

Analisis Sentimen Review Produk Acne Spot Treatment di Female Daily Menggunakan Algoritma K-Nearest Neighbor

Sanggir Dyah Fritama^{*}, Yudhi Raymond Ramadhan, Mutiara Andayani Komara

Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Purwakencana, Indonesia

Email: ^{1,*}Sanggirayah89@wastukencana.ac.id, ²yudhi.raymond@wastukencana.ac.id, ³mutiara@wastukencana.ac.id

Email Penulis Korespondensi: Sanggirayah89@wastukencana.ac.id

Abstrak—Jerawat selain berdampak negatif pada kulit, juga memberikan dampak psikologis yaitu rasa malu serta berkurangnya kepercayaan diri yang dapat dicegah dengan memberikan pengetahuan pengobatan untuk mengatasi jerawat. Salah satu produk digunakan untuk pengobatan jerawat adalah acne spot treatment. Banyaknya iklan dimedia sosial terkait produk acne spot treatment dapat mempengaruhi pembelian yang lebih besar, tetapi belum tentu produk tersebut bagus. Melihat ulasan langsung dari pengguna acne spot treatment dapat meningkatkan informasi terkait produk tersebut bagus atau tidak. Salah satu website yang menyediakan ulasan terkait produk kecantikan adalah Female Daily, Tetapi banyaknya ulasan tidak memungkinkan membaca secara keseluruhan tentunya hal ini memakan waktu. Analisis sentimen digunakan mengatasi masalah tersebut yang bertujuan mengklasifikasikan ulasan acne spot treatment menjadi Positif atau negatif. Dalam penelitian ini digunakan metode klasifikasi K-Nearest Neighbor karena memiliki konsep sederhana yang mudah diaplikasikan dan dimengerti. Hasil sentimen didapatkan ulasan yang paling banyak mengandung sentimen positif adalah Whitelab sebanyak 1.190 ulasan dan yang paling banyak mengandung sentimen negatif adalah Skin Game sejumlah 173 ulasan. Hasil klasifikasi menggunakan KNN akurasi terbaik adalah brand Whitelab sebesar 97%, kemudian Skin Game memperoleh accuracy 81%, dan nilai accuracy paling rendah adalah brand ERHA sebesar 75%.

Kata Kunci: Jerawat; Acne Spot Treatment; Female Daily; Analisis Sentimen; K-Nearest Neighbor

Abstract—Acne, besides having a negative impact on the skin, also has a psychological impact, namely embarrassment and reduced self-confidence, which can be prevented by providing medical knowledge to treat acne. One of the products used for acne treatment is acne spot treatment. The large number of advertisements on social media related to acne spot treatment products can influence larger purchases, but the product is not necessarily good. Seeing direct reviews from acne spot treatment users can increase information regarding whether the product is good or not. One website that provides reviews related to beauty products is Female Daily, but the large number of reviews does not allow reading in its entirety. Of course, this takes time. Sentiment analysis is used to overcome this problem, which aims to classify acne spot treatment reviews as Positive or negative. In this study, the classification method used is K-Nearest Neighbor because it has a simple concept that is easy to apply and understand. The sentiment results showed that the reviews that contained the most positive sentiments were Whitelab with 1,190 reviews, and the ones that contained the most negative sentiments were Skin Games with 173 reviews. The results of the classification using KNN are that the best accuracy is the Whitelab brand at 97%, then the Skin Game gets 81% accuracy, and the lowest accuracy value is the ERHA brand at 75%.

Keywords: Acne; Acne Spot Treatment; Female Daily; Sentimen Analysis; K-Nearest Neighbor

1. PENDAHULUAN

Jerawat adalah kondisi di mana pori-pori tersumbat menyebabkan kantong nanah meradang [1]. Jerawat secara stabil menempati merupakan tiga kondisi kulit teratas yang sering terjadi, menurut *Global Burden of Disease* (GBD), Jerawat terjadi pada 85 persen orang dewasa muda antara usia 12 hingga 25 tahun [2]. Di Indonesia sendiri jerawat adalah penyakit kulit umum yang terjadi sekitar 85% hingga 100% selama hidup seseorang yang ditemui pada perempuan yang berumur 14 sampai 17 tahun dan pada laki - laki berumur 16 sampai 19 tahun [3]. Kemudian berdasarkan *survey* yang dilakukan oleh *Himalaya* merek kesehatan dan perawatan menemukan 77% dari 1.000 wanita berjerawat di 10 kota besar di Indonesia yang menjadi responden, menyatakan bahwa 86% mengalami jerawat setiap bulan dan 32% setiap minggu [4]. Jerawat selain berdampak negatif pada kulit, jerawat juga memiliki dampak psikologis yaitu rasa malu dan kurang percaya diri [5]

Terdapat beberapa produk yang bisa mengobati timbulnya jerawat salah satunya adalah *Acne Spot Treatment*. *Acne Spot Treatment* merupakan obat yang digunakan untuk mengatasi masalah jerawat dengan cara mengoleskannya pada area yang timbul jerawat [6]. Adanya iklan di media sosial terkait suatu produk tentunya ini mempengaruhi minat beli yang lebih karena dengan iklan di media sosial dan info mengenai suatu produk terkait *best product* hal ini yang membuat masyarakat tertarik dalam pembelian produk [7] selain itu, iklan yang menggunakan konsep kecocokan dianggap kredibel dan diterima secara umum pada tingkat yang mudah dipahami konsumen. [8]. Ketika sudah membeli dan menggunakan produk tersebut, hasilnya belum tentu sesuai dengan iklan yang ditampilkan di media sosial, untuk itu pentingnya melihat suatu ulasan dari pengguna sebelum menggunakan produk *acne spot treatment*, agar tidak mudah terpengaruh oleh iklan yang menampilkan seolah olah produk tersebut bagus, karena dengan adanya *review* dari pengguna dapat menjadi pengetahuan tambahan yang dapat mempengaruhi asumsi dan pilihan konsumen tentang penjual atau produk yang bersangkutan [9]. Dalam *Website Female Daily* setiap orang yang sudah bergabung di *Female Daily* dapat menulis ulasan dan memberikan bintang atau rating terhadap suatu produk. Namun dengan banyaknya ulasan dalam suatu produk tidak memungkinkan membaca ulasan secara keseluruhan satu persatu hal ini pastinya memakan waktu [10].

