Jurnal J-Innovation Vol.13, No.2, Desember 2024
P-ISSN 2338-2082
E-ISSN 2808-5620

Evaluasi Sentimen Review Produk Roundup Menggunakan Algoritma Support Vector Machine

Mohamad Khoiron[1], Dian Ahkam Sani [2], Mohammad Zoqi Sarwani [3], Muhammad Mahrus Ali [4], Khoirul Anwar [5], Muhammad Udin[6]
[123] Program Studi Rekayasa Perangkat Lunak, Fakultas Teknologi Informasi, Universitas Merdeka Pasuruan
[456]Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Merdeka Pasuruan Jl. H. Juanda No. 68, Pasuruan 67129, Jawa Timur, Indonesia
Email: [1]khoiron@unmerpas.ac.id, [2]dianahkam@unmerpas.ac.id, [3]zoqi.sarwani@unmerpas.ac.id, [4]mahrusali@unmerpas.ac.id, [5]khoirulanwar@unmerpas.ac.id, [6]imelekacong28@gmail.com

ABSTRACT

In today's digital era, more and more internet users are sharing their experiences and opinions about certain products. Sentiment analysis can be used to extract valuable information from the data generated by the shopee application users. This study aims to conduct a sentiment analysis of Roundup product reviews. The method used is the Support Vector Machine (SVM). SVM is an effective machine learning method for classifying text based on positive or negative sentiments. The purpose of this study is the SVM model which can be used to perform sentiment analysis automatically on Roundup product reviews. The results of this analysis can provide important insights for Roundup producers in understanding consumer perceptions of their products. In addition, this research can also be a guide for consumers in choosing and understanding weed killer products that suit their needs and preferences. In this study, the accuracy value was 80%, the precision value was 80%, the recall value was 100% and the value F1 score of 88.89%.

Keywords: *Sentiment analysis, Product review, Roundup, Support Vector Machine.*

ABSTRAK

Pada era digital saat ini, semakin banyak pengguna internet yang berbagi pengalaman dan pendapat mereka tentang produk-produk tertentu. Analisis sentimen dapat digunakan untuk menggali informasi berharga dari data yang dihasilkan oleh para pengguna aplikasi shopee. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap review produk Roundup.Metode yang digunakan adalah Support Vector Machine (SVM). SVM adalah salah satu metode pembelajaran mesin yang efektif dalam melakukan klasifikasi teks berdasarkan sentimen positif atau negatif. Tujuan dari penelitian ini adalah model SVM yang dapat digunakan untuk melakukan analisis sentimen secara otomatis pada review produk Roundup. Hasil analisis ini dapat memberikan wawasan penting bagi produsen Roundup dalam memahami persepsi konsumen terhadap produk mereka. Selain itu, penelitian ini juga dapat menjadi panduan bagi konsumen dalam memilih dan memahami produk pembasmi rumput liar yang sesuai dengan kebutuhan dan preferensi mereka.dalam penelitian ini menghasilkan nilai akurasi sebesar 80% , nilai Presisi sebesar 80%, nilai Recall sebesar 100% dan nilai F1 Score sebesar 88.89%.

Kata kunci: *Analisis sentiment, Review produk, Roundup, Support Vector Machine.*

I. PENDAHULUAN

Roundup adalah herbisida spektrum luas yang banyak digunakan untuk membunuh tanaman yang tidak diinginkan baik di bidang pertanian maupun pertanian lanskap nonpertanian. Perkiraan penggunaan di AS adalah antara 38 dan 48 juta pound per tahun [1]. Sebagian besar produk Roundup banyak di gunakan petani indonesia. Untuk membasmi rumput liar dilahan tanaman seperti jagung,padi,dan halaman rumah. melalui media sosial warga petani bisa tahu tentang produk Roundup penggunaa dan cara pemakain produk Roundup. untuk tanaman rumput liar yang ingin dibasmi oleh petani.

Media sosial salah kunci petani untuk mengetahui perkembangan tentang bidang pertanian dan sebagai media pembelajaran dalam teknologi pertanian. Jenis media sosial yang dikenal oleh masyarakat sangat beragam. Menurut hasil riset oleh *'We Are Social'* diantara banyaknya jenis media sosial tersebut, media sosial yang diminati orang Indonesia saat ini diantaranya Youtube, Facebook, Instagram, dan Twitter [2]. Untuk masyarakat petani lebih suka menggunakan media sosial you tube sebagai media belajar. Dalam media pembelajaran petani lebih suka ke praktek. Sehingga seiringnya waktu warga petani sering mendengar tentang produk Roundup yang di gunakan petani untuk membasmi rumput liar di media sosial youtube.

