

Analisis Sentimen Text Dengan Metode CNN Study Kasus Tempat Wisata Makassar

Safutri Kamal¹, Fahrim Irhamna Rachman², Titin Wahyuni³

^{1,2,3} Informatika, Universitas Muhammadiyah Makassar, Makassar, 90221, Indonesia
105841107820@student.unismuh.ac.id

Abstract

This research aims to evaluate and determine the extent to which the CNN (Convolutional Neural Network) method can produce accurate sentiment predictions for reviews of Makassar tourist attractions. This sentiment analysis method uses review data collected from the Google Maps platform. In this research, a preprocessing stage was carried out to clean the data, such as cleaning, transform cases, tokenizing, stopwords and stemming. Next, the dataset was divided into training data and test data with scenarios 90 : 10, 80 : 20 and 70 : 30 to train and test the model with three review categories, namely positive, negative and neutral. The results of sentiment analysis show that the CNN method has good abilities in predicting positive, negative and neutral sentiment in reviews of Makassar tourist attractions. A high level of accuracy at the training stage shows that the model is able to learn well from the dataset provided. Although the accuracy rate at the validation stage was slightly lower, it still reached an adequate figure, indicating that the model has quite good generalization ability in classifying sentiment in these reviews. The highest accuracy results were obtained with Training Accuracy which increased to obtain a training accuracy value of 95%, and Validation Accuracy obtained a value of 73%.

Keywords : Sentiment analysis, Convolutional Neural Network, Makassar tourist attractions, Google Maps reviews

Abstrak

Penelitian ini bertujuan untuk mengevaluasi dan menentukan sejauh mana metode CNN (Convolutional Neural Network) dapat menghasilkan prediksi sentimen yang akurat terhadap ulasan mengenai tempat wisata Makassar. Metode analisis sentimen ini menggunakan data ulasan yang dikumpulkan dari platform Google Maps. Dalam penelitian ini, dilakukan tahap preprocessing untuk membersihkan data, seperti cleaning, transform cases, tokenizing, stopword dan stemming. Selanjutnya, dilakukan pembagian dataset menjadi data latih dan data uji dengan scenario 90 : 10, 80 : 20 dan 70 : 30 untuk melatih dan menguji model dengan tiga kategori ulasan yaitu positif, negatif dan netral. Hasil dari analisis sentimen menunjukkan bahwa metode CNN memiliki kemampuan yang baik dalam memprediksi sentimen positif, negatif, dan netral pada ulasan mengenai Tempat Wisata Makassar. Tingkat akurasi yang tinggi pada tahap pelatihan menunjukkan bahwa model mampu belajar dengan baik dari dataset yang disediakan. Meskipun tingkat akurasi pada tahap validasi sedikit lebih rendah, tetapi masih mencapai angka yang memadai, menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup baik dalam mengklasifikasikan sentimen pada ulasan-ulasan tersebut. Diperoleh hasil akurasi tertinggi dengan Training Accuracy yang meningkat memperoleh nilai akurasi training 95%, serta Validation Accuracy memperoleh nilai 73%.

Kata Kunci : Analisis sentimen, Convolutional Neural Network, Tempat wisata Makassar, Ulasan Google Maps

1. Pendahuluan

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen[1].

Obyek wisata merupakan manifestasi dari kreativitas manusia, gaya hidup masyarakat, keberagaman seni budaya, dan warisan sejarah suatu bangsa. Dalam membangun obyek wisata tersebut harus memperhatikan keadaan sosial ekonomi masyarakat setempat, sosial budaya daerah setempat, nilai- nilai agama, adat istiadat, lingkungan hidup, dan obyek wisata itu sendiri[2]. Convolutional Neural Network (CNN) merupakan metode yang sukses dalam pengenalan pola, dikembangkan oleh Lecun pada tahun 1988. CNN memiliki arsitektur dengan puluhan hingga ratusan layer. Setiap layer melakukan pembelajaran dengan menggunakan output dari layer sebelumnya sebagai input. Feature maps kemudian dikurangi ukurannya melalui operasi pooling. Proses ini berulang hingga fitur terdalam diekstraksi. Struktur umum dari CNN melibatkan convolutional layer, pooling layer, dan fully connected network[3].

Metode Deep Learning merupakan subset dari Machine Learning yang bekerja dengan mengekstraksi data menggunakan Neural Network. Model Deep Learning yang biasa digunakan untuk menganalisis data berupa teks adalah CNN, RNN, LSTM, dan Gated Recurrent Unit[4]. Deep Learning dirancang untuk terus menganalisa data seperti pada otak manusia dalam mengambil keputusan. Agar kemampuan Deep Learning semakin mempunyai maka Deep Learning menggunakan Algoritma Artificial Neural Network (ANN), yang terinspirasi dari jaringan biologis otak manusia[5].

