

Article history

Received Jan 11, 2024

Accepted Dec 29, 2024

Published Dec 29, 2024

PENERAPAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK DALAM ANALISIS SENTIMEN PENGARUH BRAND IMAGE DAN LABEL HARGA (STUDI ANALISIS: PRODUK SKINCARE SKINTIFIC)

Ekarini Lathifah¹, Aditya Dwi Putro Wicaksono¹, Andreas Rony Wijaya¹

¹Teknik Informatika, Institut Teknologi Telkom Purwokerto

email: 20102096@ittelkom-pwt.ac.id, aditya@ittelkom-pwt.ac.id, andreas@ittelkom-pwt.ac.id

Abstract

There are various kinds of products that are included in cosmetic products, namely personal care, make up, fragrance including perfume, hair care, and skincare. Skincare has become one of the primary needs for women in Indonesia today, because skincare can maintain healthy skin. Skincare is a beauty product that is used by users to clean dirt on the face. In deciding to choose skincare products, of course, consumers are influenced by various factors such as skincare quality, brand image, price, and others. In addition, reviews of skincare products are also important as an effort by cosmetic companies to attract consumers' buying interest. One method in deep learning to analyze is the Convolutional Neural Network (CNN). Sentiment analysis is carried out as an effort to evaluate and determine consumer satisfaction with skincare products as well as materials for service improvement. This study uses the CNN method which in this model has several stages such as data scraping, data preprocessing which consists of data cleansing & case folding, stemming, tokenizing, filtering (stopword removal), labeling process, modeling, and model evaluation. In this study, the data used is scraping data on the Tokopedia website about skincare skintific products. The data will be processed with the CNN model to obtain an accuracy value resulting from the performance of the model.

Keywords: Analysis, Algorithm, Brand Image, CNN, Sentiment, Skincare.

Abstrak

Ada berbagai macam produk yang termasuk kedalam produk kosmetik yaitu *personal care, make up, pewangi* termasuk *parfum, hair care, dan skincare*. *Skincare* sudah menjadi salah satu kebutuhan primer bagi perempuan di Indonesia saat ini, karena dengan *skincare* dapat menjaga kesehatan kulit. *Skincare* merupakan produk kecantikan yang digunakan oleh penggunanya untuk membersihkan kotoran pada wajah. Dalam memutuskan memilih produk *skincare*, tentunya konsumen dipengaruhi oleh berbagai faktor seperti kualitas *skincare, brand image, dan harga*. Selain itu, *review* mengenai produk *skincare* juga penting sebagai upaya perusahaan kosmetik untuk menarik minat beli konsumen. Salah satu metode dalam *deep learning* untuk menganalisis adalah *Convolutional Neural Network (CNN)*. Sentimen analisis dilakukan sebagai upaya dalam mengevaluasi dan mengetahui kepuasan konsumen terhadap produk *skincare* serta sebagai bahan peningkatan pelayanan. Penelitian ini menggunakan metode CNN dimana pada model ini memiliki beberapa tahapan seperti *scraping data, preprocessing data* yang terdiri dari *data cleansing & case folding, stemming, tokenizing, filtering (stopword removal)*, proses *labelling, modelling, dan evaluasi model*. Pada penelitian ini data yang digunakan merupakan data hasil *scraping* pada website tokopedia tentang produk *skincare skintific*. Data tersebut akan diolah dengan model CNN untuk memperoleh nilai akurasi yang dihasilkan dari kinerja model tersebut.

Kata Kunci: Analisis, Algoritma, Brand Image, CNN, Sentimen, Skincare.

1. PENDAHULUAN

Perusahaan kosmetik di Indonesia ada lebih dari 760 perusahaan, hal tersebut menunjukkan bahwa Indonesia merupakan pangsa besar bagi perusahaan kosmetik dalam memasarkan produknya [1]. Hal tersebut juga dibuktikan dengan semakin bertumbuhnya pendapatan pasar perawatan kulit Indonesia yang mencapai US\$ 1,991,7 juta pada tahun 2020 dan diperkirakan akan meningkat dalam dari 2020 hingga 2025 [2]. Dalam memasarkan produknya, industri kecantikan memerlukan media sosial sebagai sarana dalam mempromosikannya untuk dapat memikat konsumen [3]. Ada berbagai macam produk yang termasuk kedalam produk kosmetik yaitu produk perawatan pribadi, makeup, pewangi seperti parfum, hair care, dan skincare [1]. Skincare menjadi salah satu kebutuhan primer bagi perempuan di Indonesia saat ini, karena dengan skincare dapat menjaga kesehatan kulit [4].

Menurut Nilforoushzadeh dalam [5] skincare merupakan produk kecantikan yang digunakan oleh penggunaannya untuk membersihkan kotoran pada wajah. Penggunaan skincare dapat membantu seseorang untuk merawat wajah mereka agar tidak muncul jerawat, kulit kering, kemerahan, dan lain-lain. Jenis produk skincare yang beredar di pasaran banyak macamnya, seperti scarlett, avoskin, wardah, white lab, skintific, dan sebagainya. Berdasarkan salah satu artikel online yang dimuat oleh [compas.co.id](https://www.compas.co.id), Skintific berhasil masuk ke top 7 brand pada e-commerce Tokopedia dan Shopee periode 1 Mei 2023 – 5 Mei 2023. Hal tersebut membuktikan bahwa produk skincare Skintific cukup diminati oleh konsumen. Selain itu, dalam artikel [compas](https://www.compas.co.id) lainnya juga menyebutkan bahwa Skintific menjadi fenomenal dengan inovasi produk 5x Ceramide-nya. Hasilnya, banyak konsumen yang penasaran dan tergiur untuk mencobanya.

Dalam keputusannya untuk membeli sebuah produk skincare, tentunya konsumen dipengaruhi oleh berbagai faktor seperti kualitas skincare, brand image, harga [6]. Kuatnya brand image mampu membuat pelanggan yakin terhadap kualitas dari produk yang mereka beli dan tentunya akan memberikan informasi tersebut melalui review yang mereka tulis di internet [7]. Selain brand image, harga juga menjadi salah satu faktor bagi konsumen untuk memutuskan membeli suatu produk. Pengaruh harga terhadap

keputusan pembelian sangat penting, karena tingkat harga yang ditetapkan oleh perusahaan dapat menjadi tolak ukur permintaan suatu produk [8].

