

SKRIPSI

ANALISIS SENTIMEN WISATA KULINER DAN BELANJA DI  
KOTA BANDUNG DENGAN *DEEP LEARNING NEURAL  
NETWORK*



Mario Bagus Prasetya

NPM: 6181801071

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2023



**UNDERGRADUATE THESIS**

**SENTIMENT ANALYSIS CULINARY AND SHOPPING  
TOURISM IN THE CITY OF BANDUNG USING DEEP  
LEARNING NEURAL NETWORK**



**Mario Bagus Prasetya**

**NPM: 6181801071**

**DEPARTMENT OF INFORMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2023**

# LEMBAR PENGESAHAN

## ANALISIS SENTIMEN WISATA KULINER DAN BELANJA DI KOTA BANDUNG DENGAN *DEEP LEARNING NEURAL NETWORK*

Mario Bagus Prasetya

NPM: 6181801071

Bandung, 5 Juli 2023

Menyetujui,

Pembimbing

Digitally signed  
by Cecilia Esti  
Nugraheni

Dr.rer.nat. Cecilia Esti Nugraheni

Ketua Tim Penguji  
Digitally signed  
by Rosa de Lima  
E. Padmowati

Rosa De Lima, M.T.

Anggota Tim Penguji  
Digitally signed  
by Keenan  
Adiwijaya Leman

Keenan Adiwijaya Leman, M.T.

Mengetahui,

Ketua Program Studi  
Digitally signed  
by Mariskha Tri  
Adithia

Mariskha Tri Adithia, P.D.Eng

## PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

### **ANALISIS SENTIMEN WISATA KULINER DAN BELANJA DI KOTA BANDUNG DENGAN *DEEP LEARNING NEURAL NETWORK***

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,  
Tanggal 5 Juli 2023



Mario Bagus Prasetya  
NPM: 6181801071

## ABSTRAK

Kota Bandung adalah kota yang memiliki potensi pariwisata yang sangat tinggi di Indonesia. Menurut [Litbang Kompas dan Kementerian Pariwisata tahun 2016](#), Bandung unggul dalam beberapa aspek Indeks Pariwisata Indonesia. Berdasarkan data dari [Badan Pusat Statistik Provinsi Jawa Barat](#), jumlah kunjungan wisatawan ke objek wisata yang ada di kota Bandung mencapai 5.864.721 ditahun 2018. Pada tahun 2019 mencapai 2.442.250 dan pada tahun 2021 mencapai 393.223. Kunjungan yang banyak dari wisatawan ini memberikan dampak terhadap industri wisata kuliner dan belanja di Kota Bandung. Kementerian Pariwisata telah menetapkan kota Bandung sebagai salah satu destinasi unggulan wisata kuliner dan belanja (wiskulbel) di Indonesia.

Walaupun kota Bandung merupakan salah satu destinasi unggulan wisata kuliner dan belanja, tentu masih memiliki kekurangan dan permasalahan. Kekurangan dan permasalahan ini dapat dilihat dari opini wisatawan yang terdapat pada *review* diinternet. Penelitian ini dilakukan untuk menganalisis sentimen yang terdapat diinternet terhadap wisata kuliner dan belanja di Kota Bandung serta membuat model klasifikasi menggunakan *Deep Learning*.

Pada penelitian ini, dilakukan tahapan-tahapan dari *Text Mining* untuk analisis sentimen. Tahapan ini meliputi pengumpulan data, *text preprocessing*, *feature extraction*, *pattern discovery* dan evaluasi model. Tahapan pertama, yaitu mengumpulkan data menggunakan *tools webscraping*, yaitu *WebHarvy*. Data bersumber dari 2 *website*, yaitu [TripAdvisor](#) dan [Google Review](#). Selanjutnya adalah melakukan *preprocessing* teks yang meliputi *case folding*, *tokenization*, *stopword removing*, *stemming* dan pengubahan kata slang. Tahap selanjutnya adalah melakukan pelabelan data menggunakan salah satu teknik analisis sentimen, yaitu pendekatan secara *lexicon-based*. Data yang sudah dilabeli akan divisualisasikan untuk mendapatkan *insight* dari data. Tahapan selanjutnya adalah pembuatan model. Diawali dengan mengubah teks menjadi vektor menggunakan *Word Tokenizer* dari *library keras*. Selanjutnya model klasifikasi sentimen dibuat menggunakan algoritma *Deep Learning*. Model ini selanjutnya dievaluasi dan diimplementasikan kedalam perangkat lunak berbasis *website*.