2. Metode Penelitian

2.1 Analisa Sentimen

Sentimen adalah pendapat atau pandangan yang didasarkan pada perasaan yang berlebih-lebihan terhadap sesuatu (bertentangan dengan pertimbangan pikiran). Sentimen terdapat pada pernyataan atau kalimat yang memiliki pendapat. Sentimen digunakan untuk mengetahui perasaan yang diberikan terhadap topik atau objek[6]. Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral. Dalam analisis sentimen, proses analisis dapat mencakup teks ulasan, forum, tweet, atau blog, dengan data preprocessing mencakup proses tokenization, stopword, penghapusan, stemming, identifikasi sentimen, dan klasifikasi sentimen[7].

2.2 Supervised Learning

Tujuan dari supervised learning adalah untuk melatih model komputer yang dapat mempelajari pola-pola dalam data dan melakukan prediksi akurat terhadap data yang belum diketahui. Dengan menggunakan algoritma supervised learning, model dapat digunakan untuk mengklasifikasikan data ke dalam kategori yang tepat atau untuk melakukan prediksi numerik[8].

2.3 Tensorflow

Tensorflow adalah library perangkat lunak yang dikembangkan oleh Tim Google Brain Mesin Cerdas Google Asosiasi, yang bertujuan untuk melakukan pembelajaran mesin dan jaringan syaraf dalam penelitiannya.

2.4 Flowchart

Flowchart adalah bagan yang menunjukkan alur atau alur dalam suatu program atau prosedur sistem secara logis. Flowchart (bagan alir) adalah sebuah ilustrasi berupa diagram alir dari algoritma-algoritma dalam suatu program, yang menyatakan arah aliran dari program tersebut[9]. Setelah proses membuat flowchart selesai, maka giliran programmer yang akan menerjemahkan desain logis tersebut kedalam bentuk program dengan berbagai bahasa pemrograman yang telah disepakati[10].

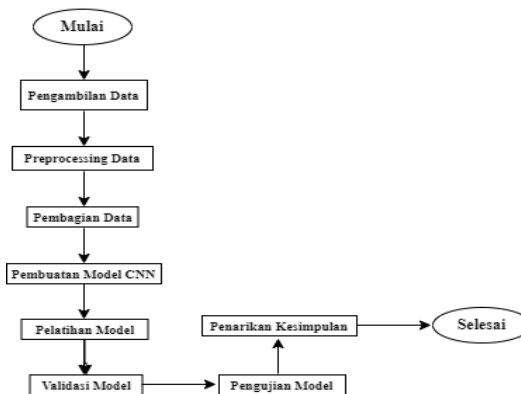
2.5 Scikit – Learn

Scikit – Learn adalah modul Python yang menyediakan berbagai macam

algoritma machine learning. Scikit-learn menggunakan task-oriented interface yang konsisten sehingga dapat dengan mudah dilakukan perbandingan antara metode. Scikit-learn tersedia dalam bentuk library python[11]. Penekanan diberikan pada kemudahan penggunaan, kinerja, dokumentasi, dan konsistensi API. Ini memiliki ketergantungan minimal dan didistribusikan dibawah lisensi BSD yang disederhanakan, mendorong penggunaannya baik dalam aturan akademis dan komersial[12].

2.6 Perancangan Sistem

Perancangan sistem sangat penting dalam pembangunan suatu system karena menguraikan bagaimana suatu sistem dibangun dari tahap perencanaan hingga tahap pembuatan fungsi-fungsi yang diperlukan untuk pengoperasian sistem.



Gambar 1. Diagram alur Proses Penelitian Analisis Sentimen Text Obyek Wisata

2.7 Teknik Pengujian Sistem

Teknik pengujian sistem yang akan digunakan pada pengujian ini adalah menggunakan skenario dengan memisahkan data menjadi dua bagian, yaitu data training dan testing. Teknik ini bertujuan untuk menguji keakuratan dan efektivitas metode *Convolutional Neural Network* (CNN) dalam menganalisis sentimen terhadap teks yang berkaitan dengan tempat wisata di Makassar.

