data Using Naive Bayes Algorithm,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 6, no. 4, pp. 120–125, 2018, [Online]. Available: <http://www.ijritcc.org>
19. C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, “Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes,” *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.
20. C. A. Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, “The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm,” *Procedia Comput. Sci.*, vol. 234, pp. 156–163, 2024, doi: 10.1016/j.procs.2024.02.162.
21. S. Zhao, “Thumb Up or Down? A Text-Mining Approach of Understanding Consumers through Reviews,” *Decis. Sci.*, vol. 52, no. 3, pp. 699–719, 2021, doi: 10.1111/deci.12349.
22. H. AL-Saif and H. Al-Dossari, “Detecting and classifying crimes from arabic twitter posts using text mining techniques,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 10, pp. 377–387, 2018, doi: 10.14569/IJACSA.2018.091046.