Indonesian Journal of Computer Science
ISSN 2302-4364 (*print*) dan 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam, No. 1, Padang, Indonesia, Telp. (0751) 7056199, 7058325
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Terhadap Produk Skincare

Jasmarizal[1], Rahmaddeni[2], Junadhi[3], M. Khairul Anam[4]
jasmarizal@gmail.com[1], rahmaddeni@sar.ac.id[2], junadhi@sar.ac.id[3], khairulanam@sar.ac.id[4]
[1,2,3,4]STMIK Amik Riau, Teknik Informatika, Pekanbaru, Riau, Indonesia

Informasi Artikel

Diterima : 3 Jan 2024
Direview : 8 Jan 2024
Disetujui : 20 Feb 2024

Kata Kunci

Support Vector Machine, Analisis Sentimen, Skincare

Abstrak

Perawatan kulit telah menjadi aspek yang signifikan dalam pola hidup modern. Kesadaran masyarakat terhadap penampilan dan kesehatan kulit semakin meningkat, mendorong permintaan terus berkembang untuk produk skincare. Konsumen sering menghadapi kesulitan dalam memilih produk yang sesuai dengan jenis kulit mereka, di mana ulasan dari pengguna lain bisa menjadi panduan berharga, namun juga berpotensi menyebabkan kebingungan jika tidak dikelola dengan baik. Mengetahui sentimen konsumen terhadap produk skincare tidak hanya membantu produsen dan pengecer memahami penerimaan produk, tetapi juga memberikan arahan bagi konsumen lain dalam pengambilan keputusan. Kemajuan dalam teknologi analisis sentimen memungkinkan penelitian yang lebih efisien dan akurat terhadap pandangan konsumen mengenai produk skincare. Analisis sentimen dapat dijalankan secara otomatis menggunakan algoritma dan model kecerdasan buatan, di mana Support Vector Machine (SVM) menjadi salah satu metode yang efektif dalam permasalahan klasifikasi. SVM memberikan wawasan mendalam mengenai sentimen yang terkandung dalam ulasan konsumen. Dataset yang digunakan mengandung komentar dan ulasan dari pengguna terkait produk skincare MS Glow, dengan total 3.006 data. Proses selanjutnya melibatkan tahap pre-processing data, yang mencakup langkah-langkah seperti Case Folding, Normalisasi Data, Tokenisasi, Filtrasi Stop Words, dan Stemming. Pada tahap pemodelan, SVM digunakan untuk mengklasifikasi sentimen atau opini pengguna terhadap produk skincare tersebut. Hasil akhir menunjukkan bahwa model dengan ketidakseimbangan kelas mengalami overfitting, di mana performa model optimal hanya pada data pelatihan dan kurang efektif pada data uji. Namun, dengan melatih model menggunakan kelas yang seimbang dan menerapkan teknik SMOTE, ditemukan hasil optimal, mencapai akurasi sebesar 99.60% dan nilai f1-score sebesar 98.55%.

Indonesian Journal of Computer Science
ISSN 2302-4364 (*print*) dan 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam, No. 1, Padang, Indonesia, Telp. (0751) 7056199, 7058325
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Keywords

Support Vector Machine, Sentiment Analysis, Skincare

Abstract

Skin care has become a significant aspect of modern lifestyle. Public awareness of the appearance and health of skin is increasing, driving demand to continue to grow for skincare products. Consumers often face difficulties in choosing products that suit their skin type, where reviews from other users can be a valuable guide, but also have the potential to cause confusion if not managed properly. Knowing consumer sentiment towards skincare products not only helps manufacturers and retailers understand product acceptance, but also provides direction for other consumers in making decisions. Advances in sentiment analysis technology enable more efficient and accurate research into consumer views regarding skincare products. Sentiment analysis can be carried out automatically using algorithms and artificial intelligence models, where Support Vector Machine (SVM) is an effective method in classification problems. SVM provides deep insight into the sentiment contained in consumer reviews. The dataset used contains comments and reviews from users regarding MS Glow skincare products, with a total of 3,006 data. The next process involves the data pre-processing stage, which includes steps such as Case Folding, Data Normalization, Tokenization, Stop Words Filtration, and Stemming. At the modeling stage, SVM is used to classify user sentiment or opinions regarding the skincare product. The final results show that the model with class imbalance experiences overfitting, where the model performance is optimal only on the training data and is less effective on the test data. However, by training the model using balanced classes and applying the SMOTE technique, optimal results were found, achieving an accuracy of 99.60% and f1-score of 98.55%.