Sebagai solusi dari permasalahan tersebut, digunakan analisis sentimen untuk mengklasifikasikan ulasan menjadi ulasan positif dan negatif sehingga dapat memberikan wawasan mengenai *acne spot treatment* tersebut bagus atau tidaknya [11], Selain itu, banyak ulasan juga berisi ulasan yang kategori emosionalnya tidak dapat diklasifikasikan dengan

jelas [12]. Analisis Sentimen adalah proses penggunaan teks analisis yang bersumber dari pendapat – pendapat, perasaan dan emosi yang disajikan dalam teks dan diperlukan untuk mengklasifikasikan informasi berbentuk teks ke dalam kategori positif dan negatif [13].

Dalam penelitian ini, analisis sentimen menggunakan salah metode klasifikasi yaitu *K-Nearest Neighbor*, karena algoritma ini relatif sederhana dan efektif dalam mengklasifikasikan teks, dianggap sederhana karena metode ini bersifat *self-learning* yang memungkinkannya mengeksplorasi struktur data yang ada dan mengklasifikasikan datanya sendiri berdasarkan tetangga terdekat [10]. Cara Kerja *K-Nearest Neighbor* menentukan jarak terpendek antara data *testing* yang akan dievaluasi dengan k tetangga terdekat pada data *training* [14].

Adapun beberapa penelitian yang berkaitan mengenai penelitian yaitu yang dilakukan oleh Abyan Ghiffarie, *dkk* pada tahun 2019 yang berjudul “Analisis Sentimen Terhadap Produk *The Body Shop Tea Tree Oil*” hasil dari penelitian tersebut adalah Produk *Tea Tree Oil The Body Shop* di *Femaledaily* menjadi salah satu produk yang banyak diminati, dengan tingkat akurasi menggunakan Naïve Bayes sebesar 61.51% [15]. Penelitian ini berkaitan dengan sumber data ulasan yang diambil oleh peneliti yaitu pada situs *Female Daily*

Penelitian berikutnya dilakukan oleh Elly Indrayuni, *dkk* yang berjudul “Implementasi Algoritma *Naive Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbors* untuk Analisa Sentimen Aplikasi Halodoc” mendapatkan hasil nilai akurasi pada *Naïve Bayes* sebesar 92.50%, SVM sebesar 93.00 % dan KNN sebesar 95.00%, penggunaan parameter C dan epsilon pada algoritma *Support Vector Machine* dan parameter k pada algoritma *K-Nearest Neighbors* yang mempengaruhi nilai akurasi yang diperoleh, sehingga KNN menghasilkan nilai akurasi yang tertinggi [16]. Penelitian ini berkaitan dengan algoritma yang diambil oleh peneliti yaitu *K-Nearest Neighbor* karna dilihat dari penelitian tersebut akurasi tertinggi didapatkan dengan Algoritma KNN.

Penelitian lainnya yaitu dilakukan oleh Restu Ramadhan, *dkk* yang berjudul “Analisis Sentimen Pada Ulasan Aplikasi *Maxim* di *Google Play* dengan *K-Nearest Neighbor*” dengan pembagian data latih dan data uji yaitu 90:10 nilai k = 5 mendapatkan hasil nilai *accuracy* 90,23%, [13]. Penelitian ini berkaitan dengan algoritma pada penelitian ini yaitu *K-Nearest Neighbor*.

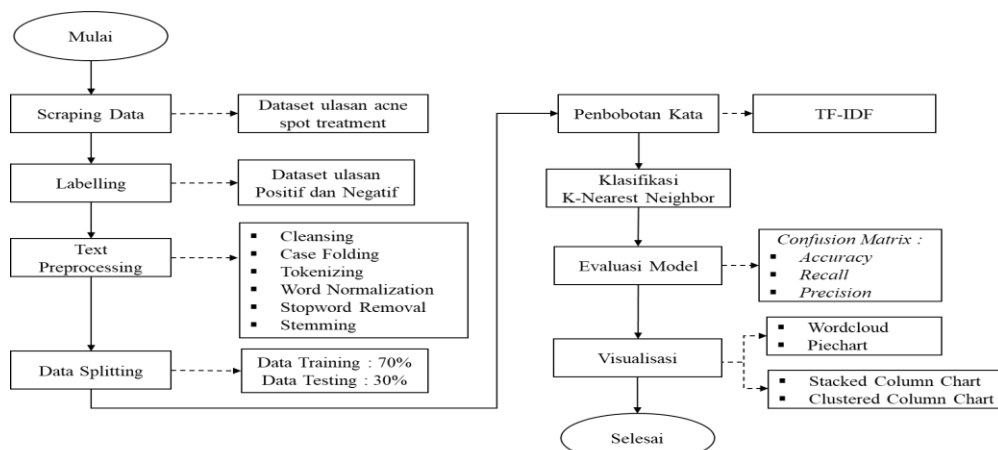
Penelitian selanjutnya dilakukan pada tahun 2021 oleh Aluisius Dwiki Adhi Putra & Safitri Juanita yang berjudul “Analisis Sentimen pada Ulasan Pengguna Aplikasi Bibit dan Bareksa dengan Algoritma KNN” hasil penelitian memperoleh nilai akurasi pada Aplikasi bibit 85,14% dan aplikasi bareksa 81.70% dengan menggunakan split data 60:40 [17]. Penelitian tersebut mengacu dengan sentimen yang digunakan pada penelitian ini yaitu Sentimen Positif dan Negatif. Penelitian berikutnya oleh Jeremy Andre Septian, *dkk* pada tahun 2019 yaitu “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan *K-Nearest Neighbor*”, Akurasi menggunakan range nilai k = 1 sampai 30 yang didapatkan akurasi terbaik pada k = 23 sebesar 79,99% [18]. Penelitian tersebut berkaitan nilai acuan yang digunakan yaitu menggunakan parameter k.