Pemanfaatan Roundup ini tidak hanya untuk digunakan untuk membantu petani, akan tetapi sebagai saranan pembantu masyarakat biasa yang dirumah memiliki halaman yang banyak rumput liarnya.dari manfaat itu masyrakat sendiri masih ada yang takut dalam penggunaan produk Roundup. Untuk agen toko yang menjual Roundup masih ada yang cenderung untuk penggunanya pada produknya. Dikarenakan takut tanahnya tidak bisa ditanami. Dan warga petani sendiri masih memilihki hasrat sentimen positif dan negatif terhadap produk Roundup. Maka dari perlu peningkatan untuk ketertarian terhadap produk roundup. Agar Agen toko bisa mengetahui tentang produk Roundup yang dijualnya. karena apabila terdapat lebih banyak sentimen negatif dari pada sentimen positif maka akan membuat produk menjadi buruk di mata petani dan para pembeli akan semakin berkurang.

Pemanfaatan Roundup ini tidak hanya untuk digunakan untuk membantu petani, akan tetapi sebagai saranan pembantu masyarakat biasa yang dirumah memiliki halaman yang banyak rumput liarnya.dari manfaat itu masyrakat sendiri masih ada yang takut dalam penggunaan produk Roundup. Untuk agen toko yang menjual Roundup masih ada yang cenderung untuk penggunanya pada produknya. Dikarenakan takut tanahnya tidak bisa ditanami. Dan warga petani sendiri masih memilihki hasrat sentimen positif dan negatif terhadap produk Roundup. Maka dari perlu peningkatan untuk ketertarian terhadap produk roundup. Agar Agen toko bisa mengetahui tentang produk Roundup yang dijualnya. karena apabila terdapat lebih banyak sentimen negatif dari pada sentimen positif maka akan membuat produk menjadi buruk di mata petani dan para pembeli akan semakin berkurang.

Dari permasalahan ini penulis akan menggunakan perangkat lunak sebagai menganalisa hasrat tersebut. untuk mendapatkan sentimen apa saja yang terjadi pada masyarakat baik sentimen positif maupun sentimen negatif. penelitian yang dilakukan [3]. Mereka melakukan penelitian menggunakan metode Naive bayes dengan pembobotan Log TF dan TF, pembobotan TF memiliki nilai akurasi yang lebih baik dibandingkan dengan pembobotan Log TF. Pembobotan TF memiliki nilai akurasi sebesar 55% sedangkan pembobotan Log TF hanya memiliki akurasi sebesar 52%. Sedangkan untuk melabeli review menggunakan Kapa Measure, pada pengujian ini Kappa Measure pada setiap penilai memiliki nilai sama yaitu 0.8. Pada penelitian selanjutnya dilakukan oleh [4] mereka mengimplementasikan metode menggunakan Naïve Bayes Classifier

sebagai pembelajaran mesin untuk secara otomatis mengklasifikasikan komentar positif atau negatif pada kosmetik produk. Hasil pengujian menunjukkan bahwa klasifikasi aplikasi memiliki akurasi dan presisi tinggi pada positif dan komentar negatif.

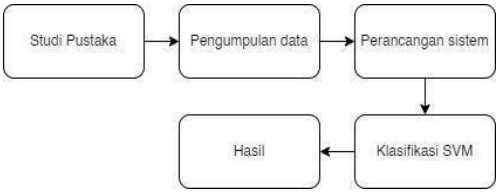
Analisis sentimen ialah proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini [5]. Analisis sentimen dilakukan guna menilai opini dan kecenderungan sebuah opini terhadap suatu topik baik negatif maupun positif [6]. Analisis sentimen dapat diterapkan pada opini semua bidang seperti ekonomi, politik, sosial dan hukum. Media sosial Twitter ini membuka jendela bagi para peneliti untuk mempelajari emosi, suasana hati, dan pendapat public melalui analisis sentimen [7]. Dalam penerapan analisis sentimen menggunakan metode machine learning terdapat beberapa metode yang sering digunakan seperti KNN, Naive Bayes, Random Forest dan Support Vectore Machine (SVM) Salah satu penelitian yang memanfaatkan data media sosial Twitter dengan metode machine learning adalah Lukmana dkk, dalam penelitiannya digunakan data politik dari twitter terkait sentimen calon presiden jokowi dan prabowo untuk dilakukan analisis sentimen dengan algoritma SVM, hasil akurasi yang diperoleh dari analisis sentimen yang dilakukan adalah 86% [8]. Sementara itu pada tahun 2019, pada peneletian selanjutnya dilakukan analisis sentimen terkait dengan efektifitas kebijakan pemerintah dalam Sistem Ganjil Genap di Tol Bekasi menggunakan SVM. Data yang diperoleh berasal dari multiplatform media sosial; twitter, youtube, facebook dan instagram. Total dataset yang digunakan adalah 440 dataset dengan polaritas 220 positif dan 220 negatif. Penelitian tersebut menghasilkan akurasi sebesar 78,18% [9].