A. Pendahuluan

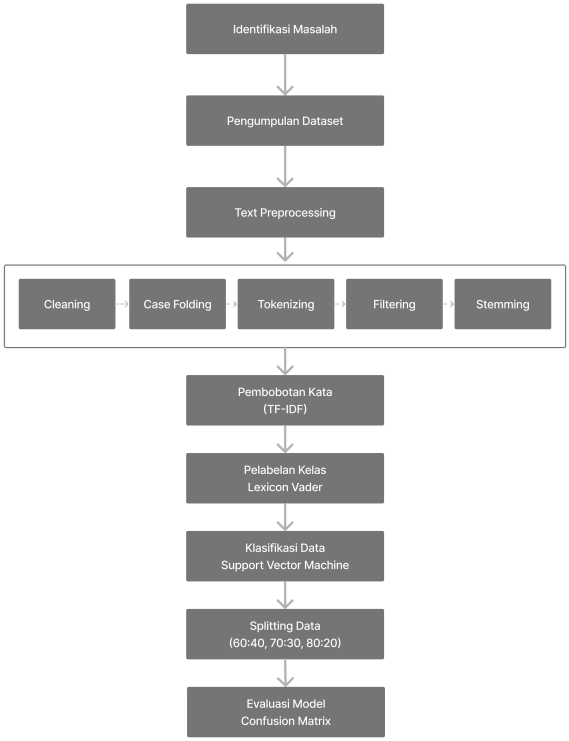
Indonesia sedang tren dalam perawatan kulit atau yang saat ini dikenal dengan sebutan skincare terus berkembang hingga saat ini. Perkembangan pada berbagai jenis perawatan kulit disebabkan oleh banyak faktor, diantaranya yaitu banyaknya permasalahan kulit yang muncul, perkembangan teknologi, banyaknya penelitian mengenai zat–zat tertentu, dan semakin banyak orang yang menyebarkan berbagai informasi mengenai pentingnya menjaga kesehatan kulit. Meskipun arti dari skincare secara umum yaitu perawatan kulit yang mencakup seluruh tubuh, namun kini skincare yang banyak digunakan dan dipahami masyarakat lebih banyak mengacu pada perawatan kulit wajah. Terdapat beberapa kriteria yang perlu diperhatikan dalam memilih produk skincare, hal pertama yang harus dilakukan yaitu mengetahui jenis kulit. Terdapat beberapa jenis kulit yaitu normal, berminyak, kering, kombinasi, maupun sensitif. Selain itu, pemilihan skincare juga didasarkan oleh permasalahan kulit yang dimiliki sehingga dapat menyesuaikan produk skincare yang akan digunakan. Jenis produk skincare yang digunakan secara umum yaitu sabun pembersih wajah (*facial wash*), toner, pelembab (*moisturizer*), serum, dan tabir surya (*sunscreen*).

Terdapat berbagai macam cara dalam melakukan perawatan kulit. Beberapa orang cenderung untuk berkonsultasi kepada dokter kulit atau berkunjung ke klinik kecantikan. Namun, dalam melakukan hal tersebut terdapat beberapa permasalahan yang dialami yaitu biaya yang relatif mahal, tidak ada nya klinik kecantikan di daerah tertentu. Permasalahan tersebut juga menjadi penyebab orang-orang mulai beralih untuk melakukan perawatan kulit secara mandiri dengan memilih produk skincare yang dijual di pasaran sebagai solusi untuk menyelesaikan permasalahan pada kulit wajahnya. Melihat adanya masalah pada klasifikasi produk perawatan kulit yang disebabkan banyaknya produk skincare di pasar dan peluang teknologi dalam membantu permasalahan tersebut maka solusi yang ditawarkan adalah membuat media informasi berupa sistem klasifikasi produk skincare.

Support Vector Machine merupakan salah satu metode terbaik yang bisa dipakai dalam permasalahan klasifikasi, konsep SVM bermula dari masalah klasifikasi dua kelas sehingga membutuhkan training set positif dan negatif [1]. Beberapa penelitian yang berkaiatan dengan metode Support Vector Machine seperti yang dilakukan oleh (Muhammad Avi Majid Kaaffah, dkk. 2020) dengan judul Sistem Klasifikasi Ukuran Baju Dengan Metode Support Vector Machine (SVM) yang menghasilkan akurasi algoritma SVM sebesar 92%. Pada penelitian [2] yang berjudul Klasifikasi Sentimen Ulasan Film Menggunakan Support Vector Machine, Information Gain, dan N-Grams. Algortima SVM menunjukkan nilai akurasi tertinggi sebesar 86% untuk unigram dan 76,4% untuk bigram pada klasifikasi SVM dengan menggunakan Information Gain sebagai seleksi fiturnya. Berikutnya penelitian yang dilakukan oleh [3] dengan judul Klasifikasi Review Produk Kecantikan Pada Aplikasi Sociolla Menggunakan Algoritme Modified K-Nearest Neighbor (MK-NN) dengan Pembobotan BM25. Hasil evaluasi pengujian dengan 5-fold cross validation dihasilkan rata-rata nilai akurasi, precision, recall, dan f-measure tertinggi sebesar 51,00%, 50,90%, 52,61%, dan 51,70% pada saat nilai k=11.