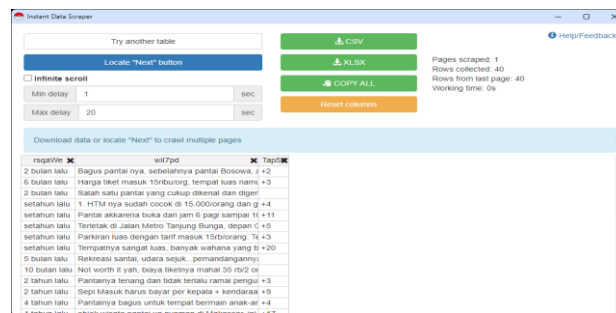
2.8 Teknik Analisis Data

Analisis data merupakan metode yang digunakan untuk mengetahui bagaimana menggambarkan data, hubungan data, semantik data dan batasan data yang ada pada suatu sistem informasi[13]. Proses analisis data dilakukan pada penelitian ini adalah sebagai berikut :

Proses analisis data pada penelitian ini memiliki 3 langkah yaitu :

- a. Reduksi Data (*Data Reduction*)
- b. Penyajian Data (*Display Data*)
- c. Penarikan Kesimpulan (*Concluding Drawing Verivation*)

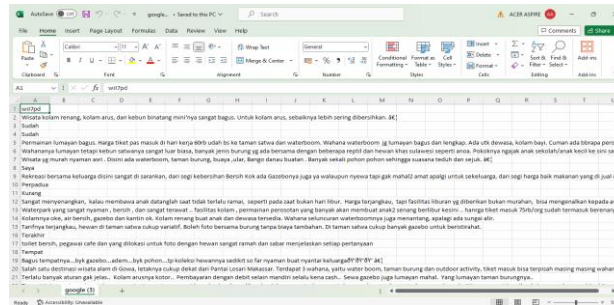
3. Hasil Dan Pembahasan



Gambar 2. Scraping Data Ulasan Menggunakan Instant Data Scraper

3.1 Pengambilan Data

Pengambilan data diakses dari ulasan google maps dalam bentuk file berekstensi.csv. Ulasan ini dapat mencakup beragam topik, mulai dari pujian terhadap pemandangan, fasilitas, atau pelayanan hingga keluhan tentang keramaian atau kebersihan. Data yang berhasil didapatkan adalah 4.500 data. Berikut adalah hasil scraping data ulasan yang telah di simpan ke excel dapat dilihat pada gambar di bawah:



Gambar 3. Dataset Ulasan

3.2 Pelabelan Data

Tahap pelabelan dilakukan untuk memberikan keterangan positif, negatif dan netral terhadap data ulasan tempat wisata makassar. Berikut pelabelan data dapat dilihat pada table berikut :

Tabel 1. Tahap Pelabelan Data

Ulasan	Label
Tempatnya sejuk. Lebih cocok untuk rekreasi anak anak. Tiket weekdays 15.000 / orang. Tempat duduk banyak. Ada pemancingan, wahana bermain anak, sewa sepeda listrik atau ATV, juga kolam renang	Positif
Sangat jelek pantai pasir kasarnya luar biasa pantainya tidak bersih	Negatif
"View pantai yang bagus, namun sayang pantainya agak kotor ketika saya datang."	Netral

3.3 Tahap Preprocessing

a. Cleaning

Pada tahap cleaning akan dilakukan proses pembersihan data dari penghapusan karakter khusus, seperti tanda baca, simbol, atau karakter non-alfanumerik.

Tabel 2. Tahap Cleaning Data

Sebelum	Sesudah
Harga Tiket murah,tempatnya bersih,dan gazebo tidak perlu bayar... 😊😊😊	harga tiket murah,tempatnya bersih dan gazebo tidak perlu bayar
Tempatnya bersih, sejuk, nyaman, dan permainan nya murah2, Mantapp 😊😊😊	tempatny bersih sejuk nyaman dan permainan nya murah mantapp
di hari-hari biasa (bukan hari libur) jumlah pengunjung yg datang masih bisa dibilang rame, cocok buat liburan keluarga 🍷🍷	di harihari biasa bukan hari libur jumlah pengunjung yg datang masih bisa dibilang rame cocok buat liburan keluarga
Rekomended untuk acara anak di bawah umur, sangat safety 😊😊😊😊😊😊😊	rekomended untuk acara anak di bawah umur sangat safety

b. Transform Cases

Tahap ini adalah proses mengubah huruf kapital menjadi huruf kecil dalam teks.