Data ulasan wisata kuliner dan belanja di kota Bandung yang dikumpulkan berjumlah 1673 ulasan. Berdasarkan penelitian yang sudah dilakukan, data berlabel positif merupakan data terbanyak dengan persentase 54%. Data berlabel negatif dengan presentase 32% dan data netral paling sedikit yaitu 14%.

Hasil evaluasi model klasifikasi sentimen menggunakan *Deep Learning* dengan data 3 label sentimen (positif, negatif dan netral) mendapatkan akurasi tertinggi menggunakan model Bidirectional Long-Short Term Memory dengan akurasi data validasi 70.72%. Model kedua adalah CNN dengan akurasi data validasi 68.13%. Model dengan akurasi data validasi terendah adalah Neural Network biasa dengan akurasi data validasi 54%.

Hasil evaluasi model klasifikasi sentimen menggunakan *Deep Learning* dengan data dengan 2 label sentimen (positif dan negatif) mendapatkan akurasi tertinggi menggunakan model Bidirectional Long-Short Term Memory dengan akurasi data validasi 81.67%. Model CNN mendapatkan akurasi data validasi sebesar 79.35%.

**Kata-kata kunci:** Analisis Sentimen, Wisata Kuliner dan Belanja di Kota Bandung, Deep Learning, Jaringan Saraf Tiruan



## ABSTRACT

The city of Bandung is a city that has very high tourism potential in Indonesia. According to [Kompas Research and Development and the Ministry of Tourism](#) in 2016, Bandung excels in several aspects of the Indonesian Tourism Index. Based on data from [Central Bureau of Statistics of West Java Province](#), the number of tourist visits to tourist objects in the city of Bandung reached 5,864,721 in 2018. In 2019 it reached 2,442,250 and in 2021 it reached 393,223. The many visits of these tourists have an impact on the culinary and shopping tourism industry in the city of Bandung. The Ministry of Tourism has designated the city of Bandung as one of the leading destinations for culinary and shopping tourism (wiskulbel) in Indonesia.

Even though the city of Bandung is one of the leading destinations for culinary and shopping tourism, of course it still has shortcomings and problems. The shortcomings and problems can be seen from opinions of tourists found in reviews on the internet. This research was conducted to analyze the sentiment found on the internet toward culinary and shopping tourism in the city of Bandung and build classification model using Deep Learning.

In this research, the stages of Text Mining were carried out for sentiment analysis. This stages include data collection, text preprocessing, feature extraction, pattern discovery and model evaluation. The first stage is collecting data using web scraping tools, namely WebHarvy. The data sourced from 2 websites, namely [TripAdvisor](#) and Google Reviews. The next step is to do text preprocessing which includes case folding, tokenization, removing stopwords, stemming and changing slang words. The next stage is to label the data using one of the sentiment analysis technique, the lexicon-based approach. Data that has been labeled will be visualized to get insight from the data. The next stage is to build model. Beginning with transform text to vector using Word Tokenizer from keras library. The pattern discovery stage using Deep Learning algorithm to build sentiment classification model. Then this model evaluated and implemented into website-based software.

The review data of culinary and shopping tourism in the City of Bandung that were collected 1673 reviews in total. Based on research that has been done, data labeled positive is the most data with percentage of 54%. Data labeled negative with a percentage of 32% and neutral data the least 14%.

Result of evaluate sentiment classification model using Deep Learning with 3 sentiment label(positive, negative and neutral) get highest accuracy using Bidirectional Long-Short Term memory with validation accuracy is 70.72%. The second model is CNN with validation data accuracy is 68.13%. Model with lowest validation accuracy is ordinary neural network with data validation accuracy 54%.

Result of evaluate sentiment classification model using Deep Learning with 2 sentiment label(positive and negative) get highest accuracy using Bidirectional Long-Short Term memory with validation accuracy is 81.67%. The CNN model get validation data accuracy is 79.35%.

**Keywords:** Sentiment Analysis, Culinary and Shopping Tourism in city of Bandung, Deep Learning, Artificial Neural Network



*Skripsi ini dipersembahkan untuk keluarga dan juga diri sendiri.*



## KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Tuhan Yesus Kristus karena berkat-Nya penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Wisata Kuliner dan Belanja di Kota Bandung dengan *Deep Learning Neural Network*”. Dalam menyelesaikan skripsi ini, penulis tidak lupa juga penulis menyampaikan terimakasih kepada:

1. Orangtua dan saudara-saudara penulis yang memberi nasihat, dukungan dan doa selama proses pengerjaan skripsi ini.
2. Ibu Dr.rer.nat. Cecilia Esti Nugraheni selaku dosen pembimbing yang telah memberikan bimbingan, waktu dan arahan selama proses pengerjaan skripsi ini.
3. Ibu Rosa De Lima, M.T. dan Bapak Keenan Adiwijaya Leman, M.T. selaku dosen penguji yang telah memberikan kritik dan saran terhadap skripsi ini.
4. Kepada teman-teman penulis semasa kuliah Michio, Alvino, Yalvi, Gian, Giovanni, David dan seluruh teman-teman jurusan Informatika angkatan 2018.
5. Semua orang yang tidak dapat disebutkan satu persatu yang telah membantu dan mendukung penulis selama penulisan skripsi.