B. Metode Penelitian

Proses melakukan sebuah penelitian data dan informasi yang bersifat objektif yang akan digunakan sebagai titik acuan dalam penelitian, dengan adanya data-data tersebut di harapkan penelitian yang di hasilkan adalah penelitian yang berkualitas. Proses dalam melakukan penelitian ini digambarkan sebagai berikut [4]:



Gambar 1. Metodologi Penelitian

Tahapan yang dilakukan dapat dijelaskan sebagai berikut:

1. Identifikasi Masalah Pembuatan sistem ini diawali dengan menentukan masalah, dalam penelitian ini dapat diidentifikasi masalahnya adalah bagaimana cara melakukan analisis sentimen pada ulasan produk skincare menggunakan metode SVM dan berapa persentase ketepatan analisis sentimen menggunakan metode SVM dalam mengklasifikasi perawatan kulit berdasarkan ulasan di akun Shopee.
2. Pengumpulan Dataset
   Pada penelitian ini menggunakan data ulasan produk kecantikan jenis skincare yang diambil dari e-commerce Shopee. Data tersebut diperoleh dengan menggunakan teknik scraping dengan bantuan aplikasi bawaan Scraper dari Google Chrome. Data yang diambil berjumlah 1000 data review. Hasil dari crawling data disimpan dalam file bertipe .csv dan kemudian dilakukan labelling untuk menentukan pendapat atau pandangan dari review yang diambil. Pada proses pelabelan dibedakan menjadi 2 kelas, yaitu kelas positif dan kelas negatif.
3. *Text Processing*

Tahap Text Processing diartikan sebagai tahap awal untuk untuk membersihkan kata-kata yang tidak perlu atau kata-kata yang tidak memiliki makna. Proses keseluruhan dilakukan menggunakan program python3, sehingga dilakukan secara otomatis [5]. Preprocessing yang akan dilakukan mencakup berbagai aktivitas, diantaranya adalah:

A. *Cleaning*
   Menghapus karakter yang tidak diinginkan seperti tanda baca, url, emoji, angka, atau karakter tertentu dari teks [6].
B. *Casefolding*
   Mengubah teks menjadi huruf kecil atau huruf besar untuk mengurangi perbedaan antara huruf besar dan kecil dalam dokumen teks [7].
C. *Tokenization*
   Memecah teks menjadi kata-kata atau frase yang lebih kecil.
D. *Stopword Removal*
   Menghapus kata-kata yang tidak penting, seperti "apakah", "ke", "luar", dan "biasanya" [8].
E. *Filtering*
   Membuat data lebih bersih dan akurat sebelum digunakan dalam proses analisis, kata-kata yang tidak sesuai dengan topik yang dianalisis harus dihilangkan. [9].
F. *Stemming*
   Mengurangi variasi kata yang sama dengan menghapus imbuhan [10].

4. Pembobotan Kata TF-IDF
   Metode TF-IDF (*Terms Frequency-Inverse Document Frequency*) merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut [11].
5. Pelabelan Lexicon Vader
   Pendekatan berbasis leksikon (*Lexicon-based*) bergantung pada leksikon sentimen, yaitu kumpulan istilah sentimen yang diketahui dan dikompilasi sebelumnya. Sumber daya yang dapat digunakan pada analisis sentimen berbasis leksikon secara umum ada 2 jenis, yaitu berbasis kamus dan berbasis corpus yang menggunakan metode statistik atau semantik untuk menemukan polaritas sentiment. Penulis mengunakan sumber daya kamus pada penelitian ini, dengan cara memanfaatkan library TextBlob pada Pyhton yang menyediakan kamus berisi leksikon sentimen [12].
6. Klasifikasi Data
   Sebelum klasifikasi data dapat dilakukan, langkah-langkah seperti pembagian data, pembuatan algoritma SVM, dan pelatihan algoritma untuk melakukan evaluasi harus dilakukan. Dengan menggunakan splitting data, data dibagi menjadi dua bagian: data latihan (training) dan data uji (testing). Penulis melakukan empat percobaan dalam penelitian ini [13].
   A. Percobaan pertama, 60% data digunakan sebagai data latih dan 40% sebagai data uji.

B. Percobaan kedua, 70% data sebagai data latih dan 30% sebagai data uji.
C. Percobaan ketiga, 80% data sebagai data latih dan 20% sebagai data uji.

7. Evaluasi Model *Confusion Matrix*

Jumlah nilai diagonal *confusion matrix* dibagi dengan jumlah total data digunakan untuk menghitung nilai akurasi metode. Untuk mengetahui kinerja model, evaluasi dilakukan dengan melihat tabel akurasi dan presisi untuk masing-masing model. Setelah data uji diuji, akan dihasilkan daftar kelas-kelas dari data uji, sebut saja prediksi kelas. Setelah mengetahui hasil, selanjutnya dapat dilihat bagaimana metode klasifikasi bekerja untuk setiap kelas melalui nilai presisi, recall, dan nilai f1. Nilai presisi, recall, dan f1-nilai masing-masing memiliki nilai 0-1. Semakin tinggi nilainya, lebih baik [14].