Tabel 3. Tahap Transform Cases

Sebelum	Sesudah
Harga Tiket murah tempatnya bersih dan gazebo tidak perlu bayar	harga tiket murah tempatnya bersih dan gazebo tidak perlu bayar
Tempatnya bersih sejuk nyaman dan permainan nya murah mantapp	tempatnya bersih sejuk nyaman dan permainan nya murah mantapp
di harihari biasa bukan hari libur jumlah pengunjung yg datang masih bisa dibilang rame cocok buat liburan keluarga	di harihari biasa bukan hari libur jumlah pengunjung yg datang masih bisa dibilang rame cocok buat liburan keluarga
Rekomended untuk acara anak di bawah umur sangat safety	rekomended untuk acara anak di bawah umur sangat safety

c. *Tokenizing*

Tahap ini melakukan proses membagi teks menjadi unit – unit yang lebih kecil seperti kata – kata, frasa atau kalimat. Ini membantu dalam pemrosesan teks lebih lanjut karena masing – masing unit ini dapat diproses secara terpisah.

Tabel 4. Tahap Tokenizing

Sebelum	Sesudah
kolamx kotor ada sampah sisa botol air mineral air tdk jernih kaporit kebanyakan sampai sepet di mulut gigi seperti terkikistehel tajam anakku terlukaberlumut juga hampir semua kolam ada yg menggunakan ban besar susah berenang	['kolamx', 'kotor', 'ada', 'sampah', 'sisa', 'botol', 'air', 'mineral', 'air', 'tdk', 'jernih', 'kaporit', 'kebanyakan', 'sampai', 'sepet', 'di', 'mulut', 'gigi', 'seperti', 'terkikistehel', 'tajam', 'anakku', 'terlukaberlumut', 'juga', 'hampir', 'semua', 'kolam', 'ada', 'yg', 'menggunakan', 'ban', 'besar', 'susah', 'berenang']
saran tolong di pisahkan kolam yg bs digunakan khusus berenang dan memakai ban apalagi ban besar yg peruntukanx buat wahana seluncuran	['saran', 'tolong', 'di', 'pisahkan', 'kolam', 'yg', 'bs', 'digunakan', 'khusus', 'berenang', 'dan', 'memakai', 'ban', 'apalagi', 'ban', 'besar', 'yg', 'peruntukanx', 'buat', 'wahana', 'seluncuran']
tempat wisata keluarga yang cukup ramai waktu berkunjung di awal pekan fasilitas yang tersedia kolam renang tempat makan aula pertemuan dll	['tempat', 'wisata', 'keluarga', 'yang', 'cukup', 'ramai', 'waktu', 'berkunjung', 'di', 'awal', 'pekan', 'fasilitas', 'yang', 'tersedia', 'kolam', 'renang', 'tempat', 'makan', 'aula', 'pertemuan', 'dll']

d. *Stopword*

Stopword adalah kata – kata umum yang sering muncul dalam teks dan dianggap tidak memberikan nilai tambah dalam analisis teks, seperti “dan”, “di”, “dari” dan lain – lain.

Tabel 5. *Stopword*

Sebelum	Sesudah
kolamx kotor ada sampah sisa botol air mineral air tdk jernih kaporit kebanyakan sampai sepet di mulut gigi seperti terkikistehel tajam anakku terlukaberlumut juga hampir semua kolam ada yg menggunakan ban besar susah berenang	kolamx kotor sampah sisa botol air mineral air tdk jernih kaporit kebanyakan sepet mulut gigi terkikistehel tajam anakku terlukaberlumut kolam yg ban susah berenang
saran tolong di pisahkan kolam yg bs digunakan khusus berenang dan memakai ban apalagi ban besar yg peruntukanx buat wahana seluncuran	saran tolong pisahkan kolam yg bs khusus berenang memakai ban ban yg peruntukanx wahana seluncuran
tempat wisata keluarga yang cukup ramai waktu berkunjung di awal pekan fasilitas yang tersedia kolam renang tempat makan aula pertemuan dll	wisata keluarga ramai berkunjung pekan fasilitas tersedia kolam renang makan aula pertemuan dll

e. *Stemming*

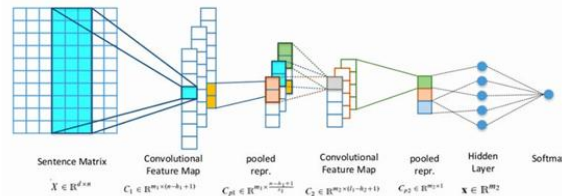
Stemming adalah proses menghilangkan imbuhan atau akhiran kata untuk menghasilkan bentuk dasar dari kata. Tujuannya untuk mengurangi kata – kata bentuk dasarnya agar variasi kata yang serupa dapat dianggap sama. Contohnya, kata “berlari”, “lari”, dan “larian” akan disistem menjadi kata dasar “lar”.