Pada penelitian ini, peneliti menggunakan data ulasan 3 produk *acne spot treatment* yang diambil dari situs *Female Daily* berdasarkan brand peraih penghargaan *Best Product Acne Treatment* versi *Female daily* yang diklasifikasikan menggunakan Algoritma *K-Nearest Neighbor* dengan perhitungan jarak *Euclidean Distance*

Berdasarkan yang telah dijabarkan diatas, maka tujuan dari penelitian ini adalah untuk mengetahui sentimen pada produk *acne spot treatment* dengan brand *Skin Game*, *Whitelab* dan *Erha*, berdasarkan ulasan yang terdapat di *website female Daily*, dengan membagi 2 kategori sentimen yaitu sentimen positif dan negatif serta untuk mengetahui respon terhadap pengguna produk *acne spot treatment* melalui pendapat untuk mengetahui kualitas produk, yang dianalisis menggunakan algoritma *K-Nearest Neighbor*.

2. METODOLOGI PENELITIAN

Kerangka penelitian dijabarkan pada Gambar 1 yang dimulai dari pengambilan data menggunakan teknik *Scraping* hingga hasil akhir yang divisualisasikan.



Gambar 1. Kerangka Penelitian

2.1 Scraping Data

Dilakukan dengan pengambilan data menggunakan *Web Scraper.io* extension gratis yang terdapat pada *browser Google Chrome*. Mengambil data ulasan *acne spot treatment* dengan brand “*Skin Game*”, “*Whitelab*” dan “*Erha*” yang terdapat pada *website female daily*. Data yang diambil adalah data tanggal dan data ulasan serta diambil dengan rentang waktu sejak 05 Januari 2019 sampai dengan 26 Maret 2023. Data yang sudah diambil disimpan dalam bentuk *Excel*.

2.2 Labelling

Labelling dilakukan secara manual [10] [19] [20] dengan menggunakan *microsoft excel*, dengan mengelompokan setiap ulasan dengan memberikan kolom tambahan disetiap ulasan yang merupakan kalimat yang mengandung sentimen positif dan negatif.

2.3 Text Preprocessing

Text Preprocessing dilakukan mempersiapkan data ulasan sebelum digunakan, berfungsi untuk mentransformasikan data ulasan tersebut ke dalam format yang lebih baik agar data ulasan tersebut berkualitas tinggi dan siap untuk digunakan pada proses selanjutnya [21], yaitu dengan dilakukan *cleaning*, *case folding*, *tokenizing*, *word normalization*, *stopword removal* dan *stemming*, dengan menggunakan bahasa pemrograman *python* dengan *library* yang disediakan *python*.

2.4 Data Splitting

Selanjutnya *dataset* ulasan yang telah dikumpulkan akan dipecah menjadi data *training* dan data *testing* dengan pembagian data *training* sebanyak 70% dan data *testing* sebesar 30% dari total *dataset* yang ada.

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

Dilakukan proses *term weighting* yaitu pemberian bobot atau nilai, dimana bobot menunjukkan pentingnya *term* (kata) terhadap dokumen. Semakin banyak kata, semakin besar nilai atau bobot kata tersebut. Rumus 1 adalah rumus dari Metode TF-IDF.

$$W = TF * (\log(\frac{D}{Df})) \quad (1)$$

Keterangan :

- W = Bobot setiap kata yang ada pada dokumen
- TF = Banyaknya kemunculan *term* di setiap dokumen
- D = Banyaknya dokumen
- Df = Banyaknya dokumen yang mengandung kata

2.6 K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor adalah mengklasifikasikan data baru berdasarkan data pembelajaran yang paling dekat dengan data uji dengan nilai k terbaik untuk algoritma ini bergantung pada data, nilai k yang tinggi mengurangi efek noise pada klasifikasi, tetapi membuat perbedaan antar klasifikasi menjadi lebih kabur [22] Nilai k yang digunakan adalah K = 9, lalu perhitungan jarak pada penelitian ini menggunakan *Euclidean Distance*. Rumus 2 untuk menghitung *Euclidean Distance* adalah sebagai berikut.

$$distance = \sqrt{\sum_{j=1}^n (X_{train}^j - X_{test})} \quad (2)$$

Keterangan :

- X_{train}^j = data *training* ke-j
- X_{test} = data *testing*
- j = baris ke-j dari tabel
- n = Banyaknya data *training*

2.7 Evaluasi (Confusion Matrix)

Evaluasi model ini dilakukan proses pengujian performa terhadap *K-Nearest Neighbor*. Evaluasi menggunakan *confusion matrix* untuk mengukur dan menghasilkan nilai *accuracy*, *recall* dan *precision*. *Accuracy* merupakan tingkat kedekatan antara nilai prediksi yang benar dengan nilai yang aktual [23]. *Precision* merupakan nilai sebenarnya antara prediksi dan data yang diinginkan sedangkan *Recall* adalah nilai keberhasilan model dalam mendapatkan kembali sebuah informasi [24].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Dimana TP (*True Positive*) adalah data yang diprediksi positif dan data sebenarnya positif. Kemudian TN (*true Negative*) adalah data yang diprediksi negatif dan sebenarnya negatif, lalu FP (*False Positive*) adalah data yang diprediksi positif tapi sebenarnya negatif dan FN (*False Negative*) adalah data yang diprediksi negatif tapi sebenarnya positif [23].

2.8 Visualisasi

Visualisasi menggunakan *software Microsoft Power BI* yang berfungsi untuk mem-visualisasikan hasil penelitian seperti menggunakan *wordcloud* dari data ulasan, dengan menggunakan *tools Stacked Colum Chart, clustered colum chart dan Pie Chart*.

3. HASIL DAN PEMBAHASAN

Analisis sentimen *acne spot treatment* di *female daily* melalui beberapa langkah pemrosesan untuk menghasilkan data yang diinginkan. Berikut adalah langkah-langkah dan hasil pengolahan data.

3.1 Scraping Data

Scraping data dilakukan pada 3 *brand* produk *acne spot treatment* memperoleh data keseluruhan sebanyak 2.618 data ulasan. Tabel 1 menunjukkan rincian data pada masing – masing *brand*.