C. Hasil dan Pembahasan

1. Pembahasan

Adapun proses yang dilakukan pada tahap implementasi penelitian yang dilakukan adalah sebagai berikut.

A. Import Library

Sebelum proses analisa data dilakukan maka terlebih dahulu dilakukan pengimporan library yang dibutuhkan saat proses analisis data. Pada gambar 2 disajikan kode program python.

```
# library yang dibutuhkan
import pandas as pd
import re
import string
from nltk.tokenize import word_tokenize
import demoji
import nltk
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.feature_extraction.text import CountVectorizer
nltk.download('words')
nltk.download('stopwords')
nltk.download('punkt')
pd.set_option('display.max_colwidth', 1)
```

Gambar 2. Import Library

B. Memuat Data

Selanjutnya dilakukan pembacaan dataset yang telah diperoleh melalui teknik crawling. Pada gambar 3 disajikan kode program python.

```
# memuat data
data = pd.read_csv('dataset/msglow_shopee.csv')
# pustaka data untuk label
lex_pos = pd.read_csv('dataset/text-library/lexicon_positive_ver1.csv')
lex_neg = pd.read_csv('dataset/text-library/lexicon_negative_ver1.csv')
# data key normalization
key_norm = pd.read_csv('dataset/text-library/key_norm.csv')
```

```
# menampilkan 5 data pertama
data.head()
```

Gambar 3. Memuat Data

Setelah kode program dijalankan, maka hasinya seperti terlihat pada gambar 4 dibawah ini.

	web-scrapper-order	web-scrapper-start-url	review	date
0	1703673315-1	https://shopee.co.id/MS-GLOW-Paket-Wajah-i.17566419.5417344403?sp_atk=8a14f9d8-c119-4029-aadc-10127ed44b74&xptdk=8a14f9d8-c119-4029-aadc-10127ed44b74	Kemasan BB cream sama Day cream yg sblumnya sama aja gak ada bedanya aku kira salah tpi dalamnya beda \nUntuk yg lainnya oke\nPengiriman cepat cman 1 hari pakai sicepat 🥰🥰🥰😘😘\nThanks MSglow	2021-11-26 15:04
1	1703673315-2	https://shopee.co.id/MS-GLOW-Paket-Wajah-i.17566419.5417344403?sp_atk=8a14f9d8-c119-4029-aadc-10127ed44b74&xptdk=8a14f9d8-c119-4029-aadc-10127ed44b74	Alhamdulillah baru banget sampe nih barang nya. Dalem nya aman buble nya banyak ga ada yang cacat sesuai baru order di toko ini sih tapi Alhamdulillah packing an nya oke. Pengiriman nya oke. Semoga cocok deh	2022-05-07 09:53
2	1703673315-3	https://shopee.co.id/MS-GLOW-Paket-Wajah-i.17566419.5417344403?sp_atk=8a14f9d8-c119-4029-aadc-10127ed44b74&xptdk=8a14f9d8-c119-4029-aadc-10127ed44b74	Kualitas produk sangat baik. Harga produk sangat baik. Kecepatan pengiriman sangat baik. Kecepatan pengiriman sangat baik. Respon penjual sangat baik. Semua baik, bahkan orang jahatpun aslinya baik, hanya saja mungkin lingkungan dan kondisilah yg merubah mereka menjadi tidak baik...	2023-07-11 17:35
3	1703673315-4	https://shopee.co.id/MS-GLOW-Paket-Wajah-i.17566419.5417344403?sp_atk=8a14f9d8-c119-4029-aadc-10127ed44b74&xptdk=8a14f9d8-c119-4029-aadc-10127ed44b74	Produk original , packing rapih double2 babble warp ,pengiriman cepat ,pengemasan cepat... Trim's MSGlow Official 😘	2021-10-29 06:09
4	1703673315-5	https://shopee.co.id/MS-GLOW-Paket-Wajah-i.17566419.5417344403?sp_atk=8a14f9d8-c119-4029-aadc-10127ed44b74&xptdk=8a14f9d8-c119-4029-aadc-10127ed44b74	Pengemasan cepat, packing aman. Produk ori, expired date masih lama. Tekstur kental, tp akan berasa watery begitu di aplikasikan ke wajah, cepat meresap. First time coba, semoga cocok & bisa terus pake.	2021-04-22 17:14