Tabel 6. Tahap Stemming

Sebelum	Sesudah
tempatnya bersih kamar mandinya bersih kolam renang ada tempat foto juga ada terus ada tempat anak bermain	tempat bersih kamar mandi bersih kolam renang foto anak main
sangat cocok untuk mengajak keluarga kesini karena kolam renang anak anak dan dewasanya luas kebersihannya juga terjaga	cocok ajak keluarga kesini kolam renang anak anak dewasa luas bersih jaga
tempat bagus luas cuma kurangnya pemanfaatan lahan kebun nya padahal bisa dijadikan lahan untuk belajar setiap tanaman yg ada disana	bagus luas kurang manfaat lahan kebun jadi lahan ajar tanam yg sana

3.4 Model Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah jenis arsitektur jaringan saraf tiruan yang secara khusus dirancang untuk memproses data, seperti teks. Model ini terdiri dari lapisan – lapisan yang disusun secara hierarkis yang mampu mengekstraksi fitur – fitur penting dari data input.



Gambar 4. Arsitektur CNN[14]

CNN adalah sebuah konstruk matematika yang biasanya disusun oleh 3 tipe layer, yaitu convolution, pooling, dan layer yang sepenuhnya terhubung. Dua layer pertama, convolution dan pooling melakukan ekstraksi fitur, sedangkan layer ketiga, menempatkan fitur yang terekstraksi menjadi hasil akhir keluaran, seperti klasifikasi[15].

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from sklearn.model_selection import train_test_split
labels = [1, 0, 2, 1, 0, 1, 1, 0, 1, 0, 1, 2, 1, 0, 1, 0, 0, 1, 1, 2, 1, 0, 2, 1, 0, 1, 1, 1, 0, 0] #
1: positif, 0: negatif, 2: netral
# Import dataset
df = pd.read_excel("DATAKU_STEM.xlsx", sheet_name='Sheet1')
df
# Preprocessing
reviews = df['ULASAN']
labels = df['LABEL_N']
reviews
# Filter out non-string elements
non_string_indices = [i for i, item in enumerate(reviews) if not isinstance(item, str)]
if non_string_indices:
    print("Non-string elements found at indices:", non_string_indices)
    reviews = [item for i, item in enumerate(reviews) if i not in non_string_indices]
    labels = labels.drop(non_string_indices)
print("Non-string element at index 1890:", reviews[1890])
reviews = [item for i, item in enumerate(reviews) if i not in non_string_indices]
labels = df['LABEL_N']
tokenizer = Tokenizer(num_words=10000)
tokenizer.fit_on_texts(reviews)
sequences = tokenizer.texts_to_sequences(reviews)
word_index = tokenizer.word_index
data = pad_sequences(sequences, maxlen=100)
# Memisahkan data menjadi data pelatihan dan data validasi
data_train, data_val, labels_train, labels_val = train_test_split(data, labels,
test_size=0.3, random_state=24)
train_dataset = tf.data.Dataset.from_tensor_slices((data_train, labels_train))
```