Berdasarkan permasalahan tersebut, metode untuk mengatasi permasalahan yang ada yaitu dengan analisis sentimen. Analisis Sentimen dilakukan sebagai upaya dalam meningkatkan kepuasan konsumen terhadap produk skincare serta sebagai bahan evaluasi untuk meningkatkan pelayanan [9]. Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis dengan mengelompokkan review pengguna menjadi opini positif, netral, atau negatif [10]. Oleh sebab itu, diperlukan sebuah program agar dapat melakukan analisis secara otomatis, yaitu dengan menggunakan algoritma Convolutional Neural Network [11]. Penggunaan algoritma ini karena algoritma CNN lebih mendominasi model secara luas dalam pengklasifikasian sebuah teks. Terdapat algoritma Convolutional Neural Network yang sangat populer yaitu algoritma CNN berbasis kata atau berbasis karakter, dimana keduanya menunjukkan hasil dari beberapa klasifikasi teks [12].

Beberapa penelitian terkait penggunaan algoritma CNN untuk analisis juga telah dilakukan oleh beberapa peneliti seperti penelitian tentang review customer pengguna indihome dan first media pada data twitter dengan model CNN yang dilakukan oleh Saleh Hasan Badjrie dkk pada tahun 2021 menyebutkan bahwa hasil analisis pada review terhadap produk IndiHome dan FirstMedia dengan algoritma CNN mendapatkan nilai tertinggi pada akurasi 98% dan 91%, dimana hasil tersebut membuktikan bahwa algoritma ini bekerja dengan efektif [12].

Penelitian lainnya dengan algoritma CNN untuk mengatuh kebijakan vaksinasi pada masyarakat yang dilakukan oleh Fany Alifian Irawan dan Dwi Anindyani Rochmah pada tahun 2022 memperoleh hasil rata-rata dari akurasi yaitu 98,66%. Sedangkan nilai rata-rata untuk precision yaitu 98.33%, recall 98.33% dan skor untuk f1-score 98.66%. Selain itu, evaluasi dengan K-Fold Cross Validation menambah keakuratan sentimen dengan perolehan nilai akurasi training terbaik yaitu 100% dan nilai akurasi testing terbaik yaitu 99.61% [9]. Penelitian yang dilakukan oleh Sukma Nindi Listyarini dan Dimas Aryo Anggoro pada tahun 2021 untuk memperoleh hasil sentimen terhadap pilkada yang diadakan di

tengah pandemi covid-19 dengan algoritma CNN didapatkan hasil model terbaik yaitu 90% dengan 4-layer convolutional dan 100 epoch [13].

Pada penelitian ini, peneliti bermaksud untuk menyusun Tugas Akhir dengan judul “Penerapan Algoritma Convolutional Neural Network (CNN) Dalam Analisis Sentimen Pengaruh Brand Image Dan Label Harga (Studi Analisis: Produk Skincare Skintific)”, dimana dalam penelitian ini bermaksud untuk memperoleh hasil analisis terhadap produk skincare skintific dengan faktor brand image dan label harga dengan mengimplementasikannya dengan menggunakan algoritma Convolutional Neural Network (CNN) guna memperoleh hasil akurasi terhadap review pengguna skincare skintific pada website Tokopedia.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini yaitu dengan algoritma *Convolutional Neural Network* (CNN). Rancangan penelitian dapat dilihat pada diagram alir penelitian yang disajikan pada gambar 2.1.

Subjek dan Objek Penelitian

Subjek Penelitian

Subjek penelitian yang digunakan adalah data ulasan produk skincare skintific yang diperoleh dari website Tokopedia sebanyak 500 data.

Objek Penelitian

Objek penelitian yang digunakan adalah algoritma CNN yang digunakan untuk analisis sentimen pada produk skincare skintific.

Alat dan Bahan

Alat dan Bahan yang digunakan dalam penelitian ini diantaranya yaitu :

Alat

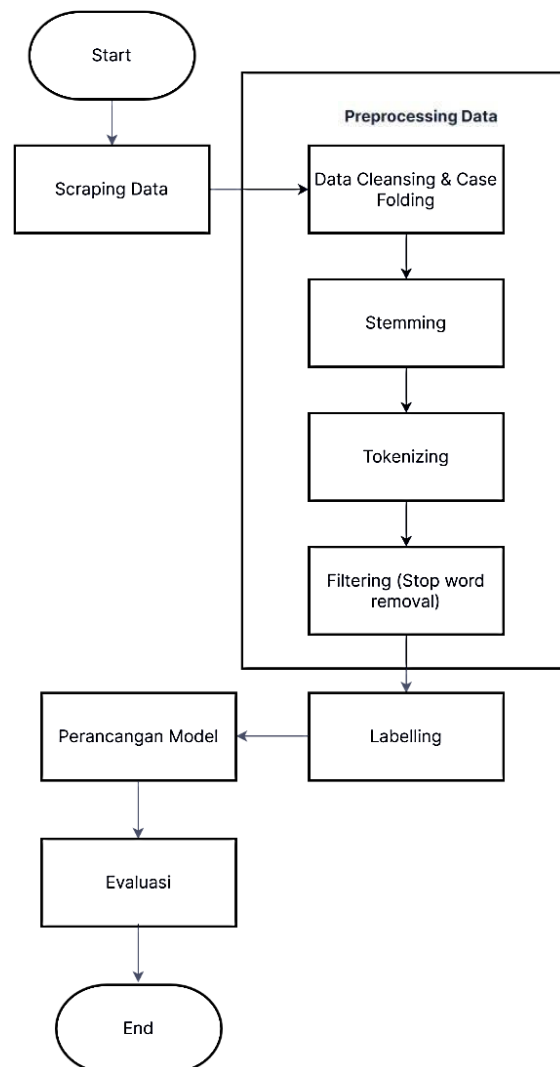
Alat yang digunakan dalam penelitian ini terdiri dari dua jenis yaitu perangkat keras (hardware) dan perangkat lunak (software), adapun perinciannya sebagai berikut :

1. Perangkat Keras (Hardware)
 - a. Device : LAPTOP-DTMHT7M3
 - b. Processor : AMD Ryzen 5 5500U with Radeon Graphics 10 GHz
 - c. RAM : 8GB
2. Perangkat Lunak (Software)

- a. Sistem Operasi : Windows 11 Home Single Language
- b. Bahasa Pemrograman : Python
- c. Aplikasi : Google Colab

Diagram Alir Penelitian

Rancangan penelitian dalam Penyusunan Tugas akhir ini dapat dilihat pada diagram alir penelitian yang disajikan pada gambar 1.