Berdasarkan uraian dan latar belakang diatas. maka dari itu penulis tertarik untuk membuat sistem untuk menganalisis sentimen masyarakat petani terhadap produk roundup.

Berdasarkan komentar dari aplikasi shopee dengan menggunakan metode support vector machine (SVM), sehingga penulis memberi judul pada proposal ini " Analisis Sentimen Review Produk Roundup (Pembasmi Rumput Liar) Menggunakan Metode Support Vector Machine (SVM).

II. METODE PENULISAN

Bab ini akan menjelaskan tentang metodologi penelitian yang digunakan pada penelitian ini, yang terdiri dari (1) Studi pustaka (2) pengumpulan data (3) Perancangan sistem, (4) Klasifikasi SVM (6) hasil



Gambar 1. Alur Penelitian

1. Studi Pustaka

Metode pengumpulan data dengan metode kepustakaan dilakukan dengan pengumpulan jurnal, literatur, paper, makalah, buku, maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya analisis sentimen menggunakan metode Support Vectore Machine (SVM).

2. Perancangan Elektrik

Pada proses ini dilakukan pengumpulan dataset berupa ulasan produk Roundup pada aplikasi shopee di toko bagindara,dimana data ulasan yang diambil adalah username,rating dan komentar yang berjumlah 200 data yang selanjutnya akan dibagi oleh sistem dengan pembagian 80:20.pada table berikut akan menjelaskan tentang dataset pada penelitian.

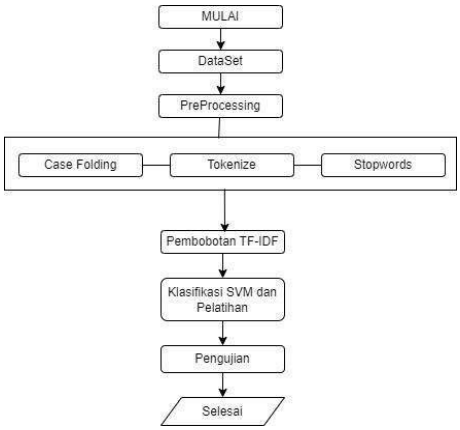
Tabel 1. *Dataset Komentar*

No	Username	Rating	Komentar	Label
1.	Rotoameerul	5	Barang sesuai dengan gambarnya bagus Trim ksh	Positif
2.	_7v6j7sup1	5	Cocok untuk: ok barang bagus	Positif
3.	s*****l	5	Alhamdulillah paket sudah diterima dengan baik,	Positif
4.	t*****d	5	Pengiriman cepat, respon cepat, oke semuanya	Positif
5.	l*****i	5	Alhamdulillah paketnya udah datang kemasan rapi	Positif

3. Perancangan Sistem *Flowchart*

Pada Tahapan dilakukan analisis sentimen dengan metode klasifikasi SVM dimulai dengan input dataset kemudian diproses pada tahapan pre-processing hingga proses klasfikasi sampai pada tingkat akurasi.

Pada tahapan Proses Dataset, data yang digunakan adalah ulasan Shopee tentang produk roundup yang diambil melalui teknik scrapping dengan bantuan Google Colab. Tahap scraping dimulai dengan menginput URL produk Shopee roundup menggunakan library re untuk pencarian, library requests untuk mengambil data melalui permintaan HTTP, dan library json untuk menyimpan hasil dalam format JSON. Data yang diperoleh adalah 200 komentar. Tahap dalam preprocessing dataset Tokenizing,Case Folding, Filtering dan Stopword.Selanjutnya, pada tahapan Term Frequency (TF), dilakukan penghitungan jumlah kemunculan kata pada dokumen.Inverse Document Frequency (IDF) adalah jumlah dokumen yang mengandung term yang dicari dalam keseluruhan dataset. Ini mendefinisikan kontribusi term tersebut dalam dokumen.Terakhir, pada tahapan Term Frequency-Inverse Document Frequency (TF-IDF), dilakukan pembobotan dengan menggabungkan metode term frequency dan inverse document frequency menggunakan rumus W = tf * idf. Metode ini berfungsi untuk mencari representasi nilai dari kumpulan data yang diberikan.