```

val_dataset = tf.data.Dataset.from_tensor_slices((data_val, labels_val))
# Define and compile the CNN model
model = keras.Sequential()
model.add(layers.Embedding(len(word_index) + 1, 100, input_length=100))
model.add(layers.Conv1D(128, 5, activation='relu'))
model.add(layers.GlobalMaxPooling1D())
model.add(layers.Dense(64, activation='relu'))
model.add(layers.Dense(3, activation='softmax')) # Output layer dengan 3 kelas
model.compile(optimizer='adam', loss='sparse_categorical_crossentropy',
metrics=['accuracy'])
model.compile(optimizer='adam', loss='sparse_categorical_crossentropy',
metrics=['accuracy'])
model.fit(train_dataset.batch(64), validation_data=val_dataset.batch(64), epochs=100)
# Lakukan prediksi pada ulasan baru
new_reviews = ["bagus makassar", "nyesel"]
new_sequences = tokenizer.texts_to_sequences(new_reviews)
new_data = pad_sequences(new_sequences, maxlen=100)
predictions = model.predict(new_data)
# Menerjemahkan prediksi
sentiment_labels = ["Positif", "Negatif", "Netral"]
for i, prediction in enumerate(predictions):
    predicted_label = sentiment_labels[np.argmax(prediction)]
    print(f"Ulasan: {new_reviews[i]} - Prediksi: {predicted_label}")
# Make predictions on validation data
predictions = model.predict(data_val)
predicted_labels = np.argmax(predictions, axis=1)
# Iterate through data_val and predicted_labels
for i, (text, true_label, predicted_label) in enumerate(zip(data_val, labels_val,
predicted_labels)):
    # Menghilangkan token padding dari teks
    text = ''.join([word for word in tokenizer.sequences_to_texts([text])[0].split() if word !=
'0'])
    print(f"Index: {i}")
    print(f"Text: {text}")
    print(f"True Label: {true_label}")
    print(f"Predicted Label: {predicted_label}")
    print("-----")

```

a. *Embedding Layer*

'layers.Embedding(len(word_index) + 1, 100, input_length=100)': Ini adalah lapisan embedding yang bertanggung jawab untuk mengubah kata-kata dalam teks menjadi vektor ruang fitur. Parameter pertama (len(word_index) + 1) adalah ukuran kosakata, yaitu jumlah kata unik dalam teks ditambah satu untuk menangani kata-kata yang tidak dikenal

b. *Convolutional Layer*

layers.Conv1D(128, 5, activation='relu)': Ini adalah lapisan konvolusi satu dimensi. Conv1D digunakan untuk mengekstrak fitur dari data urutan seperti teks. Fungsi aktivasi 'relu' (Rectified Linear Unit) diterapkan setelah konvolusi.

c. *Global Max Pooling Layer*

layers.GlobalMaxPooling1D(): Ini adalah lapisan pemadatan global yang mengambil nilai maksimum dari fitur-fitur yang diekstraksi oleh konvolusi di seluruh urutan. Ini membantu dalam merampingkan fitur untuk pengolahan lebih lanjut.

d. *Dense Layer (Hidden Layer)*

layers.Dense(64, activation='relu)': Ini adalah lapisan terhubung penuh (fully connected) dengan 64 unit neuron dan menggunakan fungsi aktivasi ReLU.

ReLU adalah fungsi aktivasi yang paling umum digunakan dalam jaringan saraf buatan saat ini. Fungsi ini didefinisikan sebagai $f(x) = \max(0, x)$, yang berarti

bahwa nilai keluaran adalah nol jika inputnya negatif dan sama dengan inputnya jika inputnya positif.

e. *Output Layer*

'layers.Dense(3, activation='softmax')': Ini adalah lapisan output. Lapisan ini memiliki 3 unit neuron yang mewakili kelas-kelas yang mungkin. Fungsi aktivasi softmax digunakan di sini untuk menghasilkan probabilitas untuk setiap kelas.

3.5 Pengujian Metode *Convolutional Neural Network (CNN)*

Pengujian metode Convolutional Neural Network (CNN) dalam konteks analisis sentimen atau klasifikasi teks mengacu pada proses evaluasi dan penilaian kinerja model CNN yang telah dibangun. Pengujian metode CNN penting untuk memastikan bahwa model yang dibangun dapat melakukan prediksi dengan akurat dan dapat diandalkan pada data baru yang tidak pernah dilihat sebelumnya. Pada tahap ini yaitu pengujian metode Convolutional Neural Network untuk menampilkan tingkat akurasi.

```
model.compile(optimizer='adam',loss='sparse_categorical_crossentropy', metrics=['accuracy'])
```

```
model.fit(train_dataset.batch(32),validation_data=val_dataset.batch(32), epochs=100)
```

Dalam kode tersebut, kita melakukan kompilasi model dan melatihnya menggunakan dataset pelatihan dan validasi. Berikut adalah penjelasan lebih rinci :

- a. Kompilasi Model : Pada langkah ini, model disusun untuk pelatihan dengan menentukan optimizer, fungsi loss, dan metrik evaluasi yang akan digunakan.
- b. Pelatihan Model : Model dilatih menggunakan fungsi fit().