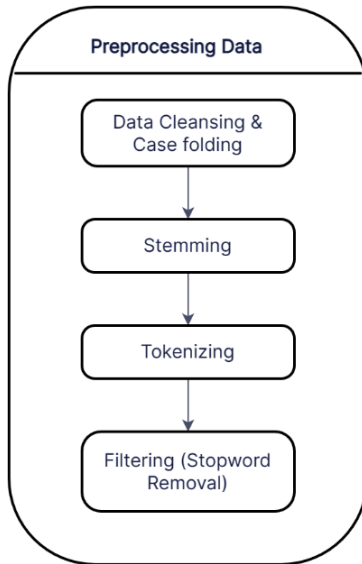
Gambar 1. Diagram Alir Penelitian

Scraping Data

Scraping data diperoleh dari review produk skintific pada website tokopedia. Hasil dari scraping tersebut nantinya disimpan ke dalam file csv.

Preprocessing Data

Dataset tersebut kemudian dilakukan proses preprocessing untuk menangani noise yang ada pada data sehingga data yang dihasilkan menjadi terstruktur. Berikut flowchart pada preprocessing data pada gambar 2.



Gambar 2. Flowchart Preprocessing Data

Case Folding

Pada tahap ini, dilakukan proses mengubah kalimat menjadi huruf kecil pada semua data.

Data Cleansing

Pada proses ini, dilakukan pembersihan data yang dapat mengganggu hasil analisis seperti karakter berulang, URL, nama pengguna, tagar, simbol, angka, dan spasi tambahan, serta tanda baca.

Stemming

Stemming merupakan proses menghilangkan afiks seperti prefiks, konfiks, dan sufiks untuk mengubah kata berimbuhan menjadi ke bentuk akarnya.

Tokenizing

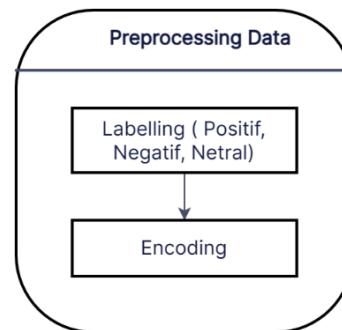
Tokenisasi merupakan proses pemecahan kalimat menjadi kata-kata sesuai dengan aturan bahasa.

Filtering (Stopword Removal)

Prosedur yang dikenal sebagai Stopwords Removal digunakan untuk menghilangkan katakata yang tidak berarti apa-apa.

Labelling

Tahap selanjutnya yaitu proses labelling. Labelling dilakukan untuk memperoleh nilai sentimen dari review produk oleh konsumen terhadap produk skincare skintific. Proses labelling dapat dilihat pada gambar 3.



Gambar 3. Flowchart Labelling Data

Perancangan Model

Model yang digunakan pada penelitian ini adalah model Algoritma *Convolutional Neural Network*.

Pelatihan dan Pengujian Model

Pengujian model dilakukan dengan matriks konvolusi, seperti precision, recall, f1-score dan akurasi. Pada confusion matrix terdapat 4 hasil pengukuran, yaitu :

1. True Negative (TN): Merupakan jumlah data review produk skincare skintific yang diklasifikasikan negatif dan terdeteksi dengan benar.
2. False Positive (FP): Merupakan jumlah data review produk skincare skintific yang diklasifikasikan negatif tetapi terdeteksi sebagai data klasifikasi positif.
3. True Positive (TP): Merupakan jumlah data review produk skincare skintific yang diklasifikasikan salah terdeteksi benar.
4. False Negative (FN): Merupakan kebalikan dari True Positive.

- 1) Akurasi : Menunjukkan hasil analisis data review yang diklasifikasi oleh model dengan tepat.
- 2) Recall : Menunjukkan banyaknya data review untuk label positif yang terklasifikasi secara benar diantara data yang diprediksi sebagai positif.
- 3) Precision : Merupakan nilai rasio antara jumlah data review berlabel positif yang

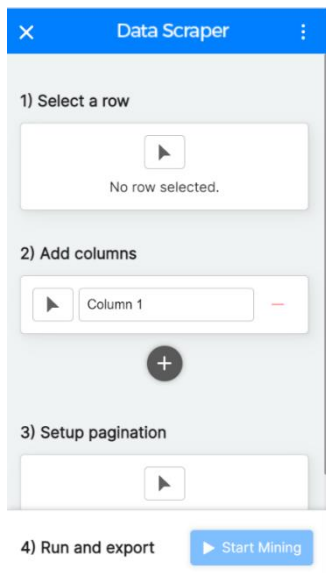
terklasifikasi secara benar diantara semua data dengan label positif.

- 4) F-1 Score : Nilai ini merupakan rata-rata dari precision dan recall

3. HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang digunakan adalah data review produk *skincare* skintific di website Tokopedia melalui proses *scraping* dengan menggunakan *tools Data Scraper*. Tampilan *tools Data Scraper* dapat dilihat pada gambar 4.

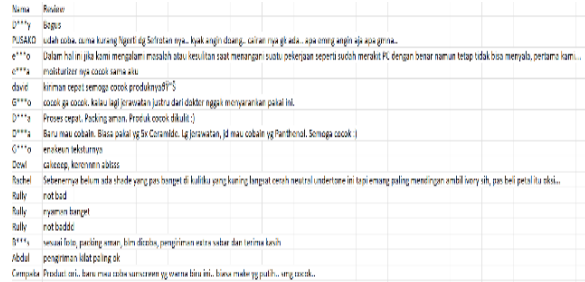


Gambar 4. Pengumpulan data dengan tools data scraper

Sebelum menggunakan *tools Data Scraper*, langkah pertama yang dilakukan adalah dengan menginstall *tools* tersebut ke dalam Google Chrome. Setelah itu, tentukan data mana yang akan diambil untuk di scrape. Pada penelitian ini, data yang diambil adalah data nama dan review produk. Selanjutnya, untuk melakukan *scraping* data dapat dimulai dengan mengklik tombol *start mining*.

Hasil Pengambilan Data

Data dari hasil *scraping* kemudian disimpan dalam bentuk file *csv* yang berjumlah 500 data untuk mempermudah penyimpanan dan penggunaan. Pada pengumpulan data ini, data yang digunakan adalah nama *reviewers* dan ulasan pengguna. Berikut hasil pengambilan data yang dapat dilihat pada gambar 5.