Gambar 2. Flowchart Sistem

4. Klasifikasi Support Vectore Machine (SVM)

Support Vector Machine (SVM) adalah sebuah metode klasifikasi yang efektif dalam memisahkan data ke dalam dua kategori berbeda. SVM mencari hyperplane optimal yang dapat memaksimalkan jarak antara kelas- kelas data, dengan menggunakan support vectors yang merupakan data-data yang berada pada atau dekat dengan margin. Kelebihan SVM antara lain kemampuannya dalam mengatasi masalah overfitting dan kemampuan untuk bekerja dengan baik dalam ruang fitur yang tinggi. SVM juga dapat mengatasi masalah non-linear dengan menggunakan fungsi kernel yang mentransformasi data ke dalam ruang fitur yang berbeda, seperti kernel linier, kernel

polynomial, dan kernel Gaussian (RBF) [10].

5. Perhitungan Manual

a. Pembobotan TF-IDF

Metode pembobotan yang diintegrasikan dari term frequency dan inverse document frequency dengan menggunakan rumus W = tf * idf metode ini berfungsi untuk mencari representasi nilai dari kumpulan data [11]. Berikut ini contoh data yang akan dihitung menggunakan TF-IDF sebagai berikut:

Tabel 2. *Data Dokumen Ulasan*

Data A	Sesuai Pesanan...Respon Cepat....	Positif
Data B	Rumput Tidak Mati	Negatif

Pada tabel 3 menggunakan contoh sebanyak 2 data yang akan dproses, dimana data tersebut akan dipisah menjadi bagian kecil untuk menghitung bobot pada setiap term/kata. Selanjutnya adalah memecah dokumen menjadi beberap kata untuk mencari term frekuensi.

Tabel 3. *Term Frekuensi*

Term	D1	D2
Sesuai	1	0
Pesanan	1	0
Respon	1	0
Cepat	1	0
Rumput	0	1
Tidak	0	1
Mati	0	1

Kemudian Langkah selanjutnya adalah membagi term frekuensi dengan Panjang dokumen untuk menormalisasi sehingga ditemukan nilai yang dijelaskan pada table dibawah ini :

Tabel 4. *Normalisasi Term Frekuensi*

D1	D2
0.25	0
0.25	0
0.25	0
0.25	0
0	0.333
0	0.333
0	0.333

Langkah selanjutnya mencari DF (dokumen frekuensi) banyaknya term yang muncul dalam setiap dokumen dan mencari IDF (jumlah df yang mengandung term) yang menggunakan rumus

$$idf_t = log\left(\frac{n}{df_t}\right)$$

T : term yang ternomalisasi
n : jumlah total dokumen dalam koleksi data
df_t : jumlah dokumen mengandung term
idf_t : Nilai normalisasi IDF dari term t

Perhitungan DF dan IDF yang dijelaskan pada table dibawah ini

Tabel 5. *DF dan IDF*

DF	IDF
1	0.301
1	0.301
1	0.301
1	0.301
1	0.301
1	0.301
1	0.301

Kemudian Langkah selanjutnya untuk mencari nilai TF-IDF nilai TF yang ternomalisasi dikalikan dengan nilai IDF yang akan dijelaskan pada table dibawah ini:

Tabel 6. *Nilai TF-IDF*

D1	D2
0.075257	0
0.075257	0
0.075257	0
0.075257	0
0	0.100343
0	0.100343
0	0.100343

Dibawah ini adalah sample dataset yang telah divektorisasi menggunakan TF-IDF

Tabel 7. *Sampel Data Vektorisasi*

X1	X2	Y
0,075	0	1
0,075	0	1
0,075	0	1
0,075	0	1
0	0,1	-1
0	0,1	-1
0	0,1	-1

b. Klasifikasi SVM

Perhitungan klasifikasi SVM dari tabel 7 hasil dari vektorisasi tf-idf maka akan dilakukan proses perhitungan manual menggunakan metode SVM.

- Menghitung matriks kernel K dengan rumus K[i, j] = X[i] . X[j], di mana K[i, j] Di mana titik (.)menunjukkan operasi perkalian dot antara vektor atribut X[i] dan X[j]. Pada kasus kernel linear, kita menghitung hasil dari produk titik antara dua vektor atribut. Dalam contoh ini, matriks desain X kita terdiri dari dua atribut: Atribut 1 dan Atribut 2. Oleh karena itu, kita akan menghitung produk titik antara vektor atribut X[i] dan X[j] sebagai elemen-elemen dari matriks kernel K [12]. Contoh perhitungan matriks kernel K berdasarkan data pelatihan yang diberikan:

  K = [X[0] • X[0], X[0] • X[1], X[0] • X[2], X[0] • X[3], X[0] • X[4], X[0] • X[5], X[0] • X[6]], [X[1] • X[0], X[1] • X[1], X[1] • X[2], X[1]• X[3], X[1] • X[4], X[1] • X[5], X[1] • [6]],[X[2] • X[0], X[2] • X[1], X[2] • [2], X[2]• X[3], X[2] • X[4], X[2] • X[5], X[2] • [6]], [X[3] • X[0], X[3] • X[1], X[3] • X[2], X [3]• X[3], X[3] • X[4], X[3] • X[5], X[3] • [6]], [X[4] • X[0], X[4] • X[1], X[4] • X[2], X[4] • X[3], X[4] • X[4], X[4] • [5], X[4] • [6]],[X [5] • X [0], X [5] • X [1], X [5] • X [2], X [5] • X [3], X [5] • X [4], X [5] • X [5], X [5] • [6]], [X [6] • X [0], X [6] • X [1], X [6] • X [2], X [6] • X [3], X [6] • X [4], X [6] • X [5], X [6] • X [6]]

Dalam kasus kernel linear, dengan cara mengalikan setiap elemen atribut pertama dengan elemen atribut pertama, dan setiap elemen atribut kedua dengan elemen atribut kedua, lalu menjumlahkan hasilnya. Jadi, menggunakan rumus ini, kita dapat menghitung matriks kernel K untuk data pelatihan yang diberikan.

Tabel 8. *Pehitungan Matix Kernel*

0,005	0	0	0	0	0	0
0	0,005	0	0	0	0	0
0	0	0,005	0	0	0	0
0	0	0	0,005	0	0	0
0	0	0	0	0,01	0,01	0,01
0	0	0	0	0,01	0,01	0,01
0	0	0	0	0,01	0,01	0,01

- Menggunakan metode optimisasi (misalnya, metode Lagrange multipliers) untuk mencari vektor Lagrange multipliers α yang meminimalkan fungsi objektif [13]:
  α = [0.0, 0.0, 0.0, 0.0, 0.2, 0.2, 0.2]

- Menghitung nilai koefisien w dengan menggunakan rumus:

  w = Σ(α[i] * y[i] * X[i])
  = (0.2*1*[0.075,0]) + (0.2*-1*[0, 0.1])
  = [0.015, -0.02]

- Nilai koefisien w ini menunjukkan arah vektor normal yang memisahkan dua kelas.
  b = y[j] - Σ(α[i] * y[i] * K[i, j])
  = 1-(0.2*1*0.005625+0.2*-1*0.0075)
  = 0.99375

  Nilai intersep b ini menunjukkan jarak garis ke titik asal (0,0) pada bidang fitur. Jadi, setelah menghitung secara manual, nilai koefisien SVM yang diperoleh adalah:
  Koefisien w: [0.015, -0.02] dan Intersep b: 0.99375

III. HASIL DAN PEMBAHASAN

Untuk melakukan proses pengambilan data, digunakan bahasa Python dan library re,request dan json yang di jalankan di Google Collaboratory yang selanjutnya akan disimpan di file csv yang akan diproses pada tahap selanjutnya.Script untuk simpan hasil scraping ditunjukkan pada gambar 2 dan hasil scraping pada tabel 1

```
# MENYIMPAN HASIL SCRAPPING
filename = 'data_ulasan.csv'
fieldnames = ['Author Username', 'Comment', 'Rating',
'Label']
with open(filename, 'w', newline='',
encoding='utf-8') as file:
writer = csv.writer(file)
    writer.writerow(fieldnames)
    writer.writerows(data_rows)
print('Data saved to', filename)
```

Gambar 3. Simpan Hasil Data

Tabel 9. *Hasil scraping Data*

No	Username	Rating	Komentar	Label
1	rotoameerul	5	Barang sesuai dengan gambarnya bagus Terima kasih	Positif
2	_7v6j7sup1	5	Cocok Untuk: ok barang bagus	Positif

Pada tahap pemrosesan data dataset akan dibagi menjadi feature dan label set. Di feature menggunakan text ulasan atau review sebagai feature. Label yang sesuai akan menjadi sentimen dari hasil ulasan. Kolom text adalah kolom ke 4 (indeks kolom dimulai dari 0) dalam dataset dan berisi teks dari data ulasan adalah kolom komentar yang berisi sentimen. Gunakan metode "iloc" dari dataframe panda untuk membuat feature set X dan label set y [14].

Pada tahap text preprocessing dilakukan preprocessing atau menyiapkan data ulasan sebelum melakukan proses klasifikasi. Tahap ini dibagi menjadi beberapa langkah, di antaranya scraping, cleaning/cleansing, normalisasi, stemming, tokenizing dan stopwords removal. Berikut merupakan proses langkah-langkah tersebut.

– Cleaning/cleansing bertujuan untuk membersihkan data reviews yang sudah diperoleh. Komponen-komponen yang dibersihkan merupakan komponen yang tidak berarti atau tidak relevan untuk proses pengklasifikasian data. Komponen tersebut di antaranya karakter spesial pada ulasan di antaranya hashtag, URL, link, tanda baca, angka, dan simbol emoji. Berikut ini merupakan tahap yang dilakukan dalam proses cleaning/ cleansing.