Gambar 4. Hasil Pemuatan Dataset

C. Cleaning Data

Proses yang dilakukan berikutnya adalah cleaning data, dimana tujuan dari proses cleaning data adalah menghapus data yang duplikasi, kata imbuhan, simbol, karakter dan data yang tidak butuhkan. Proses cleaning data yang dilakukan di antaranya case folding, tokenizing, stop word, dan stemming. Pada gambar 5 dibawah ini ditampilkan kode program pythonnya.

```
#Merubah semua huruf menjadi huruf kecil
#Menghilangkan semua tanda baca, angka, dan simbol
def casefolding(text):
    text = text.lower() # mengubah kalimat menjadi huruf kecil
    text = re.sub(r'\w+:\/{2}[\d\w-]+(\.[\d\w-]+)*(?:(?:\/[^\s/]*))*', '', text)
    text = re.sub(r'[?|$|.|@#%^&*=!_:")(-+,]', '', text)
    text = text.replace('\n', '') # menghilangkan simbol utk baris baru
    text = demoji.replace(text, '')
    text = text.strip(" ") # hapus spasi dari kiri dan kanan teks

    return text
```

```
# menerapkan fungsi
data['review'] = data['review'].apply(casefolding)
data.head()
```

Gambar 5. Proses Case Folding

Setelah proses case folding diatas dilakukan maka, hasil dari programnya disajikan pada gambar 6 dibawah ini.

	review	years	month
0	kemasan bb cream sama day cream yg sblumnya sama aja gak ada bedanya aku kira salah tpi dalamnya beda untuk yg lainnya okepengiriman cepat cman 1 hari pakai sicepat thanks msglow	2021	11
1	alhamdulillah baru banget sampe nih barang nya dalem nya aman buble nya banyak ga ada yang cacat sesuai baru order di toko ini sih tapi alhamdulillah packing an nya oke pengiriman nya oke semoga cocok deh	2022	5
2	kualitas produk sangat baik harga produk sangat baik kecepatan pengiriman sangat baik kecepatan pengiriman sangat baik respon penjual sangat baik semua baik bahkan orang jahatpun aslinya baik hanya saja mungkin lingkungan dan kondisilah yg merubah mereka menjadi tidak baik	2023	7
3	produk original packing rapih double2 babble warp pengiriman cepat pengemasan cepat trim's msglow official	2021	10
4	pengemasan cepat packing aman produk ori expired date masih lama tekstur kental tp akan berasa watery begitu di aplikasikan ke wajah cepat meresap first time coba semoga cocok bisa terus pake	2021	4

Gambar 6. Hasil Case Folding

D. Pelabelan Data

Pelabelan data merupakan proses untuk mengkategorikan kumpulan data ke dalam kelompok positif, negatif dan netral menggunakan teknik lexicon vader. Pada gambar 7 berikut ini disajikan kode program pythonnya.

```
# membaca file lexicon positive dan negative untuk pelabelan
lexicon_positive = dict()
with open('dataset/text-library/lexicon_positive_ver1.csv', 'r') as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    for row in reader:
        lexicon_positive[row[0]] = int(row[1])

lexicon_negative = dict()
with open('dataset/text-library/lexicon_negative_ver1.csv', 'r') as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    for row in reader:
        lexicon_negative[row[0]] = int(row[1])

def sentiment_analysis_lexicon_indonesia(text):
  score = 0
  for word_pos in text:
    if (word_pos in lexicon_positive):
      score = score + lexicon_positive[word_pos]
  for word_neg in text:
    if (word_neg in lexicon_negative):
      score = score + lexicon_negative[word_neg]
  Sentimen=''
  if (score > 0):
    Sentimen = 'Positive'
  elif (score < 0):
    Sentimen = 'Negative'
  else:
    Sentimen = 'Neutral'

  return score, Sentimen
```

Gambar 7. Proses Lexicon Vader

Setelah proses diatas dilakukan maka, hasil pelabelan data yang dihasilkan disajikan pada gambar 8 dibawah ini.

	review	years	month	score	sentimen
0	kemas beda salah dalam beda kirim cepat pakai sicepat terimakasih msglow	2021	11	3	Positive
1	alhamdulillah banget barang dalem aman cacat sesuai alhamdulillah packing oke kirim oke moga cocok	2022	5	15	Positive
2	kualitas produk harga produk cepat kirim cepat kirim respon jual jahat asli lingkung kondisi rubah	2023	7	18	Positive
3	produk packing rapih kirim cepat emas cepat terimakasih msglow	2021	10	10	Positive
4	emas cepat packing aman produk expired tekstur kental aplikasi wajah cepat resap coba moga cocok pakai	2021	4	2	Positive