Gambar 5. Hasil Pengambilan Data

Preprocessing Data

Data yang telah di *scraping* masih merupakan data mentah. Oleh karena itu, tahap selanjutnya adalah dengan melakukan pembersihan data untuk mengurangi *noise*. Tahapan *preprocessing* ini menggunakan *library* pada Bahasa pemrograman Python. Beberapa *library* python yang digunakan antara lain : *pandas*, *numpy*, *matplotlib*, *swifter*, *scikit-learn*, *seaborn*, *tensorflow*. Berikut penjelasan untuk setiap tahapan *preprocessing* yang dilakukan:

Casefolding

Casefolding merupakan tahap awal *preprocessing* data untuk mengubah karakter pada data menjadi huruf kecil. Dilakukannya *casefolding* ini karena tidak semua kata dalam teks konsisten menggunakan huruf kapital. Maka dari itu, dilakukan proses *penyamarataan* huruf ini. *Source code* proses *casefolding* dapat dilihat pada gambar 6.

```
#case folding
data['Review'] = data['Review'].str.lower()
print(data['Review'].head())
```

Gambar 6. Source code Casefolding

Tabel 1. Hasil Casefolding

Teks	Casefolding
Tipe Kulit Normal, Tidak Iritasi Udah coba creamnya dan cocok.....semoga dengan adanya serum jadi lebih maksimal hasilnya.....soon to be glowing 🧡🥰 pengiri man cepat, sel...	tipe kulit normal, tidak iritasi udah coba creamnya dan cocok.....semoga dengan adanya serum jadi lebih maksimal hasilnya.....soon to be glowing 🧡🥰 pengiri man cepat, sel...

Pada Tabel 1 dari teks hasil *scraping* lalu dilakukan *casefolding* terdapat perubahan, yaitu pada kalimat ‘Tidak Kulit Normal’ dan ‘Tidak Iritasi Udah’ yang semula pada awal kata

menggunakan huruf kapital, setelah dilakukan case folding menjadi huruf kecil.

Data Cleansing

Setelah dilakukan casefolding, maka tahap selanjutnya adalah melakukan data cleansing. Tujuan dari data cleansing ini adalah untuk menghapus angka, simbol, karakter berulang, yang dapat mengganggu hasil analisis. Source code proses cleansing dapat dilihat pada gambar 7.

```
#Cleaning text
def remove_punct(data):
    data = re.sub("[@A-Za-z0-9+]|([^0-9A-Za-z \t])|(\w+\/\w+\/\w+)", " ", data)
    data = re.sub(r"^\r|\s+", "", data)
    data = re.sub('/n', ' ', data)
    data = re.sub(r'((www\.[^\s]+)|(https?://[^\s]+)|(http?://[^\s]+))', ' ', data)
    data = re.sub(r'+', ' ', data)
    return data

pd.set_option("display.max_colwidth", -1)
data['Cleaning'] = data['Review'].apply(lambda x: remove_punct(x))
data.head()
```

Gambar 7. Source code data cleansing

Tabel 2. Hasil Data Cleansing

Casefolding	Data Cleansing
tipe kulit normal, tidak iritasi udah coba creamnya dan cocok.....semoga dengan adanya serum jadi lebih maksimal hasilnya.....soon to be glowing 🧡🥰 pengiri man cepat, sel...	tipe kulit normal tidak iritasi udah coba creamnya dan cocok semoga dengan adanya serum jadi lebih maksimal hasilnya soon to be glowing pengiriman cepat sel

Pada tabel 2 dari hasil casefolding lalu dilakukan data cleansing terdapat perubahan, yaitu yang semula terdapat tanda baca titik dan emoji maka pada tahap data cleansing dihilangkan.

Tokenizing

Tahapan setelah dilakukan data cleansing adalah tokenizing, yaitu memecah sebuah kalimat ke dalam kepingan-kepingan kata. Tujuan dilakukannya tokenizing adalah untuk memudahkan dalam memaknai setiap kata. Source code proses tokenizing dapat dilihat pada gambar 8.

```
#tokenization
def tokenization(text):
    text = re.split('\W+', text)
    return text

data['Tokenizing'] = data['Cleaning'].apply(lambda x: tokenization(x.lower()))
data.head()
```

Gambar 8. Source Code Tokenizing

Tabel 3. Hasil Tokenizing

Data Cleansing	Tokenizing
tipe kulit normal tidak iritasi udah coba creamnya dan cocok semoga dengan adanya serum jadi lebih maksimal hasilnya soon to be glowing pengiriman cepat sel	[tipe, kulit, normal, tidak, iritasi, udah, coba, creamnya, dan, cocok, semoga, dengan, adanya, serum, jadi, lebih, maksimal, hasilnya, soon, to, be, glowing, pengiriman, cepat, sel,]

Pada tabel 3 dari hasil data cleansing lalu dilakukan tokenizing, dimana pada tahap ini terdapat perubahan, yaitu yang semula teks tersebut menjadi satu kalimat maka pada tahap ini dilakukan pemisahan per kata.

Stopword Removal

Tahapan *stopword removal* dilakukan untuk menghilangkan kata atau *term* yang tidak memiliki arti. Tahapan ini bisa menggunakan *stoplist* yaitu dengan membuang kata yang tidak memiliki arti atau *wordlist* yaitu menyimpan kata penting. Daftar *stoplist* dibuat sebelum melakukan proses *stopword removal*. Apabila kata tersebut termasuk kedalam *stoplist* maka akan terhapus. Pada penelitian ini menggunakan `nltk.download('stopwords')`. Source code tahapan *stopword removal* dapat dilihat pada gambar 9.

```
#stopremoval
from nltk.corpus import stopwords
stopwords_ind = stopwords.words('indonesian')

def remove_stopwords(text):
    text = [word for word in text if word not in stopwords_ind]
    return text

data['Stop Removal'] = data['Tokenizing'].apply(lambda x: remove_stopwords(x))
data.head()
```

Gambar 9. Source Code Stopword Removal setelah tahap tokenizing

Pada Gambar 9 hasil data masih dalam bentuk *tokenizing* yang berbentuk per kata akan tetapi sudah di lakukan *stopword removal*. Untuk mengubahnya agar tidak menjadi bentuk per kata, maka dilakukan proses seperti Gambar 10.

```
stop_removal = data[['Stop Removal']]

def fit_stopwords(text):
    text = np.array(text)
    text = ' '.join(text)
    return text

data['Stop Removal'] = data['Stop Removal'].apply(lambda x: fit_stopwords(x))
data.head()
```

Gambar 10. Source Code stopword removal

Tabel 4. Hasil Stopword Removal

Tokenizing	Stopword Removal
[tipe, kulit, normal, tidak, iritasi, udah, coba, creamnya, dan, cocok, semoga, dengan, adanya, serum, jadi, lebih, maksimal, hasilnya, soon, to, be, glowing, pengiriman, cepat, sel,]	tipe kulit normal iritasi udah coba creamnya cocok semoga serum maksimal hasilnya soon to be glowing pengiriman cepat sel

Pada tabel 4 dari hasil tokenizing lalu dilakukan stopwords removal terdapat perubahan, yaitu terdapat penghilangan kata 'dan', 'dengan', 'adanya', 'jadi', 'lebih'.