Akhir kata, penulis ingin mengucapkan terimakasih kepada pembaca dokumen skripsi ini. Penulis berharap skripsi ini dapat berguna bagi pembaca. Penulis memohon maaf bila terdapat kesalahan dalam penulisan maupun metode yang ditulis dalam skripsi ini.

Bandung, Juli 2023

Penulis



# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>xv</b>
<b>DAFTAR ISI</b>	<b>xvii</b>
<b>DAFTAR GAMBAR</b>	<b>xxi</b>
<b>DAFTAR TABEL</b>	<b>xxiii</b>
<b>DAFTAR KODE PROGRAM</b>	<b>xxv</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan . . . . .	3
1.4 Batasan Masalah . . . . .	3
1.5 Metodologi . . . . .	3
1.6 Sistematika Pembahasan . . . . .	4
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 Text Mining . . . . .	5
2.1.1 Text Preprocessing . . . . .	5
2.1.2 Feature Selection dan Extraction . . . . .	7
2.1.3 Pattern Discovery . . . . .	8
2.2 Analisis Sentimen . . . . .	8
2.3 Deep Learning . . . . .	8
2.3.1 Activation Function . . . . .	9
2.3.2 Gradient Descent, Learning Rate, Epoch, dan Batch Size[1] . . . . .	11
2.3.3 Adam Optimization [2] . . . . .	12
2.3.4 Dropout [3] . . . . .	12
2.3.5 Loss Function . . . . .	13
2.3.6 Arsitektur Neural Network . . . . .	13
2.4 Recurrent Neural Network . . . . .	15
2.5 <i>Long Short-Term Memory</i> [4] . . . . .	16
2.6 <i>Bidirectional Long Short-Term Memory</i> . . . . .	18
2.7 <i>Convolutional Neural Network</i> . . . . .	19
2.7.1 Lapisan Konvolusional[5] . . . . .	19
2.7.2 <i>Pooling Layer</i> [5] . . . . .	20
2.8 Evaluasi model . . . . .	20
2.9 Visualisasi Data . . . . .	21
2.9.1 Bar Plot [6] . . . . .	21
2.9.2 Pie Chart [6] . . . . .	22
2.9.3 Word Cloud . . . . .	22
2.10 Wisata Kuliner dan Belanja . . . . .	23

<b>3</b>	<b>ALUR Pengerjaan dan Eksplorasi Tools</b>	<b>25</b>
3.1	Alur Pengerjaan	25
3.2	<i>Tool dan Library yang Digunakan</i>	26
3.3	Ekplorasi	26
3.3.1	Ekplorasi Pengumpulan Data	26
3.3.2	Ekplorasi Preprocessing Data	28
3.3.3	Ekplorasi Pelabelan Data	30
3.3.4	Ekstraksi Fitur Menggunakan <i>Word Tokenizer</i>	32
3.3.5	Ekplorasi <i>Deep Learning</i> dengan Fitur <i>Word Tokenizer</i> dan <i>Convolutional Neural Network</i>	32
3.3.6	Ekplorasi <i>Deep Learning</i> dengan Fitur <i>Word Tokenizer</i> dan <i>Bidirectional LSTM</i>	34
3.3.7	Ekplorasi <i>Deep Learning</i> dengan Fitur <i>Word Tokenizer</i> dan <i>Artificial Neural Network</i>	34
3.3.8	Ekplorasi Pembuatan Perangkat Lunak	35
<b>4</b>	<b>Eksplorasi Data dan Pembuatan Model</b>	<b>39</b>
4.1	Deskripsi Dataset	39
4.2	<i>Preprocessing Data</i>	40
4.2.1	<i>Case Folding</i> , Menghapus Angka dan Menghapus Tanda Baca	40
4.2.2	Ubah Kata Slang dan <i>Stemming</i>	40
4.2.3	<i>Stopword Remove</i>	41
4.3	Pelabelan Data	41
4.4	Ekplorasi dan Visualisasi Data	42
4.4.1	Visualisasi Label Sentimen	42
4.4.2	Visualisasi Kata yang Ada Dalam <i>Review</i>	42
4.5	Ekstraksi Fitur	47
4.6	Permodelan dengan <i>Deep Learning Artificial Neural Network</i>	47
4.6.1	Percobaan Jumlah <i>Hidden Layer</i>	48
4.6.2	Percobaan Fungsi Aktivasi yang Sesuai	48
4.6.3	Percobaan Jumlah Epoch Terhadap Nilai Akurasi	48
4.6.4	Evaluasi Model <i>Deep Learning Neural Network</i>	49
4.7	Pembuatan Model Menggunakan Arsitektur CNN	50
4.7.1	Percobaan Jumlah <i>Epoch</i>	50
4.7.2	Pecobaan Jumlah dan Dimensi <i>Hidden Layer</i>	50
4.7.3	Evaluasi Model <i>Convolutional Neural Network</i>	50
4.8	Pembuatan Model Menggunakan Arsitektur <i>Bidirectional LSTM</i>	52
4.8.1	Percobaan Jumlah Epoch	52
4.8.2	Pecobaan Jumlah dan Dimensi <i>Hidden Layer</i>	52
4.8.3	Evaluasi Model <i>Bidirectional LSTM</i>	53
4.9	Ekstraksi Fitur 2 Label Sentimen	54
4.10	Pembuatan Model Menggunakan Arsitektur CNN dengan 2 Label Sentimen	55
4.10.1	Percobaan Jumlah <i>Epoch</i>	55
4.10.2	Evaluasi Model CNN 2 Label	55
4.11	Pembuatan Model Menggunakan Arsitektur <i>Bidirectional LSTM</i> dengan 2 Label Sentimen	57
4.11.1	Percobaan Jumlah <i>Epoch</i>	57
4.11.2	Evaluasi Model BiLSTM 2 Label	57
4.12	Percobaan Menggunakan Regularization L1 dan L2	58
4.13	Kesimpulan Analisis Sentimen Data Wiskulbel di Bandung dan Sekitarnya	59
<b>5</b>	<b>Implementasi Perangkat Lunak</b>	<b>61</b>