Tabel 10. *Hasil Cleaning Data*

Sebelum	Sesudah
Barang sesuai dengan gambarnya bagus Trim ksh	Barang sesuai dengan gambarnya bagus trim ksh
Cocok Untuk: ok barang bagus	Cocok untuk ok barang bagus
Alhamdulillah paket sudah diterima dengan baik	Alhamdulillah paket sudah diterima dengan baik
hasil belum terlihat setelah terlihat pemakaian Semoga hasil nya cepat terlihat	Hasil belum Hasil belum terlihat setelah terlihat pemakaian semoga hasil nya cepat terlihat

– Pada data ulasan yang digunakan, terdapat banyak penggunaan huruf kapital yang tidak konsisten sehingga case folding ini dibutuhkan mengubah semua huruf menjadi huruf kecil atau disebut lowercase Script untuk proses tersebut adalah sebagai berikut:

Tabel 11. *Hasil Case Folding*

Sebelum	Sesudah
Barang sesuai dengan gambarnya bagus Trim ksh	barang sesuai dengan gambarnya bagus trim ksh
Cocok Untuk: ok barang bagus	cocok untuk ok barang bagus

Alhamdulillah paket sudah diterima dengan baik,	alhamdulillah paket sudah diterima dengan baik
hasil belum terlihat setelah pemakaian. Semoga hasil nya cepat terlihat	hasil belum terlihat setelah pemakaian semoga hasil nya cepat terlihat

– Pada proses tokenisasi ini, tokenizer melakukan tugasnya untuk membagi sebuah kalimat menjadi beberapa bagian seperti kata- kata, frasa atau elemen bermakna yang lainnya. Berikut ini adalah hasil ulasan

Tabel 12. *Hasil Tokenizing*

Sebelum	Sesudah
barang sesuai dengan gambarnya bagus trim ksh	‘barang’, ‘sesuai’,‘dengan’,‘gambarnya’,‘bagus’,’trim’ ‘ksh’
ok barang bagus	ok’, ‘barang’, ’bagus’

– Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopword dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan stopword yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya Script untuk proses tersebut adalah sebagai berikut:

Tabel 13. *Hasil Stopwords*

Sebelum	Sesudah
Barang sesuai dengan gambarnya bagus Trim ksh	Barang sesuai dengan gambarnya bagus
ok barang bagus	barang bagus

– Pada tahap selanjutnya akan dilakukan proses TF-IDF. TF-IDF berasal dari kata TF dan IDF. Term Frequency adalah sama dengan berapa kali sebuah kata muncul dalam dokumen tertentu. Itu dihitung sebagai:

TF = (Frekuensi kata dalam dokumen) / (Total kata dalam dokumen)

Inverse Document Frequency adalah untuk kata tertentu sama dengan jumlah total dokumen, dibagi dengan jumlah dokumen yang mengandung kata tertentu. Log seluruh istilah dihitung untuk mengurangi dampak division. Itu dapat dihitung sebagai berikut:

IDF = Log ((Jumlah total dokumen) / (Jumlah dokumen yang mengandung kata))

```
from sklearn.feature_extraction.text import
TfidfVectorizer
tfidfconverter                                  =
TfidfVectorizer(max_features=2000, min_df=5,
max_df=0.7,
        stop_words=stopwords.words('indonesia
n'),ngram_range=(1,3))
X1                                              =
tfidfconverter.fit_transform(processed_twee ts).toarray()
print(str(X1))
print()
```

Gambar 4. Proses Pembobotan TF_IDF

Atribut max_features menentukan jumlah kata yang paling banyak muncul untuk membuat feature vektor. Kata-kata yang jarang muncul tidak memainkan peran utama dalam klasifikasi. Karenanya kami hanya menyimpan 2000 kata yang paling sering muncul dalam dataset. Nilai min_df dari 5 menentukan bahwa kata tersebut harus muncul di setidaknya 5 dokumen. Demikian pula, nilai max_df sebesar 0,7 menetapkan bahwa kata tersebut tidak boleh muncul di lebih dari 70 persen dokumen. Alasan di balik memilih 70 persen sebagai ambang batas adalah bahwa kata-kata yang muncul di lebih dari 70 persen dokumen terlalu umum dan kecil kemungkinannya untuk berperan dalam klasifikasi sentiment Terakhir, untuk mengonversi dataset menjadi feature vektor