Stemming

Tahapan yang terakhir yaitu stemming. Tahap ini bertujuan untuk mencari kata dasar dari setiap kata hasil filtering sebelumnya. Tahapan ini dilakukan dengan proses pengembalian berbagai bentuk kata menjadi bentuk kata dasar dengan menghilangkan imbuhan baik awalan dan akhiran. Proses ini dilakukan dengan mengimport StemmerFactory dari library Sastrawi. Source code proses stemming dapat dilihat pada gambar 11.

```
#Stemming
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stemming(text):
    return stemmer.stem(text)

data['Stemming'] = data['Stop Removal'].apply(lambda x: stemming(x))
data.head()
```

Gambar 11. Source Code Stemming

Tabel 5. Hasil Stopword Removal

Stopword Removal	Stemming
tipe kulit normal iritasi udah coba creamnya cocok semoga serum maksimal hasilnya soon to be glowing pengiriman cepat sel	tipe kulit normal iritasi udah coba creamnya cocok moga serum maksimal hasil soon to be glowing kirim cepat sel

Pada tabel 5 dari hasil stopwords removal lalu dilakukan stemming terdapat perubahan, yaitu terdapat kata dasar dari kata 'semoga' menjadi 'moga', kata 'hasilnya', menjadi 'hasil', dan kata 'pengiriman' menjadi 'kirim'.

Pelabelan Data

Pelabelan data dilakukan dengan menghitung nilai polarity dari kalimat tersebut. Selanjutnya, apabila nilai polarity pada kalimat lebih dari 0.0 maka dikategorikan ke dalam sentimen positif, lalu jika polarity kalimatnya sama dengan 0.0 maka dikategorikan dalam sentimen netral, dan apabila nilai polarity kurang dari 0.0 maka dikategorikan ke dalam sentimen negatif. Berikut source code untuk pelabelan data dapat dilihat pada gambar 12 dan 13.

```
from textblob import TextBlob

def thePolarity(text):
    analysis = TextBlob(text)
    return analysis.sentiment.polarity

data['Polarity'] = data['Clean_Text'].apply(lambda x: thePolarity(x))
#(thePolarity)
data.head()
```

Gambar 12. Source Code Polarity

Selanjutnya, dari hasil label tersebut, dilakukan encoding, yaitu proses perubahan data menjadi numerik.

```
from sklearn.preprocessing import LabelEncoder

labelencoder = LabelEncoder()
data['Label'] = labelencoder.fit_transform(data['Label'])
data.head()
```

Gambar 13. Source Code Encoding

Tabel 6. Hasil labelling sebelum encoding

Clean Text	Polarity	Label
cobain pakai yg 5x ceramide lg jerawat jd cobain yg panthenol moga cocok	0	Netral
sesuai foto packing aman blm coba kirim extra sabar terima kasih	0	Netral
langsung ngefek jerawat kaget haha top banget	0.35	Positif
kering kulit dry sensitif	-0.06667	Negatif
fast respon kirim cpt brg yg kirim bnr blm coba ya smg cocok kulit muka sy	0.2	Positif
barang kualitas harga promo kemas moga serum cocok wajah bruntusan thanks seller	0.2	Positif
tekstur ringan bebas white cast respon jual top packaging aman proses pesan cepat cepat resap	0.25	Positif

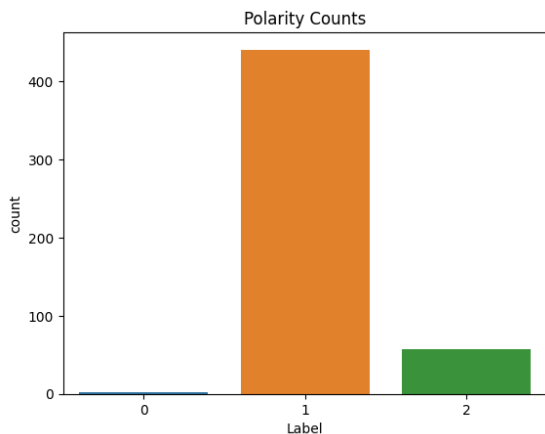
Tabel 7. Hasil labelling setelah encoding

Clean Text	Polarity	Label
cobain pakai yg 5x ceramide lg jerawat jd cobain yg panthenol moga cocok	0	1
sesuai foto packing aman blm coba kirim extra sabar terima kasih	0	1
langsung ngefek jerawat kaget haha top banget	0.35	2
kering kulit dry sensitif	-0.06667	0
fast respon kirim cpt brg yg kirim bnr blm coba ya smg cocok kulit muka sy	0.2	2
barang kualitas harga promo kemas moga serum cocok wajah bruntusan thanks seller	0.2	2
tekstur ringan bebas white cast respon jual top packaging aman proses pesan cepat cepat resap	0.25	2

Visualisasi Kata



Gambar 14. Hasil Visualisasi data semua kelas



Gambar 15 Grafik jumlah label tiap kelas

Berdasarkan gambar 15 jumlah klasifikasi sentimen tersebut untuk masing – masing kelas yaitu 2 data untuk sentiment negatif, 441 untuk sentiment netral, dan 57 data untuk sentiment positif.

Perancangan Arsitektur CNN

Sebelum melakukan pemodelan, maka dilakukan tahapan :

OverSampling

Proses *oversampling* dilakukan agar tidak terjadi data yang tidak seimbang, jumlah kelas data (*instance*) yang satu lebih sedikit atau lebih banyak disbanding dengan jumlah kelas yang lainnya.

```
Before OverSampling, counts of label '2': [0 0 0]
Before OverSampling, counts of label '1': [ 2 352 46]
Before OverSampling, counts of label '0': [398 48 354]
```

Gambar 16. Data sebelum oversampling

```
After OverSampling, counts of label '2': [0 0 0]
After OverSampling, counts of label '1': [352 352 352]
After OverSampling, counts of label '0': [704 704 704]
```

Gambar 17. Data setelah oversampling

Dari gambar 16 dan 17 terdapat perbedaan data sebelum dan setelah *oversampling*. Sebelum *oversampling* data pada masing-masing kelas tidak beraturan. Namun, setelah dilakukan *oversampling* data pada kelas 1 menjadi 352. Sedangkan untuk data pada kelas 2 menjadi 704. Data-data yang dihasilkan setelah adanya *oversampling* terbentuk untuk menyeimbangkan kelas minoritas, kemudian di-*fit* dan diterapkan pada *dataset*. Sehingga nantinya kelas minoritas memiliki jumlah yang sama dengan kelas mayoritas.