5.1	Mockup Perangkat Lunak . . . . .	61
5.1.1	<i>Mockup</i> Halaman Prediksi sentimen . . . . .	61
5.1.2	<i>Mockup</i> Halaman Eksplorasi Data . . . . .	62
5.2	Hasil Implementasi Perangkat Lunak . . . . .	63
5.2.1	Alur Program Pada Halaman Prediksi Sentimen . . . . .	63
5.2.2	Halaman Prediksi Sentimen . . . . .	64
5.2.3	Halaman Eksplorasi Data Analisis . . . . .	66
5.3	Pengujian Perangkat Lunak . . . . .	66
5.3.1	Pengujian Halaman Prediksi Sentimen Input 1 Teks . . . . .	67
5.3.2	Pengujian Halaman Prediksi Sentimen Input File . . . . .	67
<b>6</b>	<b>KESIMPULAN DAN SARAN</b>	<b>69</b>
6.1	Kesimpulan . . . . .	69
6.2	Saran . . . . .	69
	<b>DAFTAR REFERENSI</b>	<b>71</b>
	<b>A KODE PROGRAM</b>	<b>73</b>
	<b>B CONTOH DATA</b>	<b>87</b>





## DAFTAR GAMBAR

2.1	Neuron biologis(kiri) dan artificial neuron model	9
2.2	<i>Linear function</i>	9
2.3	<i>Sigmoid activation function</i>	10
2.4	<i>Tanh activation function</i>	10
2.5	<i>ReLu activation function</i>	11
2.6	Grafik <i>gradient descent</i>	11
2.7	<b>Kiri:</b> <i>Artificial Neural Network</i> tanpa <i>dropout</i> . <b>Kanan:</b> Contoh dari <i>Artificial Neural Network</i> yang diterapkan <i>dropout</i> sehingga beberapa neuron tidak aktif	13
2.8	<i>Fully connected neural network</i>	14
2.9	Ilustrasi dari arsitektur RNN	15
2.10	Gambar dari LSTM <i>Cline</i>	16
2.11	Gambar dari LSTM bagian <i>forget gate</i>	17
2.12	Gambar dari LSTM bagian <i>input gate</i>	17
2.13	Gambar dari LSTM bagian perubahan $C_t$	18
2.14	Gambar dari LSTM bagian <i>output gate</i>	18
2.15	Gambar arsitektur <i>Bidirectional LSTM</i>	19
2.16	Gambar dari convolutional layer	19
2.17	Gambar dari <i>pooling layer</i>	20
2.18	Contoh bar plot horizontal	22
2.19	Contoh bar plot vertikal	22
2.20	Contoh dari <i>pie chart</i>	22
2.21	Contoh dari <i>word cloud</i>	23
3.1	Diagram alir pengerjaan penelitian	25
3.2	Ulasan diTripadvisor	27
3.3	Ulasan diGoogle Review	27
3.4	Hasil webscraping	27
3.5	Contoh proses pelabelan data	31
3.6	Plot akurasi CNN	33
3.7	Plot <i>loss</i> CNN	33
3.8	Plot akurasi BiLSTM	34
3.9	Plot <i>loss</i> BiLSTM	34
3.10	Plot akurasi NN	35
3.11	Plot <i>loss</i> NN	35
3.12	Contoh halaman dengan menampilkan elemen kedalam website	36
3.13	Contoh halaman dengan menampilkan <i>widget-widget</i> yang ada diStreamlit	37
4.1	<i>Pie chart</i> persentase label sentimen	42
4.2	<i>WordCloud</i> data keseluruhan	43
4.3	Barplot data keseluruhan	43
4.4	Barplot kata berkategori positif	44
4.5	Barplot kata bersentimen positif	44
4.6	Barplot bigrams bersentimen positif	45