```
# Lakukan prediksi pada ulasan baru
```

```
new_reviews = ["bagus makassar", "nyesel"]
new_sequences = tokenizer.texts_to_sequences(new_reviews)
new_data = pad_sequences(new_sequences, maxlen=100)
predictions = model.predict(new_data)
```

Dalam kode tersebut, dilakukan prediksi sentimen pada ulasan baru menggunakan model CNN yang telah dilatih sebelumnya. Berikut adalah penjelasan langkah-langkahnya :

1. Membuat Ulasan Baru : Dua ulasan baru disiapkan dalam variabel new_reviews.
2. Mengonversi Ulasan Baru Menjadi Sequences : Ulasan baru dikonversi menjadi sequences menggunakan objek tokenizer yang telah dibuat sebelumnya. Sequences ini merepresentasikan urutan angka-angka berdasarkan indeks kata-kata dari vocabulary yang telah dibangun pada data pelatihan.
3. Padding Sequences : Sequences yang diperoleh dari ulasan baru dipad agar memiliki panjang yang sama dengan panjang maksimum yang telah ditentukan sebelumnya (maxlen=100) untuk konsistensi dengan input yang digunakan dalam pelatihan model.
4. Melakukan Prediksi : Data baru yang sudah diproses kemudian digunakan sebagai input untuk melakukan prediksi sentimen menggunakan model CNN yang telah dilatih sebelumnya. Fungsi predict() digunakan untuk menghasilkan probabilitas prediksi untuk setiap kelas sentimen (positif, negatif, netral) berdasarkan ulasan yang diberikan.

```
# Menerjemahkan prediksi
```

```
sentiment_labels = ["Positif", "Negatif", "Netral"]
for i, prediction in enumerate(predictions):
    predicted_label =
    sentiment_labels[np.argmax(prediction)]
    print(f"Ulasan: {new_reviews[i]} - Prediksi:
    {predicted_label}")
```

Dalam kode tersebut, dilakukan penerjemahan hasil prediksi probabilitas menjadi label sentimen yang lebih mudah dipahami. Berikut adalah penjelasan langkah-langkahnya :

1. Sentimen Labels : Sebuah list sentiment_labels didefinisikan untuk menentukan label yang sesuai dengan setiap kelas sentimen. Indeks dalam list ini harus sesuai

dengan indeks kelas sentimen dalam urutan probabilitas yang diberikan oleh model.

2. Iterasi Melalui Prediksi : Melalui loop for, setiap prediksi dalam hasil prediksi (predictions) diakses secara berurutan.
3. Menerjemahkan Prediksi : Prediksi probabilitas untuk setiap kelas sentimen diinterpretasikan. Indeks dengan probabilitas tertinggi diambil dengan menggunakan np.argmax(prediction) dan kemudian digunakan untuk mengambil label sentimen yang sesuai dari list sentiment_labels
4. Mencetak Hasil : Hasil prediksi yang telah diterjemahkan dicetak untuk setiap ulasan baru bersama dengan teks ulasan aslinya.