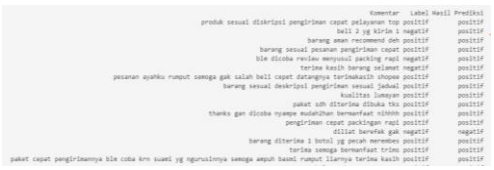
TF-IDF menggunakan metode fit_transform pada kelas Tfidf Vectorizer dan meneruskannya dengan dataset yang telah diproses sebelumnya.hasil dari proses tersbut dijelaskan pada gambar 5.

```
Komentar 4: ['pengiriman', 'cepat', ',', 'respon', 'cepat', ',', 'oke']
Nilai TF-IDF:
Kata: cepat, Nilai TF-IDF: 0.6195047364233505
Kata: oke, Nilai TF-IDF: 0.545412877154466
Kata: respon, Nilai TF-IDF: 0.4419379422404728
Kata: pengiriman, Nilai TF-IDF: 0.3513253907581676
```

Gambar 5. Hasil Pembobotan TF-IDF

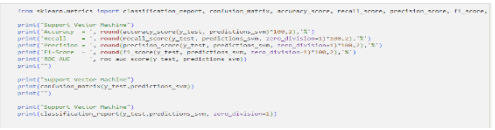
4.5 Membagi Data ke Training dan Test Set Pada tahap ini data review akan dibagi dengan perbandingan 80:20 menjadi data train dan data test digunakan untuk melatih model klasifikasi (pelatihan). Data training digunakan untuk melihat persentase algoritma klasifikasi yang berhasil mengklasifikasikan dengan benar yang ditunjukkan pada gambar 6.

Pada tahap selanjutnya setelah membagi data ke dalam training dan test set kemudian melatih model pada training set menggunakan metode Support Vector Machine yang akan melakukan proses prediksi kelas positif atau negatif dengan menguji data serta mempelajari pengetahuan yang terkandung dalam data latih. Pada data latih terdapat kelas sentimen positifi dan negatif, SVM akan mempelajari karakteristik kata-kata yang terdapat padai masing-masing kelas. Menggunakan modul sklearn untuk melatih model dengan memanggil metode "fit" pada objek classifier dan meneruskannya ke training feature set dan training label set . hasil Klasifikasi Svm pada gambar 7.



Gambar 6. Hasil Klasifiaksi SV

Pada tahap evaluasi model dilakukan Pengujian confusion matrix yang bertujuan untuk evaluasi hasil dari prediksi yang dilkakun metode klasifikasi support vector machine. Confusion script untuk confusion matrix seperti berikut ini:



Gambar 7. Proses Evaluasi Model

```
Support Vector Machine
Accuracy  =  87.5 %
Recall    =  96.15 %
Precision =  86.21 %
F1-Score  =  90.91 %
ROC AUC   =  0.8379120879120879

Support Vector Machine
              precision    recall  f1-score   support

           0       0.91      0.71      0.80        14
           1       0.86      0.96      0.91        26

    accuracy                           0.88        40
   macro avg       0.89      0.84      0.85        40
weighted avg       0.88      0.88      0.87        40
```

Gambar 8. Hasil Evaluasi Model

Pada proses evaluasi menggunakan confusion matrix menghasilkan nilai confusion matrix yang ditunjukkan pada table dibawah ini

Tabel 14. *Hasil Confusion Matrix*

Aktual	Prediksi	
	Negatif	Positif
Negatif	8	6
Positif	1	25
Akurasi	**87,5**	

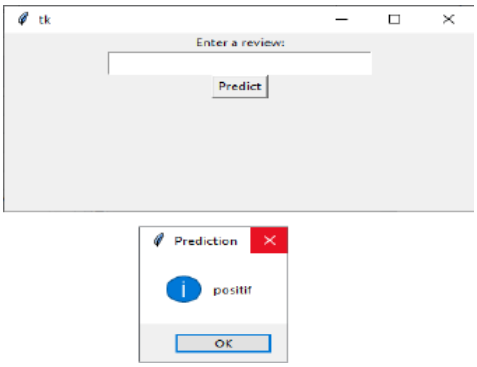
Tabel 14 menunjukkan hasil confusion matrix data ulasan produk roundup terdapat 25 data positifi yang benar diprediksi termasuk ke dalam kelas sentimen positif. dan 1 data positif yang diprediksi termasuk ke dalam kelas sentimen negatif, Kemudian terdapat 8 data negatif yang benar diprediksi termasuk ke dalam kelas sentimen negatif dan 6 data negatif yang diprediksi termasuk ke dalam kelas sentimen positif. Kemudian hasil evaluasi confusion matrix juga menghasilkan nilai evaluasi seperti f1 score, recall dan precision yang dapat dilihat pada Tabel 15.