4.7	Barplot kata berkategori negatif	45
4.8	Barplot kata bersentimen negatif	46
4.9	Barplot <i>bigrams</i> bersentimen negatif	46
4.10	Barplot kata berkategori netral	47
4.11	Confusion matrix 3 label sentimen menggunakan <i>Neural Network</i> biasa	49
4.12	plot akurasi CNN	51
4.13	plot loss CNN	51
4.14	Confusion matrix 3 label sentimen menggunakan CNN	51
4.15	Plot akurasi BiLSTM	53
4.16	Plot loss BiLSTM	53
4.17	Confusion matrix 3 label sentimen menggunakan BiLSTM	54
4.18	Plot akurasi CNN 2 label	56
4.19	Plot loss CNN 2 label	56
4.20	Confusion matrix 2 label sentimen menggunakan CNN	56
4.21	Plot akurasi CNN 2 label	57
4.22	Plot loss CNN 2 label	57
4.23	Confusion matrix 2 label sentimen menggunakan CNN	58
4.24	Plot akurasi BiLSTM dengan L1	58
4.25	Plot loss BiLSTM dengan L1	58
4.26	Plot akurasi BiLSTM dengan L2	59
4.27	Plot loss BiLSTM dengan L2	59
5.1	Mockup halaman prediksi sentimen satu teks.	61
5.2	Mockup halaman prediksi sentimen dengan input file.	62
5.3	Mockup halaman eksplorasi data.	63
5.4	Halaman prediksi sentimen satu teks.	64
5.5	Halaman hasil prediksi sentimen satu teks.	65
5.6	Halaman prediksi sentimen satu file.	65
5.7	Halaman hasil prediksi sentimen satu file.	66
5.8	Halaman eksplorasi data analisis.	66
5.9	Halaman hasil pengujian jika input berbahasa inggris.	67
5.10	Halaman hasil pengujian jika input berekstensi selain excel atau csv.	67

## DAFTAR TABEL

2.1	Contoh dari <i>tokenization</i> . . . . .	5
2.2	Tabel contoh <i>case folding</i> . . . . .	6
2.3	Tabel contoh penghapusan tanda baca . . . . .	6
2.4	Tabel awalan-akhiran . . . . .	7
2.5	Tabel contoh <i>Bag of Words</i> . . . . .	7
3.1	Hasil dari <i>Text Preprocessing</i> . . . . .	30
3.2	Hasil dari pelabelan . . . . .	31
4.1	Contoh data yang dikumpulkan . . . . .	39
4.2	Contoh sebelum-sesudah pembersihan dan <i>case folding</i> . . . . .	40
4.3	Contoh sebelum-sesudah ubah slang dan <i>stemming</i> . . . . .	40
4.4	Contoh sebelum-sesudah <i>stopword remove</i> . . . . .	41
4.5	Tabel contoh leksikon positif dan negatif . . . . .	41
4.6	Hasil pelabelan sentimen . . . . .	42
4.7	Contoh hasil dari ekstraksi fitur menggunakan <i>Keras Tokenizer</i> . . . . .	47
4.8	Hasil percobaan jumlah <i>hidden layer</i> . . . . .	48
4.9	Hasil aktivasi fungsi yang digunakan . . . . .	48
4.10	Hasil percobaan jumlah <i>epoch</i> . . . . .	49
4.11	Hasil evaluasi model <i>deep learning Neural Network</i> . . . . .	49
4.12	Hasil percobaan jumlah <i>epoch</i> . . . . .	50
4.13	Hasil percobaan jumlah <i>hidden layer</i> . . . . .	50
4.14	Hasil evaluasi model deep learning CNN . . . . .	52
4.15	Hasil percobaan jumlah <i>epoch</i> . . . . .	52
4.16	Hasil percobaan jumlah <i>hidden layer</i> . . . . .	52
4.17	Hasil percobaan jumlah <i>hidden layer</i> dengan <i>epoch 20</i> . . . . .	53
4.18	Hasil evaluasi model deep learning BiLSTM . . . . .	54
4.19	Contoh hasil dari ekstraksi fitur menggunakan <i>Keras Tokenizer</i> . . . . .	55
4.20	Hasil percobaan jumlah <i>epoch</i> . . . . .	55
4.21	Hasil evaluasi model deep learning CNN 2 label . . . . .	56
4.22	Hasil percobaan jumlah <i>epoch</i> . . . . .	57
4.23	Hasil evaluasi model <i>deep learning BiLSTM 2 label</i> . . . . .	58
B.1	Contoh data yang dikumpulkan . . . . .	87
B.2	Contoh data yang dikumpulkan . . . . .	88