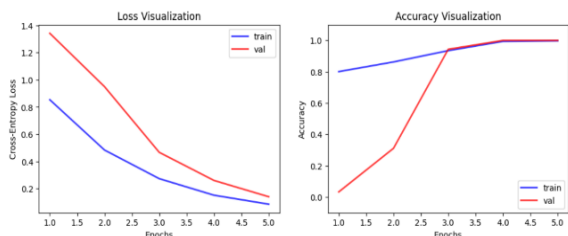
Modelling

Sebelum ke tahap *modelling*, dilakukan *splitting dataset* dengan *test_size* yaitu 0.2, artinya data yang digunakan sebagai data testing adalah 20% dari keseluruhan dataset yang ada dengan random state 42. Berikut disajikan eksperimen modelling yang dapat dilihat pada tabel 8.

Tabel 8. Experimen Modelling

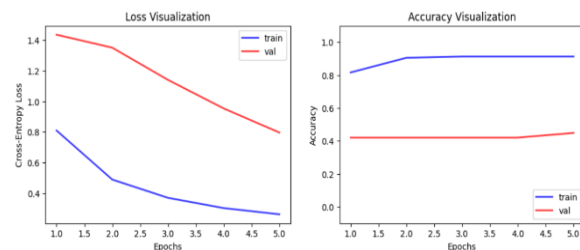
Learning rate	Susunan layer	Epoch	Validation Split	Akurasi
0.01	Embedding () Conv1D(256,3,tanh) Dropout(0.2) GlobalMaxPooling1D Dense (3, softmax)	5	0.2	0.97
0.01	Embedding () Conv1D(256,3,tanh) Dropout(0.2) GlobalMaxPooling1D Dense (3, softmax)	5	0.5	0.89
0.01	Embedding () Conv1D(256,3,tanh) Dropout(0.5) GlobalMaxPooling1D Dense (3, softmax)	5	0.2	0.94

Hasil dari experimen model dapat disimpulkan bahwa pengaruh perbedaan adanya validation split juga mempengaruhi hasil akurasi. Pada tabel, untuk model 1 dan 2 dengan validation split 0.2 pada model 1 menghasilkan nilai akurasi 0.97 (97%), sedangkan pada model 2 dengan nilai validation split 0.5 menghasilkan nilai akurasi 0.89 (89%). Selanjutnya, untuk model 3 dengan perbedaan dropout dengan model 1 tidak terlalu mempengaruhi nilai akurasi. Hal tersebut dibuktikan dengan hasil nilai akurasi yang mencapai 94% tidak berbeda jauh dengan nilai akurasi yang dihasilkan oleh model 1, yaitu 97%. Dengan kata lain, nilai akurasi untuk model 1 dan 3 dengan perbedaan dropout masih diatas 90%. Berikut ditampilkan hasil visualisasi akurasi untuk tiap model.



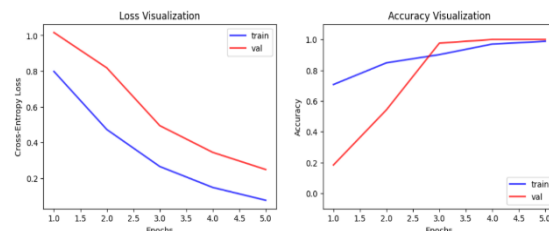
Gambar 18. Gambar accuracy dan loss visualization model 1

Pada gambar 18 grafik loss function rendah ditandai dengan gambar grafik menurun kebawah. Hal tersebut berarti kinerja model bekerja dengan baik. Kemudian, untuk grafik accuracy antara train accuracy dan validation accuracy juga menunjukkan kinerja model sudah optimal.



Gambar 19. Gambar accuracy dan loss visualization model 2

Pada gambar 19 grafik loss function tidak dapat mendeteksi model dengan baik, hal itu karena grafik menampilkan garis datar atau nilai noise yang cukup tinggi. Jadi, bisa dikatakan model mengalami underfitting. Kemudian, untuk grafik accuracy antara train accuracy dan validation accuracy juga menunjukkan model yang belum optimal.

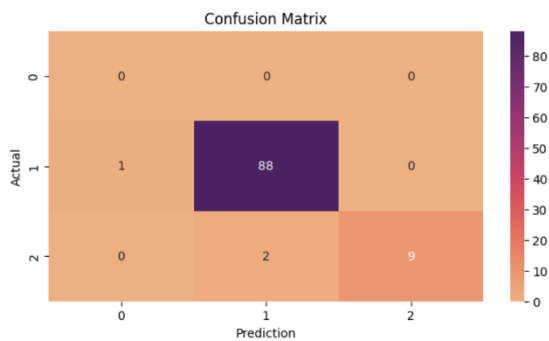


Gambar 20. Gambar accuracy dan loss visualization model 3

Pada gambar 20 grafik loss function tidak dapat mendeteksi model dengan baik, karena model mengalami underfitting. Kemudian, untuk grafik accuracy antara train accuracy dan validation accuracy model menunjukkan sudah bekerja secara optimal.

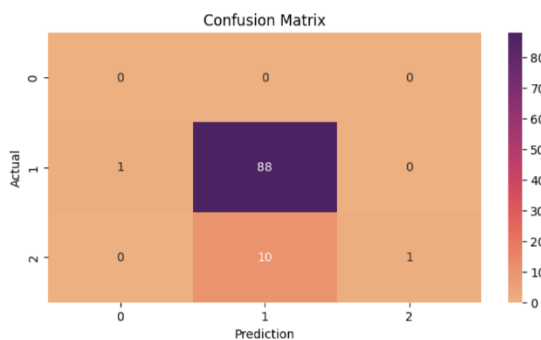
Confusion Matrix

Berikut confusion matrix pada model 1 dapat dilihat pada gambar 21.



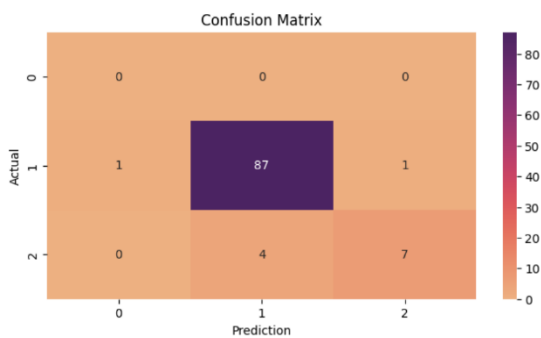
Gambar 21. Confusion Matrix model 1

Berdasarkan gambar 21 menunjukkan bahwa model mengklasifikasikan secara benar sebanyak 0 data sebagai negatif, 88 data sebagai positif, dan 9 data sebagai netral.



Gambar 22. confusion matrix model 2

Berdasarkan gambar 22 menunjukkan bahwa model mengklasifikasikan secara benar sebanyak 0 data sebagai negatif, 88 data sebagai positif, dan 1 data sebagai netral.