Tabel 1. Hasil Scraping Data

Dataset	Jumlah Data
<i>Skin Game</i>	838
<i>Whitelab</i>	1.244
<i>Erha</i>	536

3.2 Labelling Data

Hasil dari data *scraping* selanjutnya dilakukan proses *labelling*. Tabel 2 merupakan contoh *labelling* pada penelitian ini.

Tabel 2. Data hasil labelling

No	Ulasan	Sentimen
1	Lama banget meresap dikulit dan kalau dibawa tidur suka nempel di spreï bantal. Kalaupun udah meresap ini jadi peelling dan yang ditotolin jadi lebih berminyak kulitnya. Untuk mengempeskan jerawat agak lama dan ini buat area sekitarnya jadi muncul bruntusan kecil di aku.	Negatif
2	Acne spot treatment terbaik yang pernah aku coba sih sejauh ini, untuk jerawat biasa dan baru muncul 2 hari udah Kempes di aku kalau malemnya aku pakein ini, minim bekas sih, untuk teksturnya kaya clay warna pink gituu , oke banget buat dicoba!	Positif

3.3 Text Preprocessing

Text preprocessing data ulasan melalui beberapa tahapan yaitu tahapan *cleansing, case folding, tokenizing, Word Normalization, stopword removal, dan stemming*,

3.3.1 Cleaning

Cleaning ini dilakukan untuk mengurangi *noise* pada data ulasan, yaitu menghapus tanda baca, angka, simbol – simbol atau emoticon pada data ulasan yang didapatkan dari *female daily*. menggunakan *library RegEx (Regular Expresion)*. Pada Tabel 3 adalah *sample* data yang menunjukkan proses sebelum dan sesudah *cleaning*.

Tabel 3. Contoh Hasil Cleaning

Sebelum Cleaning	Sesudah Cleaning
Acne spot treatment terbaik yang pernah aku coba sih sejauh ini, untuk jerawat biasa dan baru muncul 2 hari udah Kempes di aku kalau malemnya aku pakein ini, minim bekas sih, untuk teksturnya kaya clay warna pink gituu , oke banget buat dicoba!	Acne spot treatment terbaik yang pernah aku coba sih sejauh ini untuk jerawat biasa dan baru muncul hari udah Kempes di aku kalau malemnya aku pakein ini minim bekas sih untuk teksturnya kaya clay warna pink gituu oke banget buat dicoba

3.3.2 Case Folding

Mengubah semua teks ulasan menjadi huruf kecil (*lowercase*) karena hasil data *scraping* memiliki huruf yang tidak konsisten seperti terdapat huruf kapital dan huruf kecil, untuk itu dilakukan *case folding* untuk mengubah huruf pada teks ulasan agar menjadi data yang seragam. Berikut ditunjukkan pada Tabel 4 yang merupakan contoh data hasil dilakukannya *case folding*.

Tabel 4. Contoh data hasil *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
Acne spot treatment terbaik yang pernah aku coba sih sejauh ini untuk jerawat biasa dan baru muncul hari udah Kempes di aku kalau malemnya aku pakein ini minim bekas sih untuk tekstur nya kaya clay warna pink gitu oke banget buat dicoba	acne spot treatment terbaik yang pernah aku coba sih sejauh ini untuk jerawat biasa dan baru muncul hari udah kempes di aku kalau malemnya aku pakein ini minim bekas sih untuk tekstur nya kaya clay warna pink gitu oke banget buat dicoba

3.3.3 Tokenizing

Tokenizing digunakan untuk memisahkan kalimat menjadi per-kata. Pemisahan dengan menggunakan *library nltk tokenize* dengan fungsi *word_tokenize* pada *python*. Tabel 5 adalah contoh data sebelum dilakukan *tokenizing* dan sesudah dilakukan *tokenizing*.

Tabel 5. Contoh data hasil *Tokenizing*

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
acne spot treatment terbaik yang pernah aku coba sih sejauh ini untuk jerawat biasa dan baru muncul hari udah kempes di aku kalau malemnya aku pakein ini minim bekas sih untuk tekstur nya kaya clay warna pink gitu oke banget buat dicoba	['acne', 'spot', 'treatment', 'terbaik', 'yang', 'pernah', 'aku', 'coba', 'sih', 'sejauh', 'ini', 'untuk', 'jerawat', 'biasa', 'dan', 'baru', 'muncul', 'hari', 'udah', 'kempes', 'di', 'aku', 'kalau', 'malemnya', 'aku', 'pakein', 'ini', 'minim', 'bekas', 'sih', 'untuk', 'tekstur', 'nya', 'kaya', 'clay', 'warna', 'pink', 'gituu', 'oke', 'banget', 'buat', 'dicoba']

3.3.4 Word Normalization

Word Normalization dilakukan membenarkan singkatan atau kata yang salah eja sebagai kata yang benar, mengubah dimensi kata yang memiliki arti yang sama tetapi salah eja. Contohnya “dgn” dinormalisasi menjadi “dengan”. Pada Tabel 6 adalah contoh data sebelum dilakukan *Word Normalization* dan sudah dilakukan *word normalization*.