## DAFTAR KODE PROGRAM

3.1	Kode untuk mengubah kata menjadi lower case . . . . .	28
3.2	Kode untuk menghapus angka dan tanda baca . . . . .	28
3.3	Kode untuk melakukan tokenisasi . . . . .	28
3.4	Kode untuk penghapusan stopwords . . . . .	29
3.5	Kode untuk melakukan stemming . . . . .	29
3.6	Kode untuk mengubah kata slang . . . . .	29
3.7	Kode untuk menghitung polaritas berdasarkan lexicon dan melabeli data . . . . .	30
3.8	Kode untuk mengubah teks menjadi vektor . . . . .	32
3.9	Kode memanggil library Deep Learning CNN . . . . .	32
3.10	Kode untuk membuat model CNN . . . . .	33
3.11	Kode untuk membuat model BiLSTM . . . . .	34
3.12	Kode untuk membuat model Neural Network . . . . .	35
3.13	Kode untuk menampilkan elemen kehalaman website . . . . .	35
3.14	Kode untuk visualisasi data kedalam website . . . . .	36
3.15	Kode untuk membuat widget pada Streamlit . . . . .	36
A.1	Preprocessing_Text.py . . . . .	73
A.2	Label_Sentimen_Lexicon.py . . . . .	74
A.3	Preprocessing_Label_Data.py . . . . .	75
A.4	Visualisasi_Data.py . . . . .	75
A.5	Visualisasi_Evaluasi_Model.py . . . . .	77
A.6	Pembuatan_Model_Deep_Learning.py . . . . .	78
A.7	Prediksi_Sentimen.py . . . . .	80
A.8	Eksplorasi_Data.py . . . . .	82
A.9	App.py . . . . .	84



# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Kota Bandung adalah kota yang memiliki potensi pariwisata sangat tinggi di Indonesia. Penelitian yang dilakukan [Litbang Kompas dan Kementerian Pariwisata tahun 2016](#) mengembangkan Indeks Pariwisata Indonesia (IPI) dimana data berasal dari laporan jurnalistik 25 kabupaten/kota yang lulus seleksi data sekunder. Indeks Pariwisata Indonesia ini berskala 0-5. Menurut hasil penelitian tersebut, Bandung unggul dalam beberapa aspek Indeks Pariwisata Indonesia. Aspek-aspek tersebut antara lain aspek tata kelola menempati posisi ke-3 dengan indeks 3.68, aspek potensi (jumlah) wisata alam dan buatan menempati posisi ke-2 dengan indeks 3.45 dan aspek infrastruktur pendukung posisi ke-2 dengan indeks 4.12.

Dari segi kunjungan wisatawan, kota Bandung juga bisa dikatakan kota dengan jumlah kunjungan yang cukup besar. Berdasarkan data dari [Badan Pusat Statistik Provinsi Jawa Barat](#), jumlah kunjungan wisatawan ke objek wisata yang ada di kota Bandung mencapai 5.864.721 ditahun 2018. Pada tahun 2019 mencapai 2.442.250 dan pada tahun 2021 mencapai 393.223. Dengan jumlah kunjungan yang sangat banyak ini tentu memberikan dampak terhadap industri wisata kuliner dan belanja di kota Bandung. Bandung terkenal akan wisata kuliner dan belanjanya. Kementerian Pariwisata telah menetapkan kota Bandung sebagai salah satu destinasi unggulan wisata kuliner dan belanja (wiskulbel) di Indonesia [7].