Tabel 15. *Nilai Evaluasi*

Ukuran	Nilai
Recall	96,15%
Precision	86,21%
F1 Score	90,91%

Pada tabel 15 nilai evaluasi menghasilkan perhitungan nilai recall sebesar 96,15%, nilai precision 86,21% dan nilai f1 score 90.91%.

Pengujian menggunakan ulasan atau komentar yang diinput kedalam script yang akan menghasilkan prediksi sesuai dengan prediksi sistem.



Gambar 9. Pengujian Manual

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan maka diperoleh beberapa kesimpulan sebagai berikut:

1. Dari 200 sampel ulasan didapatakan gambaran data ulasan produk roundup memiliki data kelas positif sebanyak 165buah ulasan dan kelas negatif sebanyak 35 ulasan.
2. Hasil penerapan Support Vector Machine pada analisis sentiment produk roundup mendapatkan nilai evaluasi pada pembagian data latih & data uji 80:20 dengan niali akurasi 87,5% Presisi 86,21%, Recall 96,15% dan F1 Score sebesar 90.91%.

Berdasarkan kesimpulan di atas, berikut adalah beberapa saran yang dapat diberikan:

1. Pengembangan Model
   Untuk meningkatkan kinerja model, terutama dalam hal presisi dan akurasi, dapat dilakukan eksplorasi lebih lanjut terhadap parameter-parameter Support Vector Machine (SVM), seperti kernel, nilai C, dan gamma, serta mencoba algoritma lain untuk membandingkan hasilnya.
2. Peningkatan Kualitas Data
   Sebaiknya dilakukan evaluasi dan peningkatan kualitas data ulasan, seperti melakukan preprocessing yang lebih mendalam (misalnya, penghapusan stopwords, stemming, atau normalisasi teks). Hal ini dapat membantu mengurangi noise dalam data sehingga hasil analisis lebih akurat.

DAFTAR PUSTAKA

[1] Mukhroji, “Analisis Pengaruh Faktor-Faktor Positioning Produk Terhadap Citra Merek Pada Produk Roudup Biosorb 480 SL di Brebes Selatan,” *Jurnal Akuntansi*, vol. 11, no. 2, pp. 152–169, 2017.

[2] I. S. Yudis Bowo Pratama, “PEMANFAATAN INSTAGRAM SEBAGAI MEDIA PROMOSI (Studi Deskriptif Kualitatif Pada Akun Instagram @Kawanbaru.Co),” *Sustainability (Switzerland)*, vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI

[3] I. S. Joko Winahyu, “Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes,” *Kumpulan*

Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI), vol. 10, no. 2, p. 206, 2021, doi: 10.23887/karmapati.v10i2.36609.

[4] S. A. F. Clarisa Hasya Yutika, Adiwijaya, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.

[5] P. Arsi*1 and R. Waluyo2, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[6] Imam Fahrur Rozi, Sholeh Hadi Pramono, and Erfan Achmad Dahlan, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *Jurnal EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.

[7] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier," *Journal of Information System Research (JOSH)*, vol. 3, no. 4, pp. 524–531, 2022, doi: 10.47065/josh.v3i4.1944.

[8] R. A. Kurniawan, F. I. N. G, and M. R. Pribadi, "Sentimen Analisis Presiden Terpilih Menggunakan Algoritma Support Vector Machine," *The Indonesian Journal of Computer Science Research*, vol. 3, no. 14, pp. 68–73, 2024.

[9] H. S. Utama, D. Rosiyadi, B. S. Prakoso, and D. Ariadarma, "Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan Algoritma Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 243–250, 2019, doi: 10.29207/resti.v3i2.1050.

[10] J. F. Trevor Hastie, Robert Tibshirani, "The Elements of Statistical Learning Data Mining, Inference, and Prediction," *J Am Geriatr Soc*, vol. 32, no. 6, pp. 441–444, 1984, doi: 10.1111/j.1532-5415.1984.tb02220.x.

[11] H. S. Christopher D. Manning, Prabhakar Raghavan, *Introduction to information retrieval*, vol. 46, no. 05. 2009. doi: 10.5860/choice.46-2715.

[12] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *Ann Stat*, vol. 36, no. 3, pp. 1171–1220, 2008, doi: 10.1214/009053607000000677.

[13] S. Shalev-Shwartz and S. Ben-David, "Understanding machine learning: From theory to algorithms," *Understanding Machine Learning: From Theory to Algorithms*, vol. 9781107057, pp. 1–397, 2013, doi: 10.1017/CBO9781107298019.

[14] F. Fitriyani and T. Arifin, "Penerapan Word N-Gram Untuk Sentiment Analysis Review Menggunakan Metode Support Vector Machine (Studi Kasus: Aplikasi Sambara)," *Sistemasi*, vol. 9, no. 3, p. 610, 2020, doi: 10.32520/stmsi.v9i3.954.