Gambar 23. confusion matrix model 3

Berdasarkan gambar 23 menunjukkan bahwa model mengklasifikasikan secara benar sebanyak 0 data sebagai negatif, 87 data sebagai positif, dan 7 data sebagai netral.

Classification Report

Tabel 9. Hasil Classification Report Model 1

	Precision	Recall	F1-score
0	0.00	0.00	0.00
1	0.98	0.99	0.98
2	1.00	0.82	0.90
Accuracy (%)			
97			

Berdasarkan tabel 9 dapat disimpulkan bahwa hasil review produk skincare skintific didapatkan hasil precision, recall, dan f1-score untuk kelas kelas 0 atau negatif yaitu 0%. Sedangkan untuk nilai precision kelas 1 (netral) dan kelas 2 (positif) yaitu 0.98 dan 1.00. Artinya model dalam meprediksi semua ulasan produk skincare skintific yang diklasifikasikan kedalam kelas netral dan positif, 98% dan 100% nya benar diklasifikasikan ke dalam kelas tersebut.

Untuk nilai recall kelas 1 (netral) dan kelas 2 (positif) yaitu 0.99 dan 0.82. Artinya model mampu mengklasifikasikan 99% dan 82% dari seluruh ulasan produk skincare skintific dengan tepat. Lalu, untuk nilai f1-score pada kelas netral yaitu 98% dan untuk kelas positif yaitu 90%, kedua kelas tersebut nilainya mendekati 1 berarti bahwa semua prediksi benar.

Tabel 10. Hasil Classification Report Model 2

	Precision	Recall	F1-score
0	0.00	0.00	0.00
1	0.90	0.99	0.94
2	1.00	0.09	0.17
Accuracy (%)			
89			

Berdasarkan tabel 10 dapat disimpulkan bahwa hasil review produk skincare skintific didapatkan hasil precision, recall, dan f1-score untuk kelas kelas 0 atau negatif yaitu 0%. Sedangkan untuk nilai precision kelas 1 (netral) dan kelas 2 (positif) yaitu 0.90 dan 1.00. Artinya model dalam meprediksi semua ulasan produk skincare skintific yang diklasifikasikan kedalam kelas netral dan positif, 90% dan 100% nya memang benar diklasifikasikan ke dalam kelas tersebut.

Untuk nilai recall kelas 1 (netral) dan kelas 2 (positif) yaitu 0.99 dan 0.09. Artinya model mampu mengklasifikasikan 99% dan 9% dari seluruh ulasan produk skincare skintific dengan

tepat. Lalu, untuk nilai f1-score pada kelas netral yaitu 94%, dimana nilainya mendekati 1 berarti bahwa semua prediksi benar. Untuk kelas positif nilai f1-score yaitu 17%, dimana nilai tersebut mendekati 0 maka nilai f1-score pada kelas positif buruk.

Tabel 11. Hasil Classification Report Model 3

	Precision	Recall	F1-score
0	0.00	0.00	0.00
1	0.96	0.98	0.97
2	0.88	0.64	0.74
Accuracy (%)			
94			

Berdasarkan tabel 11 dapat disimpulkan bahwa hasil review produk skincare skintific didapatkan hasil precision, recall, dan f1-score untuk kelas 0 atau negatif yaitu 0%. Sedangkan untuk nilai precision kelas 1 (netral) dan kelas 2 (positif) yaitu 0.96 dan 0.88. Artinya model dalam memprediksi semua ulasan produk skincare skintific yang diklasifikasikan kedalam kelas netral dan positif, 96% dan 88% nya memang benar diklasifikasikan ke dalam kelas tersebut.

Untuk nilai recall kelas 1 (netral) dan kelas 2 (positif) yaitu 0.98 dan 0.64. Artinya model dalam mengklasifikasikan ulasan produk skincare skintific yang benar-benar kedalam kelas netral dan positif, 98% dan 64% nya diklasifikasikan dengan benar. Lalu, untuk nilai f1-score pada kelas netral yaitu 97% dan untuk kelas positif yaitu 74%, kedua kelas tersebut nilainya mendekati 1 berarti bahwa semua prediksi benar.

4. PENUTUP

Kesimpulan

Hasil klasifikasi sentimen menggunakan algoritma CNN pada *review* produk untuk *skincare* skintific di website Tokopedia cenderung bersentimen netral.

Hasil Akurasi terbaik didapatkan pada percobaan pertama dengan nilai akurasi 97% dan *hyperparameter batch size 32, learning rate 0.01, dan epoch 5* dengan susunan *layer* yaitu, *Embedding(), Conv1D(256,3,tanh), Dropout(0.2), GlobalMaxPooling1D, Dense(3, softmax)*.

Perbedaan adanya *validation split* dan *dropout* mempengaruhi hasil akurasi model.

Saran

Peneliti memberikan saran untuk penelitian selanjutnya agar dataset yang digunakan dapat diperbanyak dan mencoba algoritma lain atau membandingkan dengan algoritma yang lainnya.

5. REFERENSI

- [1] D. M. Dwitari and L. Kusdibyo, "Mengukur Sikap dan Minat Beli Konsumen Terhadap Produk Skin Care dengan Menggunakan Brand Image Merek Lokal," *Ind. Res. Work. Natl. Semin.*, vol. 10, no. 1, pp. 686–696, 2019.
- [2] A. Z. N. Fauzia and A. Sosianika, "Analisis Pengaruh Brand Image, Perceived Quality, dan Country of Origin Terhadap Minat Beli Produk Skincare Luar Negeri," *Pros. Ind. Res. Work. Natl. Semin.*, vol. 12, pp. 1068–1072, 2021.
- [3] K. Kussudyarsana and L. Rejeki, "Pengaruh Media Sosial Online Dan Media Promosi Offline Terhadap Pemilihan Merek Produk Skincare Dan Klinik Kecantikan," *J. Manaj. Daya Saing*, vol. 22, no. 1, pp. 1–11, 2020, doi: 10.23917/dayasaing.v22i1.10701.
- [4] N. Rohadatul Anisa, "Pengaruh Label Halal dan Inovasi Produk Terhadap Minat Beli Ulang (Survei Online Pada Pengikut Instagram @SafiIndonesia)," *J. Indones. Sos. Sains*, vol. 2, no. 12, pp. 2138–2146, 2021, doi: 10.36418/jiss.v2i12.453.
- [5] D. Marlina, K. T. Basuki, Z. M. Zaki, and ..., "Sentiment Analysis on Natural Skincare Products," *J. Data ...*, vol. 2022, no. August, 2022, [Online].
- [6] D. Hertina and D. Wulandari, "Pengaruh harga, kualitas produk dan brand image terhadap keputusan pembelian," *Fair Value J. Ilm. Akunt. dan Keuang.*, vol. 4, no. 12, pp. 5379–5384, 2022, doi: 10.32670/fairvalue.v4i12.1988.
- [7] B. P. Stevy, "Brand image," *New Sci.*, vol. 165, no. 2229, p. 57, 2019.
- [8] A. Hapsila and I. Astarina, "Pengaruh Label Halal, Brand Image Dan Harga Terhadap Keputusan Pembelian Produk Kosmetik