Tabel 6. Contoh Hasil data *Word Normalization*

Sebelum <i>Word Normalization</i>	Sesudah <i>Word Normalization</i>
['acne', 'spot', 'treatment', 'terbaik', 'yang', 'pernah', 'aku', 'coba', 'sih', 'sejauh', 'ini', 'untuk', 'jerawat', 'biasa', 'dan', 'baru', 'muncul', 'hari', 'udah', 'kempes', 'di', 'aku', 'kalau', 'malemnya', 'aku', 'pakein', 'ini', 'minim', 'bekas', 'sih', 'untuk', 'tekstur', 'nya', 'kaya', 'clay', 'warna', 'pink', 'gituu', 'oke', 'banget', 'buat', 'dicoba']	['acne', 'spot', 'treatment', 'terbaik', 'yang', 'pernah', 'aku', 'coba', 'sih', 'sejauh', 'ini', 'untuk', 'jerawat', 'biasa', 'dan', 'baru', 'muncul', 'hari', 'sudah', 'kempes', 'di', 'aku', 'kalau', 'malemnya', 'aku', 'pakaikan', 'ini', 'minim', 'bekas', 'sih', 'untuk', 'tekstur', 'nya', 'kayak', 'clay', 'warna', 'pink', 'begitu', 'oke', 'banget', 'buat', 'dicoba']

3.3.5 Stopword Removal

Proses *Stopword Removal* berfungsi menghilangkan kata pada data teks ulasan yang tidak memiliki makna yaitu kata – kata penghubung seperti kata ‘itu’, ‘ke’, ‘yang’, ‘ko’, dan lain sebagainya, akan dihapus dengan menggunakan *library NLTK (Natural Language Toolkit)* yang menyediakan kata-kata *stopword* dalam bahasa Indonesia. Contoh data ulasan sebelum dilakukan *Stopword* dan sesudah dapat dilihat pada Tabel 7.

Tabel 7. Contoh Hasil data *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
['acne', 'spot', 'treatment', 'terbaik', 'yang', 'pernah', 'aku', 'coba', 'sih', 'sejauh', 'ini', 'untuk', 'jerawat', 'biasa', 'dan', 'baru', 'muncul', 'hari', 'sudah', 'kempes', 'di', 'aku', 'kalau', 'malemnya', 'aku', 'pakaikan', 'ini', 'minim', 'bekas', 'sih', 'untuk', 'tekstur', 'nya', 'kayak', 'clay', 'warna', 'pink', 'begitu', 'oke', 'banget', 'buat', 'dicoba']	['acne', 'spot', 'treatment', 'terbaik', 'coba', 'jerawat', 'muncul', 'kempes', 'malemnya', 'pakaikan', 'minim', 'bekas', 'tekstur', 'kayak', 'clay', 'warna', 'pink', 'oke', 'banget', 'dicoba']

3.3.6 Stemming

Tahap *stemming* diproses untuk pembentukan kata dasar dengan cara mengubah atau menghilangkan kata yang memiliki imbuhan seperti ‘mem’, ‘per’, ‘ber’, ‘an’ dengan menggunakan *library Sastrawi*. Contoh hasil pada proses *stemming* ditunjukkan pada Tabel 8.

Tabel 8. Contoh hasil data *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
['acne', 'spot', 'treatment', 'terbaik', 'coba', 'jerawat', 'muncul', 'kempes', 'malemnya', 'pakaian', 'minim', 'bekas', 'tekstur', 'kayak', 'clay', 'warna', 'pink', 'oke', 'banget', 'dicoba']	['acne', 'spot', 'treatment', 'baik', 'coba', 'jerawat', 'muncul', 'kempes', 'malemnya', 'pakai', 'minim', 'bekas', 'tekstur', 'kayak', 'clay', 'warna', 'pink', 'oke', 'banget', 'coba']

3.4 Data *Splitting*

Pada data *Splitting*, data *training* akan digunakan untuk *machine learning* mempelajari data untuk klasifikasi KNN, sedangkan data *testing* digunakan melakukan prediksi pada data. Pembagian data menggunakan *Library Scikit-learn* dengan fungsi *model selection* dan *train_test_split*. pada Tabel 9 memperlihatkan jumlah data hasil pembagian data setiap *brand* dan Gambar 2 adalah *code* pembagian data.

Tabel 9. Jumlah data hasil *Splitting*

Dataset	Data Training	Data Testing
Skin Game	586	252
Whitelab	870	374
Erha	475	161

```
from sklearn import model_selection
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = model_selection.train_test_split(data['ulasan_clean'], data['Sentimen'], test_size=0.3, random_state=25)
```

Gambar 2. *Code* data *Splitting*

3.5 Term Frequency-Inverse Document Frequency (TF-IDF)

Pada tahapan pembobotan kata pada dokumen ulasan *Female Daily*, perhitungan dilakukan menggunakan menggunakan *library scikit-learn TfidfVectorizer*. Pada Gambar 3 merupakan implementasi *code* penggunaan TF-IDF.

```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vect_7030 = TfidfVectorizer(max_features=5000)
tfidf_vect_7030.fit(data['ulasan_clean'].astype('U').values)
train_X_tfidf_7030 = tfidf_vect_7030.transform(df_train70['ulasan_clean'].astype('U').values)
test_X_tfidf_7030 = tfidf_vect_7030.transform(df_test30['ulasan_clean'].astype('U').values)

print(train_X_tfidf_7030)
✓ 0.1s
(0, 2812) 0.20623540037003815
(0, 2745) 0.11056516870529724
(0, 2260) 0.3158132739932566
(0, 2140) 0.19583543361552166
(0, 2022) 0.10700958159745443
(0, 2006) 0.22850844249002678
```

Gambar 3. *Code* TF-IDF

3.6 K-Nearest Neighbor (KNN)

Selanjutnya dilakukan klasifikasi menggunakan *K-Nearest Neighbor*. Klasifikasi KNN menggunakan *library Scikit-Learn* dengan fungsi *KNeighborsClassifier*, dengan nilai *k* tetangga dimasukan dengan nilai *k=9*, kemudian metrik jarak yang digunakan adalah *euclidean distance*. Gambar 4 merupakan *code* yang digunakan dalam pengklasifikasian KNN.