Walaupun kota Bandung merupakan salah satu destinasi unggulan wisata kuliner dan belanja, tentu masih memiliki kekurangan dan permasalahan yang dihadapi. Salah satu cara untuk mengetahui kekurangan dan permasalahan tersebut adalah dengan menganalisis opini dari wisatawan kota Bandung. Opini adalah suatu respon, tanggapan atau penilaian seseorang terhadap suatu hal atau peristiwa yang sedang terjadi. Opini dapat bernilai positif, netral atau negatif. Nilai opini yang bernilai positif, netral atau negatif ini disebut sentimen. Dengan mengetahui dan menganalisis sentimen dari tiap wisatawan terhadap wiskulbel di kota Bandung maka dapat digunakan untuk membantu para *stakeholder* dalam penyusunan strategi yang tepat untuk memajukan wisata kuliner dan belanja kota Bandung.

Sekarang terdapat banyak opini masyarakat tentang wisata kuliner dan belanja yang ditulis di internet seperti pada *website TripAdvisor*, *Google Review* dan masih banyak lagi. Opini yang terdapat di internet berjumlah sangat banyak sehingga diperlukan bantuan komputer untuk mengolahnya. Salah satu teknik untuk mengolah data opini tersebut adalah teknik Analisis Sentimen yang merupakan sub-bidang ilmu dari *Text Mining*. Analisis Sentimen, juga disebut penambangan opini, adalah bidang ilmu yang menganalisis opini orang banyak, penilaian, sikap dan emosi terhadap entitas seperti produk, jasa, organisasi, individu, isu, acara dan topik[8].

Karena data yang diambil dari internet berupa data teks yang mana data tersebut adalah data tidak terstruktur, maka perlu adanya proses pra-pengolahan data sehingga bisa dibaca oleh mesin[9]. Analisis sentimen ini menggunakan teknik *Text Mining* yang disebut *Natural Language Processing* untuk mengubah data yang tidak terstruktur yaitu data teks menjadi data yang terstruktur. Proses NLP ini meliputi *case-folding*, *tokenization*, *stopword removing*, *stemming* dan beberapa teknik NLP lain[10].

Setelah melakukan proses pra-pengolahan data menggunakan teknik NLP, proses selanjutnya adalah melakukan teknik Analisis Sentimen untuk melabeli data. Seperti pada penjelasan sebelumnya bahwa sentimen adalah opini yang mengandung emosi atau kata positif, negatif atau netral maka fitur dari data ini merupakan kata-kata yang diubah menjadi vektor untuk dianalisis apakah termasuk ke dalam sentimen positif, negatif atau netral. Ada empat pendekatan yang digunakan yaitu : pendekatan secara *Machine learning*, pendekatan secara *Lexicon-based*, pendekatan secara *Rule-based* dan pendekatan secara Statistik [11]. Karena data pada penelitian ini belum ada labelnya, maka pendekatan yang digunakan dapat menggunakan *Lexicon-based* untuk mengetahui jenis sentimen dari setiap ulasan. Leksikon yang digunakan pada penelitian berdasarkan hasil dari penelitian dari Wahid Devid[12].

Setelah memberi label untuk setiap opini, dibuat model klasifikasi sentimen menggunakan algoritma *Deep Learning* berdasarkan data yang sudah dilabeli oleh daftar leksikon. Dilakukan proses ekstraksi fitur menggunakan *Word Tokenizer* untuk mengubah teks menjadi vektor agar fitur yang ada didalam teks bisa digunakan oleh algoritma *Deep Learning Neural Network*. Setiap kata yang ada dalam teks *review* diubah menjadi angka dari 1-n dimana n adalah banyaknya kata/*term* yang ada dalam data sentimen wisata kuliner dan belanja di Kota Bandung.

Setelah data teks diubah menjadi vektor melalui proses ekstraksi fitur, selanjutnya adalah melakukan pembuatan model klasifikasi menggunakan algoritma *Deep Learning Neural Network* serta melakukan evaluasi terhadap model yang dibuat. Akhirnya model dapat membedakan opini masyarakat yang bersentimen positif, negatif atau netral yang mana model ini lebih cepat dalam melakukan klasifikasi sentimen daripada manusia karena dilakukan oleh mesin. Dari hasil klasifikasi dapat diperoleh kata-kata bersentimen yang dapat dijadikan *insight*.

Secara singkat yang dilakukan pada penelitian ini adalah pengumpulan data opini atau *review* masyarakat untuk diolah dan dianalisis sehingga menemukan *insight* yang berguna sebagai usulan strategi pengelolaan wiskulbel di kota Bandung. Pada penelitian ini juga dilakukan Analisis Sentimen untuk melabeli data teks menjadi positif, negatif atau netral. Selain itu juga dilakukan pembuatan dan evaluasi model untuk mengklasifikasikan teks opini termasuk opini positif, negatif atau netral menggunakan teknik *Deep Learning Neural Network*. Model yang telah dibuat diimplementasikan pada *User Interface* berbasis *website* agar dapat digunakan dengan lebih mudah menggunakan *library* Streamlit. Teknologi dan *tools* yang digunakan pada penelitian ini adalah *library-library python* untuk mengolah dan menganalisis data seperti *numpy*, *tensorflow*, *keras*, *pandas*, *nlTK* dan *library-library python* lainnya.