Gambar 8. Hasil Lexicon Vader

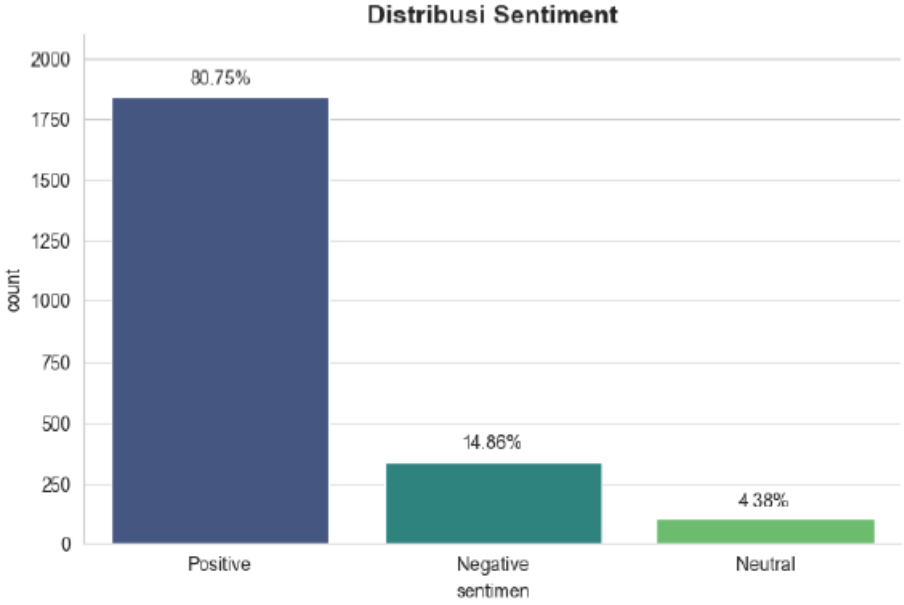
E. Visualisasi Distribusi Sentimen

Selanjutnya dilakukan proses visualisasi untuk melihat berapa jumlah label yang memiliki data positif, negatif dan netral. Pada gambar 9 dibawah ini kode program python.

```
# visualisasi distribusi jumlah sentimen terhadap produk skincare
plt.figure(figsize=(8,5))

sns.set_style('whitegrid')
ax = sns.countplot(data=df, x='sentimen', palette='viridis')
sns.despine(ax=ax, top=True, right=True)

for bar in ax.patches:
    height = bar.get_height()
    ax.annotate(f'{height / float(len(df)) * 100:.2f}%',
                (bar.get_x() + bar.get_width() / 2, height),
                ha='center', va='center',
                size=10,
                xytext=(0, 10),
                textcoords='offset points',)

plt.title('Distribusi Sentiment', size=14, fontweight='bold')
plt.ylim(0, 2100)
plt.show()
```

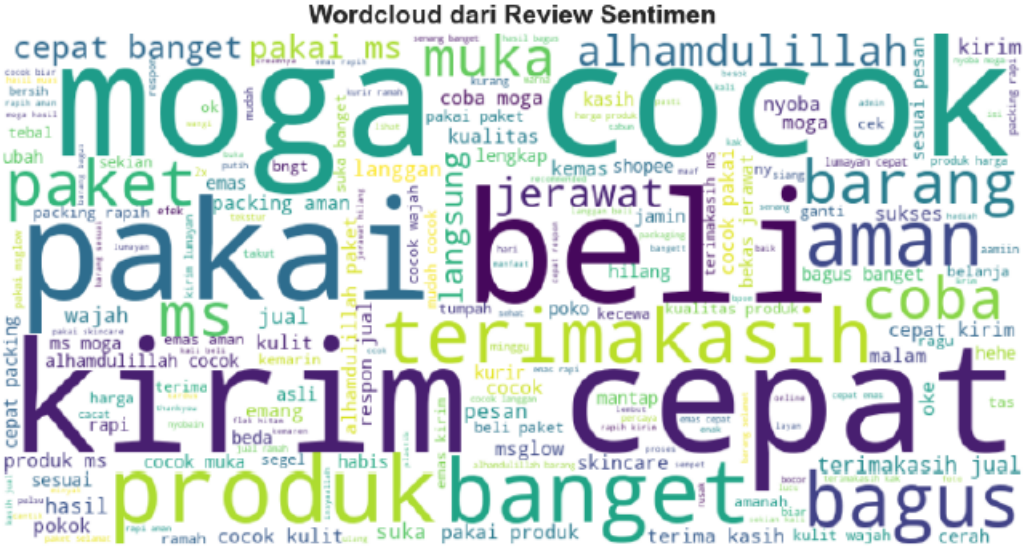
Gambar 9. Visualisasi Distribusi Sentimen

Dari hasil kode program tersebut maka dihasilkan visualisasinya seperti disajikan pada gambar 10. Dari dataset yang berjumlah 2000 maka diperoleh label yang memiliki positif berjumlah 80.75%, negatif 14,86% dan netral 4,38%.



Gambar 10. Visualisasi Distribusi Sentimen

Prose berikutnya dilakukan visualisasi dengan menampilkan kata yang sering muncul pada review tersebut. Pada gambar 11 terlihat bahwa kata yang sering muncul adalah moga, cocok, pakai, beli, kirim, cepat, produk, bagus, banget dan terimakasih.