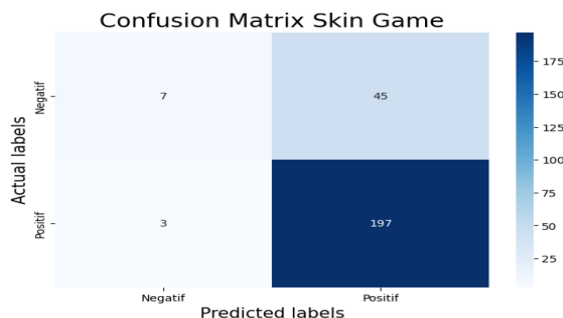
```
# Inisialisasi model
classifier_knn = KNeighborsClassifier(n_neighbors=9, weights='distance', metric='euclidean')
classifier_knn.fit(train_X_tfidf_7030, df_train70['Sentimen'])
# prediksi data testing
pred_knn_tv_3 = classifier_knn.predict(test_X_tfidf_7030)
print(classification_report(df_test30['Sentimen'], pred_knn_tv_3))
```

Gambar 4. *Code* K-Nearest Neighbor

3.7 Evaluasi (*Confusion Matrix*)

Tahap evaluasi menggunakan *Confusion matrix* dengan bahasa pemrograman *Python* menggunakan *library scikit-learn* dengan fungsi *confusion_matrix* yang akan menghitung nilai *Accuracy*, *Precision* dan *Recall*. Evaluasi *confusion matrix* yang dihasilkan dari *brand Skin Game* menggunakan *K-Nearest Neighbor* yaitu nilai *Accuracy* sebesar 81%, Gambar 5 menunjukkan *confusion matrix* beserta *Accuracy*, *Precision* dan *Recall* yang berasal dengan sekumpulan data *testing brand acne spot treatment Skin Game*.

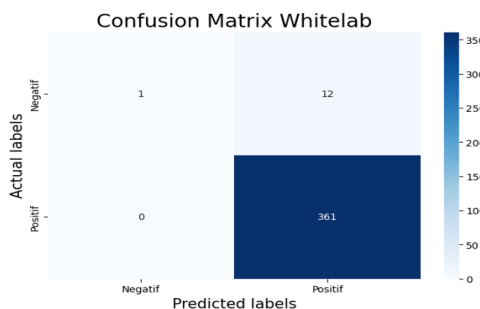
	precision	recall	f1-score	support
Negatif	0.70	0.13	0.23	52
Positif	0.81	0.98	0.89	200
accuracy			0.81	252
macro avg	0.76	0.56	0.56	252
weighted avg	0.79	0.81	0.75	252



Gambar 5. Confusion matrix Skin Game

Pada brand acne spot Whitelab dengan menggunakan K-Nearest Neighbor didapatkan Accuracy sebesar 97%. Pada Gambar 6 yang menunjukkan confusion matrix beserta nilai Accuracy, Precision dan Recall yang bersumber dengan data testing acne spot treatment Whitelab.

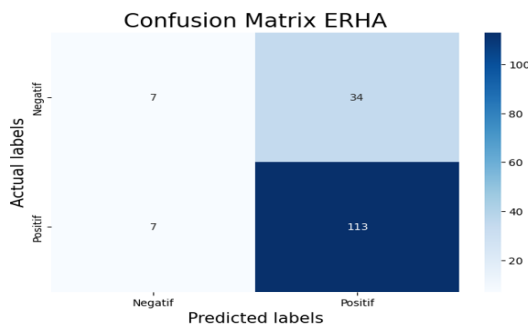
	precision	recall	f1-score	support
Negatif	1.00	0.08	0.14	13
Positif	0.97	1.00	0.98	361
accuracy			0.97	374
macro avg	0.98	0.54	0.56	374
weighted avg	0.97	0.97	0.95	374



Gambar 6. Confusion Matrix Whitelab

Pada brand acne spot ERHA dengan menggunakan algoritma K-Nearest Neighbor didapatkan Accuracy sebesar 75%. Dijabarkan pada Gambar 7 yang menunjukkan confusion matrix beserta nilai Accuracy, Precision dan Recall yang bersumber dengan sekumpulan data uji brand acne spot treatment ERHA.

	precision	recall	f1-score	support
Negatif	0.50	0.17	0.25	41
Positif	0.77	0.94	0.85	120
accuracy			0.75	161
macro avg	0.63	0.56	0.55	161
weighted avg	0.70	0.75	0.70	161



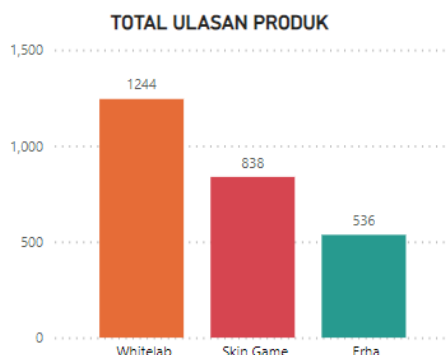
Gambar 7. Confusion Matrix ERHA

Berdasarkan hasil *evaluasi* dari perhitungan menggunakan *confusion matrix*, bahwa *brand acne spot treatment* yang memperoleh nilai *accuracy* tertinggi adalah *brand Whitelab* dengan nilai *accuracy* sebesar 97%.

3.8 Visualisasi

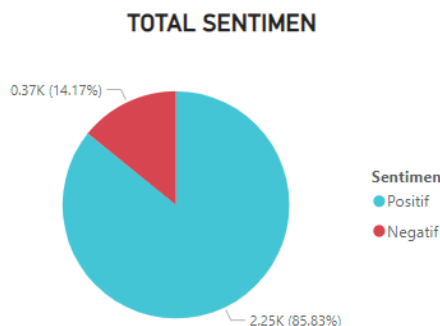
Visualisasi dilakukan untuk dilakukan untuk menampilkan data dengan beberapa visualisasi grafik atau bagan yang ada pada *tools* yang disediakan *Microsoft Power BI*.

- a. *Stacked Colum Chart*, pada bagan ini digunakan untuk menunjukan grafik jumlah keseluruhan ulasan *produk acne spot treatment* yang didapatkan dari *website female daily* dapat dilihat pada Gambar 8.



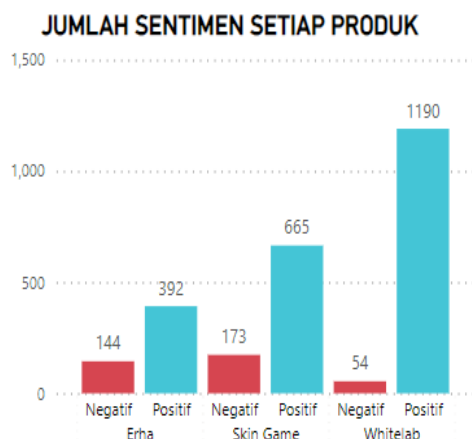
Gambar 8. *Stacked Colum Chart Total Ulasan*

- b. *Pie Chart*, pada bagan ini menunjukan total sentimen dari keseluruhan *brand* baik sentimen negatif maupun sentimen positif pada Gambar 9.