- Rk,” *J. Manaj. dan Bisnis*, vol. 9, no. 1, pp. 41–50, 2020.
- [9] F. A. Irawan and D. A. Rochmah, “Penerapan Algoritma CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19,” *J. Inform.*, vol. 9, no. 2, pp. 148–158, 2022, doi: 10.31294/inf.v9i2.13257.
- [10] A. Rakhman and M. R. Tsani, “Analisis Sentimen Review Media Massa,” *Smart Comp*, vol. 8, no. 2, pp. 78–82, 2019.
- [11] P. L. Parameswari and Prihandoko, “Penggunaan Convolutional Neural Network Untuk Analisis Sentimen Opini Lingkungan Hidup Kota Depok Di Twitter,” *J. Ilm. Teknol. dan Rekayasa*, vol. 27, no. 1, pp. 29–42, 2022, doi: 10.35760/tr.2022.v27i1.4671.
- [12] S. H. Badjrie, O. N. Pratiwi, and H. D. Anggana, “Analisis Sentimen Review Customer Terhadap Produk Indihome Dan First Media Menggunakan Algoritma Convolutional Neural Network Review Analysis Sentiment Customer Product Indihome And First Media Using Convolutional Neural Network,” *eProceedings ...*, vol. 8, no. 5, pp. 9049–9061, 2021, [Online].
- [13] S. N. Listyarini and D. A. Anggoro, “Analisis Sentimen Pilkada di Tengah Pandemi Covid-19 Menggunakan Convolution Neural Network (CNN),” *J. Pendidik. dan Teknol. Indones.*, vol. 1, no. 7, pp. 261–268, 2021, doi: 10.52436/1.jpti.60.
- [14] I. M. A. Susilayasa, A. A. I. Eka Karyawati, L. G. Astuti, L. A. A. Rahning Putri, I. G. Arta Wibawa, and I. K. Ari Mogi, “Analisis Sentimen Ulasan E- Commerce Pakaian Berdasarkan Kategori dengan Algoritma Convolutional Neural Network,” *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 1, p. 1, 2022, doi: 10.24843/jlk.2022.v11.i01.p01.
- [15] E. Y. Hidayat and D. Handayani, “Penerapan 1D-CNN untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female Daily Review,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 3, pp. 153–163, 2023, doi: 10.25077/teknosi.v8i3.2022.153-163.
- [16] I. Miati, “Pengaruh Citra Merek (Brand Image) Terhadap Keputusan Pembelian Kerudung Deenay (Studi pada Konsumen Gea Fashion Banjar),” *Abiwara J. Vokasi Adm. Bisnis*, vol. 1, no. 2, pp. 71–83, 2020, doi: 10.31334/abiwara.v1i2.795.
- [17] M. S. Anam, D. L. Nadila, T. A. Anindita, and R. Rosia, “Pengaruh Kualitas Produk, Harga dan Brand Image terhadap Keputusan Pembelian Produk Hand and Body Lotion Merek Citra,” *Jesya (Jurnal Ekon. Ekon. Syariah)*, vol. 4, no. 1, pp. 120–136, 2020, doi: 10.36778/jesya.v4i1.277.
- [18] I. Bima, I. N. Nurfarida, and E. Sarwoko, “Determinan keputusan pembelian di Shopee: online customer review, brand image dan promosi,” *Manag. Bus. Rev.*, vol. 5, no. 2, pp. 239–250, 2021, doi: 10.21067/mbr.v5i2.5554.
- [19] D. C. Putri and C. K. Dewi, “Pengaruh Brand Image Terhadap Purchase Intention Melalui Brand Loyalty Pada Billionaires Project,” *eProceedings Manag.*, vol. 8, no. 3, 2021,
- [20] F. Edrin and N. Fhatiyah, “Pengaruh Brand Ambassador Dan Brand Image Terhadap Keputusan Pembelian Konsumen (Kasus Pada Event Free Photocard BTS Tokopedia),” *Proceeding Semin. Nas. Bisnis Seri IV*, pp. 78–84, 2022.
- [21] V. Wu and Z. Chen, “Saliency map generation based on saccade target theory,” *Proc. - IEEE Int. Conf. Multimed. Expo*, pp. 529–534, 2017, doi: 10.1109/ICME.2017.8019456.
- [22] B. Review, “*Journal of Technopreneurship*,” vol. 1, no. 1, pp. 58–68, 2020.
- [23] L. R. H. Irwanto, “Mengenal Arti Skincare dan Tahapan Merawat Kulit,” *J. Komun.*, vol. 11, no. 2, pp. 119–128, 2020,
- [24] Y. Yuliska, D. H. Qudsi, J. H. Lubis, K. U. Syaliman, and N. F. Najwa, “Analisis Sentimen pada Data Saran Mahasiswa Terhadap Kinerja Departemen di Perguruan Tinggi Menggunakan Convolutional Neural Network,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, p. 1067, 2021, doi: 10.25126/jtiik.2021854842.
- [25] B. Kurniawan, A. Ari Aldino, and A. Rahman Isnain, “Sentimen Analisis Terhadap Kebijakan Penyelenggara Sistem

- Elektronik (Pse) Menggunakan Algoritma Bidirectional Encoder Representations From Transformers (Bert),” *J. Teknol. dan Sist. Inf.*, vol. 3, no. 4, pp. 98–106, 2022,
- [26] A. R. Maulana and N. Rochmawati, “Opinion Mining Terhadap Pemberitaan Corona di Instagram menggunakan Convolutional Neural Network,” *J. Informatics Comput. Sci.*, vol. 2, no. 01, pp. 53–59, 2020, doi: 10.26740/jinacs.v2n01.p53-59.
- [27] F. M. Ihsan and M. N. Fauzan, “Identifikasi Audio Ancaman Menggunakan Metode Convolutional Neural Network,” *J. Sist. dan Teknol. Inf.*, vol. 10, no. 4, pp. 446–452, 2022, doi: 10.26418/justin.v10i4.52433.
- [28] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text classification algorithms: A survey,” *Inf.*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.