Gambar 11. Wordcloud Review Sentimen

2. Hasil

A. Implementasi Model SVM

Proses berikutnya yang dilakukan adalah implementasi model SVM dengan kode program disajikan pada gambar 12 dibawah ini.

```
# inisialisasi train_size
train_sizes = np.arange(0.5, 0.91, 0.1)

# list untuk menyimpan performa model
train_acc = []
test_acc = []
model_name = []
trains_size = []
train_f1scores = []
test_f1scores = []

for size in train_sizes:
    # splitting data
    features_train, features_test, target_train, target_test = train_test_split(x_tf_idf, target, stratify=target, train_size=size, random_state=12345)

    # Inisialisasi model SVM
    model = SVC(kernel='linear', random_state=12345)
    name = model.__class__.__name__
    #
    model.fit(features_train, target_train)
    #
    target_pred_train = model.predict(features_train)
    target_pred_test = model.predict(features_test)
    # f1_score
    f1scores_test = f1_score(target_test, target_pred_test, average='macro')
    f1scores_train = f1_score(target_train, target_pred_train, average='macro')
    # accuracy
    accuracy_train = accuracy_score(target_train, target_pred_train)
    accuracy_test = accuracy_score(target_test, target_pred_test)
    # menghitung rata-rata akurasi dari cross-validation
    train_acc.append(accuracy_train)
    test_acc.append(accuracy_test)
    model_name.append(name)
    trains_size.append(f'{size:.0%}')
    train_f1scores.append(f1scores_train)
    test_f1scores.append(f1scores_test)
```

Gambar 12. Model SVM

Selanjutnya hasil dari kode program pada gambar 13 diatas adalah sebagai berikut.

	model_name	train_size	train_accuracy	train_f1-score	test_accuracy	test_f1-score
0	SVC	50%	0.950877	0.753377	0.838738	0.438452
1	SVC	60%	0.954678	0.763664	0.837897	0.435400
2	SVC	70%	0.954887	0.763078	0.855474	0.476391
3	SVC	80%	0.958882	0.791518	0.859956	0.487305
4	SVC	90%	0.957602	0.780969	0.886463	0.542432

Gambar 13. Hasil Model SVM

B. Pengujian

Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi. Evaluasi dengan confusion matrix menghasilkan nilai accuration, precision, recall dan f-measure. Pada gambar 14 dibawah ini disajikan kode program python dari proses pengujian confusion matrix.

```
# cclassification report untuk balanced class dengan smote
for size in train_sizes:
        # splitting data
        features_train, features_test, target_train, target_test = train_test_split(X_resampled, y_resampled, stratify=y_resampled, train_size=size, random_state=12345)

        # Inisialisasi model SVM
        model = SVC(kernel='linear', random_state=12345)
        #
        model.fit(features_train, target_train)
        #
        target_pred_test = model.predict(features_test)

        print(f'Data Training {size:.0%}'.center(55, '-'))
        print(classification_report(target_test, target_pred_test))
        print(55*'=')
```

Gambar 14. Pengujian Model SVM dengan Confusion Matrix

Dari proses confusion matrix diatas maka, diperoleh hasilnya seperti pada gambar 15 dibawah ini. Proses pengujian data yang dilakukan

menggunakan beberapa kombinasi yaitu 60:10, 70:30, 80:20, dan 90:10. Dari proses pengujian yang dilakukan maka diperoleh hasil akurasi yang tertinggi adalah 99% dengan kombinasi data traning dan testing 90:10.

```
======================================================
-------------------Data Training 60%-------------------
              precision    recall  f1-score   support

    Negative       0.95      1.00      0.97       737
     Neutral       0.98      1.00      0.99       737
    Positive       1.00      0.93      0.96       737

    accuracy                           0.97      2211
   macro avg       0.98      0.97      0.97      2211
weighted avg       0.98      0.97      0.97      2211

======================================================
-------------------Data Training 70%-------------------
              precision    recall  f1-score   support

    Negative       0.95      0.99      0.97       552
     Neutral       0.99      1.00      0.99       553
    Positive       0.99      0.94      0.97       553

    accuracy                           0.98      1658
   macro avg       0.98      0.98      0.98      1658
weighted avg       0.98      0.98      0.98      1658

======================================================
-------------------Data Training 80%-------------------
              precision    recall  f1-score   support

    Negative       0.96      1.00      0.98       368
     Neutral       0.98      1.00      0.99       369
    Positive       1.00      0.95      0.97       369

    accuracy                           0.98      1106
   macro avg       0.98      0.98      0.98      1106
weighted avg       0.98      0.98      0.98      1106

======================================================
-------------------Data Training 90%-------------------
              precision    recall  f1-score   support

    Negative       0.97      1.00      0.99       184
     Neutral       0.98      1.00      0.99       184
    Positive       1.00      0.96      0.98       185

    accuracy                           0.99       553
   macro avg       0.99      0.99      0.99       553
weighted avg       0.99      0.99      0.99       553

======================================================
```