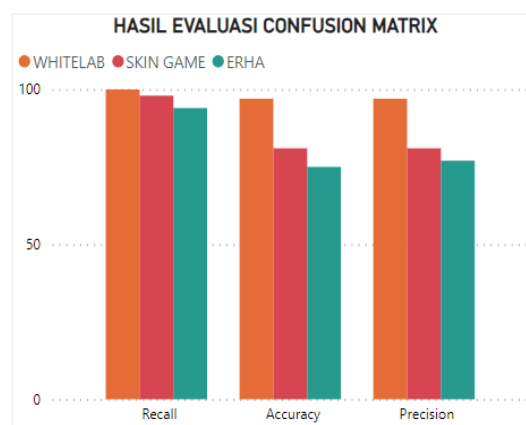


Gambar 9. *Pie Chart Total Sentimen*

- c. *Clustered Column Chart*, pada grafik ini menunjukan hasil *labelling* pada data ulasan *setiap brand* produk *acne spot treatment* yang ditunjukan pada Gambar 10. Serta pada Gambar 11 menampilkan grafik dari hasil perhitungan menggunakan *confusion matrix* nilai *accuracy*, *precision* dan *recall* untuk *setiap brand*.

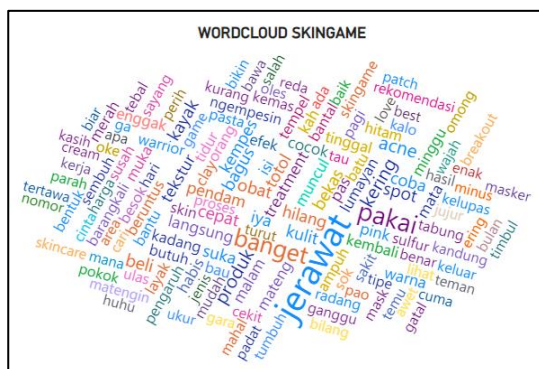


Gambar 10. Sentimen setiap produk

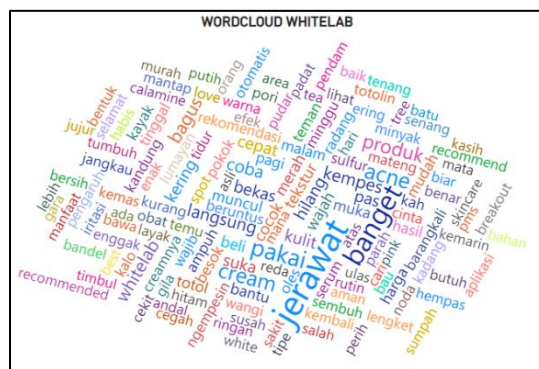


Gambar 11. Hasil Evaluasi *Confusion matrix*

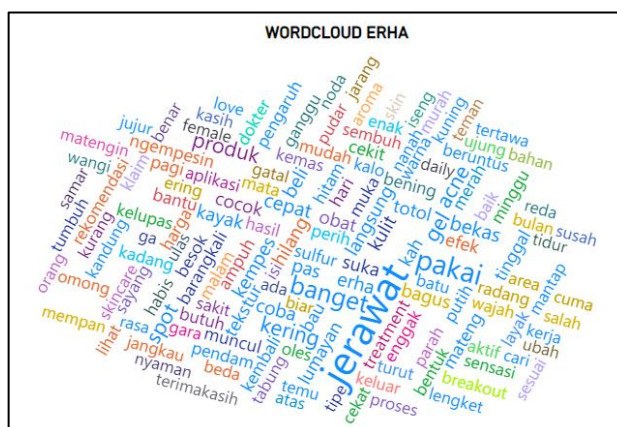
- d. *WordCloud*, berfungsi untuk menunjukan kata – kata yang sering muncul dalam ulasan produk *acne spot treatment* dengan 150 kata teratas yang divisualisasikan baik yang mengandung sentimen positif maupun sentimen negatif.



Gambar 12. Wordcloud Skin Game



Gambar 13. Wordcloud Whitelab



Gambar 14. Wordcloud ERHA

3. KESIMPULAN

Hasil penelitian yang telah dilakukan tentang analisis sentimen *review* produk *acne spot treatment* dari 3 *brand* dengan menggunakan *K-Nearest Neighbor*. Dari data hasil *labelling*, ulasan yang paling banyak mengandung sentimen positif adalah *acne spot treatment* dengan *brand Whitelab* sejumlah 1.190 dan yang paling banyak mengandung sentimen negatif adalah *brand Skin Game* sebanyak 173. Setelah dilakukan *Text Preprocessing*, *splitting* data dan pembobotan kata menggunakan TF-IDF dan klasifikasi menggunakan *K-Nearest Neighbor* dengan perbandingan data *training* dan data *testing* 70:30 serta penggunaan *euclidean distance* dan nilai K tetangga = 9, diperoleh nilai akurasi tertinggi adalah *Acne Spot treatment brand Whitelab* dengan nilai *accuracy* 97%, *Precision* 97% dan *recall* 100%, kemudian peringkat kedua adalah *Skin Game* dengan nilai *accuracy* 81%, *precision* 81% dan *Recall* 98%, dan nilai *accuracy* paling rendah adalah *brand ERHA* yang memperoleh nilai *Accuracy* 75%, *Precision* 77% dan *recall* 94%. Nilai ini menunjukkan algoritma *K-Nearest Neighbor* dapat bekerja dengan baik dalam pengklasifikasian 3 *brand acne spot treatment* ini. Adapun beberapa yang bisa dilakukan untuk penelitian kedepannya seperti menambahkan variasi beberapa *brand acne spot treatment* lagi karena masih banyak produk *acne spot treatment* kemudian kedepannya akan lebih banyak *brand acne spot treatment* terbaru yang lebih *variatif* serta dapat menggunakan algoritma *machine learning* lainnya dalam menganalisis sentimen.