```
# Make predictions on validation data
predictions = model.predict(data_val)
predicted_labels = np.argmax(predictions, axis=1)
# Iterate through data_val and predicted_labels
for i, (text, true_label, predicted_label) in
enumerate(zip(data_val, labels_val, predicted_labels)):
# Menghilangkan token padding dari teks
text = ''.join([word for word in
tokenizer.sequences_to_texts([text])[0].split() if
word != '0'])
print(f"Index: {i}")
print(f"Text: {text}")
print(f"True Label: {true_label}")
print(f"Predicted Label: {predicted_label}")
print("-----")
# Create DataFrame from data_val, labels_val, and
predicted labels
df_results = pd.DataFrame({'Text': [' '.join([word for
word in
tokenizer.sequences_to_texts([review])[0].split() if
word != '0']) for review in data_val], 'True Label':
labels_val, 'Predicted Label': predicted_labels})
# Save DataFrame to Excel file
df_results.to_excel('hasil_prediksi.xlsx', index=False)
```

Hasil prediksi kemudian ditampilkan dalam bentuk teks dan disimpan ke dalam file Excel. Berikut adalah penjelasan langkah-langkahnya :

1. Melakukan Prediksi : Prediksi dilakukan pada data validasi (data_val) menggunakan model CNN yang telah dilatih sebelumnya. Probabilitas prediksi untuk setiap kelas sentimen diperoleh menggunakan fungsi predict(), dan label prediksi yang sesuai diambil dengan menggunakan np.argmax() dengan parameter axis=1.
2. Iterasi Melalui Data Validasi dan Prediksi : Melalui loop for, setiap data validasi, label asli, dan label prediksi diakses secara berurutan. Selama iterasi ini, token padding dihilangkan dari teks dengan memeriksa kata-kata yang bukan token padding.
3. Mencetak Hasil Prediksi : Setiap data validasi beserta label asli dan prediksi dicetak ke layar.
4. Membuat DataFrame : DataFrame df_results dibuat dengan kolom-kolom 'Text', 'True Label', dan 'Predicted Label' yang berisi teks ulasan, label asli, dan label prediksi, masing-masing.
5. Menyimpan DataFrame ke Excel : DataFrame df_results disimpan ke dalam file Excel dengan nama 'hasil_prediksi.xlsx' tanpa menyertakan indeks.

4. Kesimpulan

Dari hasil penelitian yang telah dilakukan adalah bahwa dari pengujian dataset sebanyak 4500 ulasan tentang Tempat Wisata Makassar menggunakan metode CNN dengan tiga kelas sentimen: positif, negatif, dan netral. Hasilnya, akurasi tertinggi dicapai dengan nilai akurasi pelatihan sebesar 95% dan akurasi validasi sebesar 73%. Hal ini

menunjukkan bahwa metode CNN mampu memprediksi sentimen dengan baik, meskipun akurasi validasinya sedikit lebih rendah. Kesimpulannya, CNN dapat diandalkan untuk analisis sentimen ulasan wisata, memberikan kontribusi positif dalam evaluasi kualitas layanan. Penelitian ini juga menyarankan integrasi model ke dalam aplikasi nyata, seperti aplikasi pemesanan hotel atau wisata, untuk menguji kelayakan model dalam skenario dunia nyata.

Referensi

- [1] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," vol. 2, no. 1, pp. 32–41, 2017.
- [2] I. Asriandy, "Strategi Pengembangan Obyek Wisata Air TERJUN BISSAPU DI KABUPATEN BANTAENG," *Univ. Hasanudin*, p. 82, 2016, [Online]. Available: <https://core.ac.uk/download/pdf/77625485.pdf>
- [3] B. A. I. G. N. Agung, "Implementasi Deep Learning untuk Image Clasification menggunakan Arsitektur Convolutional Neural Network (CNN) pada Citra Sampah Hotel (Studi Kasus: Hotel ...," 2023, [Online]. Available: http://eprints.unram.ac.id/id/eprint/41624%0Ahttp://eprints.unram.ac.id/41624/2/JURNAL_TUGAS_AKHIR_-_F1D019037_-_I.G.N.A._BAYU_ADHIPRAMANA_-_TEKNIK_INFORMATIKA.pdf
- [4] A. K. QORITA, "Analisis Sentimen Berbasis Aspek Pada Ulasan Tempat Wisata Diy," 2022, [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/39202%0Ahttps://dspace.uui.ac.id/bitstream/handle/123456789/39202/18523214.pdf?sequence=1>
- [5] A. Peryanto, A. Yudhana, and R. Umar, "Rancang Bangun Klasifikasi Citra Dengan Teknologi Deep Learning Berbasis Metode Convolutional Neural Network," *Format J. Ilm. Tek. Inform.*, vol. 8, no. 2, p. 138, 2020, doi: 10.22441/format.2019.v8.i2.007.
- [6] Muhammad Haris Diponegoro, Sri Suning Kusumawardani, and Indriana Hidayah, "Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 131–138, 2021, doi: 10.22146/jnteti.v10i2.1417.
- [7] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [8] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [9] J. C. Mestika, M. O. Selan, and M. I. Qadafi, "Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python," vol. 99, no. 99, pp. 216–219, 2022.
- [10] Z. Muhamad, "Jurnal Teknik Informatika Atmaluhur," *J. Tek. Inform. Atmaluhur*, vol. 6, no. 1, p. 40, 2018.
- [11] R. Rosaly and A. Prasetyo, "Flowchart Beserta Fungsi dan Simbol-Simbol," *J. Chem. Inf. Model.*, vol. 2, no. 3, pp. 5–7, 2020.
- [12] Y. Riadi Silitonga, Munawar, and I. Noor Hapsari, "Analisis Dan Penerapan Datamining Untuk Mendeteksi Berita Palsu (Fake News) Pada Social Media Dengan Memanfaatkan Modul Scikit Learn," *Undergrad. Theses Inf. Syst.*, 2019.
- [13] I. Pelham, "Erd2," *Secret. Pathw.*, vol. 5, pp. 135–135, 2023, doi: 10.1093/oso/9780198599425.003.0085.
- [14] D. T. Hermanto, A. Setyanto, and E. T. Luthfi, "Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online," *Creat. Inf. Technol. J.*, vol. 8, no. 1, p. 64, 2021, doi: 10.24076/citec.2021v8i1.264.
- [15] A. F. Hidayatullah and R. A. N. Nayoan, "Analisis Sentimen Berbasis Fitur pada Ulasan Tempat Wisata Menggunakan Metode Convolutional Neural Network(CNN)," 2019, [Online]. Available: www.cnet.com.