Gambar 15. Hasil Pengujian SVM dengan Confusion Matrix

D. Simpulan

Dataset komentar atau review pengguna atau pelanggan terhadap produk skincare ms glow berjumlah sebanyak 3.006. Akan tetapi dataset ini terdapat *missing value* dan dilakukan proses penghapusan karena data tersebut merupakan data kualitatif yang berisi komentar dari setiap pelanggan. Kemudian dilakukan tahap data pre-processing, dimana meliputi langkah-langkah berikut ini: Case Folding Data Normalization Tokenizing Filtering Stop Words, dan Stemming. Kemudian dilakukan modeling, algoritma yang digunakan adalah Support Vector Machine untuk mengklasifikasi sentimen atau opini pengguna skincare. Hasil akhir yang didapatkan setelah melatih model yang imbalance class dan balanced class mendapati bahwa model dengan imbalance class mengalami over fitting dimana artinya performa model hanya bagus untuk data training dan untuk data latih

model tidak bekerja dengan baik. Setelah kita melatih model dengan balanced class dengan teknik SMOTE kita mendapati hasil model yang maksimal, dimana nilai akurasi sebanyak 99.60% dan nilai f1-score sebanyak 98.55%.

E. Ucapan Terima Kasih

Terima Kasih penulis sampaikan kepada Kampus STMIK Amik Riau yang telah meluangkan waktunya untuk membantu kami dalam mensukseskan penelitian ini serta lainnya yang ikut terlibat membantu dalam penelitian ini.

F. Referensi

[1] Pratama, A., Cahya, R. & Ratnawati, D. E., 2017. Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. Pengembangan Teknologi Informasi dan Ilmu Komputer, Volume 2, pp. 1704- 1708.

[2] Rizky Hilman Faturrahman, dkk. 2022. Klasifikasi Sentimen Ulasan Film Menggunakan Support Vector Machine, Information Gain, dan N-Grams. Vol.9, No.3 pp. 1928-1933

[3] Alfita Nuriza, 2020. Klasifikasi ReviewProduk Kecantikan Pada Aplikasi Sociolla Menggunakan AlgoritmeModified K-Nearest Neighbor (MK-NN)dengan Pembobotan BM25. Vol. 4, No. 10, Oktober 2020, hlm. 3426-3431

[4] Muhammad Muadin, dkk. 2023. Implementasi Metode Support Vector Machine Pada Opinion Mining Masyarakat Terkait ChatGPT. Vol. 7, No.1, Juni 2023, Hlm 78-84

[5] Junadhi, Agustin, Rifqi, M., & Anam, M. K. (2022). Sentiment Analysis of Online Lectures using K-Nearest Neighbors based on Feature Selection. Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), 11(3), 216–225. https://doi.org/10.23887/janapati.v11i3.51531

[6] Alhaq, Z., Mustopa, A., & Santoso, J. D. (2021). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter.

[7] Dwi Cahyo, M. P., Widodo, & Prasetya Adhi, B. (2019). Kinerja Algoritma Support Vector Machine dalam MenentukanKebenaran Informasi Banjir di Twitter. PINTER : Jurnal Pendidikan Teknik Informatika Dan Komputer, 3(2), 116–121. https://doi.org/10.21009/pinter.3.2.5

[8] Fitriyana, V., Hakim, L., Candra Rini Novitasari, D., Hanif Asyhar, A., Studi Matematika, P., Sains Dan Teknologi, F., Sunan Ampel Surabaya, U., & Timur, J. (2023). Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine. In Jurnal Buana Informatika (Vol. 14, Issue 1).

[9] R. Maulana, P. A. Rahayuningsih, W. Irmayani, D. Saputra, and W. E. Jayanti, "Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain," J. Phys. Conf. Ser., vol. 1641, no. 1, pp. 0–6, 2020, doi: 10.1088/1742-6596/1641/1/012060.

[10] N. Octaviani Faomasi Daeli, "Sentiment Analysis on Movie Reviews Using Information Gain and KNearest Neighbor," J. Data Sci. Its Appl., 2020.

[11] O. M. Bafadhal dan A. D. Santoso, "Memetakan Pesan Hoaks Berita Covid-19 Di Indonesia Lintas Kategori, Sumber, Dan Jenis Disinformasi," Bricol. J.

Magister Ilmu Komun., vol. 6, no. 02, hal. 235, 2020, doi: 10.30813/bricolage.v6i02.2148.

[12] E. Mailoa, "Analisis Node dengan Centrality dan Follower Rank pada Twitter," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 4, no. 5, hal. 937–942, 2020, doi: 10.29207/resti.v4i5.2398.

[13] M. Hanafiah, A. Herdiani, W. Astuti, dan M. Kom, "Klasifikasi Spam Tweet Pada Twitter Menggunakan Metode Naïve Bayes (Studi Kasus : Pemilihan Presiden 2019)," in e-Proceeding of Engineering, 2019, vol. 6, no. 2, hal. 9111–9120.

[14] S. Anggelia dan A. Syaifudin, "SENTIMEN WARGANET MAHASISWA TERHADAP COVID-19," LITERASI, vol. 5, hal. 49–57, 2021, doi: http://dx.doi.org/10.25157/literasi.v5i1.5149.