REFERENCES

- [1] N. Sifatullah and Z. Zulkarnain, "Jerawat (Acne vulgaris): Review penyakit infeksi pada kulit," *Pros. Semin. Nas. Biol.*, no. November, pp. 19–23, 2021, [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/psb/article/view/22212%0Ahttp://journal.uin-alauddin.ac.id/index.php/psb/article/download/22212/12470>
- [2] P. Cintia, N. Artasih, W. Mulianingsih, S. Nirmala, and L. Mariam, "HUBUNGAN PERILAKU MEMBERSIHKAN WAJAH DENGAN KEJADIAN ACNE VULGARIS PADA MAHASISWA LAKI-LAKI Email : putucintianandaartasih@gmail.com," vol. 13, no. October 2022, pp. 267–275, 2023.
- [3] A. Syahputra, S. Anggreni, Y. Handayani, and M. Rahmadhani, "Pengaruh Makanan Akibat Timbulnya Acne Vulgaris (Jerawat) pada Mahasiswa Mahasiswi FK UISU Tahun 2020," *J. Kedokt. STM*, vol. 4, no. 2, pp. 75–82, 2021.
- [4] D. Mutiah, "Survei: 77 Persen Perempuan Berjerawat di Indonesia Jadi Korban Acne Shaming," *liputan6*, 2020. <https://www.liputan6.com/lifestyle/read/4337072/survei-77-persen-perempuan-berjerawat-di-indonesia-jadi-korban-acne-shaming> (accessed Apr. 10, 2023).
- [5] N. Aini, I. Herdiani, and B. Brahmantia, "Hubungan Acne Vulgaris (Av) Dengan Kepercayaan Diri Pada Mahasiswa Universitas Muhammadiyah Purwokerto Angkatan 2021," *J. Kesehat. Tambusai*, vol. 3, no. 3, pp. 434–441, 2022, doi: 10.31004/jkt.v3i3.6595.

- [6] A. Q. Azhari, D. Mayasari, and R. Rusli, "Formulasi Sediaan Gel Total Jerawat Berbahan Aktif Ekstrak Kulit Buah Naga Merah (*Hylocereus polyrhizus*) Terhadap Bakteri *Staphylococcus aureus* dan *Propionibacterium acnes*," *Proceeding Mulawarman Pharm. Conf.*, vol. 14, pp. 359–365, 2021, doi: 10.25026/mpc.v14i1.603.
- [7] C. M. A. Sari, D. Purbowati, F. Fadlillah, and Fashhan, "Literature Review Keputusan Pembelian Melalui Minat Beli: Promosi Iklan Tv Dan Media Sosial," *J. Manaj. Pendidik. Dan Ilmu Sos.*, vol. 3, no. 1, pp. 175–190, 2022, doi: 10.38035/jmpis.v3i1.864.
- [8] A. Fazriatul and A. C. Februadi, "Menelusuri Persepsi Acne Fighter Terhadap Konsep Iklan Yang Tidak Sesuai," *IRWNS; Ind. Res. Work. Natl. Semin.*, pp. 13–14, 2022.
- [9] A. Halim and M. J. Tyra, "Pengaruh Online Consumer Review Dan Influencer Terhadap Keputusan Pembelian Produk Di Marketplace Shopee," *Bul. Ekon. Manajemen, Ekon. Pembangunan, Akunt.*, vol. 18, no. 1, p. 99, 2021, doi: 10.31315/be.v18i1.5630.
- [10] A. N. S. Rahayu, T. I. Hermanto, and I. M. Nugroho, "Sentiment Analysis Using K-Nearest Neighbor Based on Particle Swarm Optimization According To Sunscreen'S Reviews," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1639–1646, 2022, doi: 10.20884/1.jutif.2022.3.6.425.
- [11] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.
- [12] 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.
- [13] R. Ramadhan, M. Afdal, I. Permana, and M. Jazman, "Analisis Sentimen pada Ulasan Aplikasi Maxim di Google Play Store dengan K-Nearest Neighbor," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 3, p. 715, 2023, doi: 10.30865/jurikom.v10i3.6396.
- [14] T. Setiyorini, ; Rizky, and T. Asmono, "Implementation of K-Nearest Neighbor and Gini Index Method in Classification of Student Performance," *Techno Nusa Mandiri J. Comput. Inf. Technol.*, vol. 16, no. 2, pp. 121–126, 2019, [Online]. Available: <https://ejournal.nusamandiri.ac.id/index.php/techno/article/view/747>
- [15] A. Ghiffarie, K. D. A. Salsabila, and R. P. Baistama, "Analisis Sentimen Terhadap Produk The Body Shop Tea Tree Oil," *J. Teknol. Manaj. Inform.*, vol. 5, no. 1, 2019.
- [16] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc," *Fakt. Exacta*, vol. 14, no. 2, p. 64, 2021, doi: 10.30998/faktorexacta.v14i2.9697.
- [17] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [18] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [19] A. W. Sari, T. I. Hermanto, and M. Defriani, "Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine," vol. 8, no. 3, pp. 1366–1378, 2023.
- [20] N. Legiawati, T. I. Hermanto, and Y. R. Ramadhan, "Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naïve Bayes Berbasis PSO," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 930, 2022, doi: 10.30865/jurikom.v9i4.4629.
- [21] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [22] H. Rofiq, K. C. Pelangi, and Y. Lasena, "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes," *J. Manaj. Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 8–15, 2020, [Online]. Available: <http://mahasiswa.dinus.ac.id/docs/skripsi/jurnal/19417.pdf>
- [23] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [24] Sudioanto, P. Wahyuningtias, H. W. Utami, U. A. Raihan, and H. N. Hanifah, "Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study : Internet Selebgram Rachel Vennya Escape From Quarantine) Perbandingan Metode Random Forest Dan Support Vector Machine Pada Analisis Sentimen Twitt," *Jutif*, vol. 3, no. 1, pp. 141–145, 2022.