## 1.2 Rumusan Masalah

Berdasarkan deskripsi dari latar belakang sebelumnya, rumusan masalah yang dapat dibuat adalah sebagai berikut :

1. Bagaimana cara mengumpulkan data opini-opini masyarakat dan turis terhadap wiskulbel kota Bandung yang ada pada internet?
2. Bagaimana cara menyiapkan dan mentransformasi data opini terhadap wiskulbel yang berbentuk teks agar dapat dianalisis?
3. Bagaimana cara melakukan Analisis Sentimen pada data opini masyarakat dan turis terhadap wiskulbel kota Bandung yang sudah diolah sebelumnya?
4. Bagaimana membuat model *Deep Learning* untuk melakukan klasifikasi sentimen dari data wiskulbel kota Bandung?
5. Bagaimana mengevaluasi model yang dibuat?
6. Bagaimana mengimplementasikan model analisis sentimen pada perangkat lunak berbasis *User Interface*?



## 1.3 Tujuan

Berdasarkan rumusan masalah, tujuan penelitian ini adalah sebagai berikut:

1. Mempelajari teknik *web scraping* untuk mengambil data dari *website* yang menyediakan opini wiskulbel kota Bandung yaitu *Tripadvisor* dan *Google Review*.
2. Mempelajari dan mengimplementasikan teknik-teknik *Natural Language Processing* untuk pra-pengolahan dan mentransformasi data teks menjadi fitur yang sesuai untuk *Deep Learning Neural Network*.
3. Mempelajari teknik *Natural Language Processing Analisis Sentimen* untuk memberi label pada teks berdasarkan leksikon kata positif dan negatif.
4. Mempelajari dan mengimplementasikan *Deep Learning Neural Network* untuk membuat model klasifikasi sentimen berdasarkan data ulasan wiskulbel kota Bandung yang telah dilabeli menggunakan leksikon kata positif dan negatif.
5. Melakukan evaluasi model seperti akurasi, presisi dan *recall* agar mengetahui tingkat ketepatan dari model.
6. Mempelajari, merancang dan membuat perangkat lunak berbasis *User Interface* yang menerapkan model klasifikasi menggunakan *Deep Learning Neural Network*.

## 1.4 Batasan Masalah

Penelitian ini memiliki batasan masalah sebagai berikut:

1. Data yang diambil adalah data dari wisata kuliner dan belanja kota Bandung dan sekitarnya yaitu kota-kota yang berbatasan langsung dengan kota Bandung (Menurut Website resmi [kota Bandung](#) wilayah yang berbatasan langsung adalah Kabupaten Bandung Barat dan Kabupaten Bandung).
2. Data yang diambil menggunakan Bahasa Indonesia.
3. Kriteria pemilihan restoran adalah restoran Bandung dan sekitarnya yang memiliki ciri khas terhadap Bandung dari segi makanan atau bisa juga terhadap tempat yang cukup ramai dikunjungi karena wisatanya seperti Lembang, Dago dan tempat-tempat lain yang menarik wisatawan.
4. Data untuk wisata belanja dibatasi hanya berdasarkan beberapa tempat wisata belanja rekomendasi dari *website Tripadvisor* seperti *Factory Outlet*, *PVJ*, *Ciwalk*, *Paskal* dan beberapa rekomendasi *Tripadvisor*.
5. Pengambilan data diambil mulai September 2022 sampai April 2023.

## 1.5 Metodologi

Metodologi yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mempelajari metode atau *tools* pengumpulan data pada *website*.
2. Melakukan studi literatur *Text Mining* yang meliputi *Text Preprocessing*, *Feature Extraction*, pelabelan data dengan teknik analisis sentimen, algoritma *Deep Learning Neural Network* untuk membuat model klasifikasi sentimen dan evaluasi model.
3. Mengumpulkan data teks opini atau *review* masyarakat tentang wisata kuliner dan belanja di Kota Bandung.
4. Mempelajari dan melakukan eksplorasi teknologi yang digunakan seperti bahasa pemrograman Python dan *library-library*.
5. Melakukan *Text Preprocessing* terhadap data meliputi pembersihan, pengolahan dan transformasi data agar dapat dianalisis lebih lanjut menggunakan bahasa pemrograman Python.
6. Melakukan tahapan *Text Mining* dibidang Analisis Sentimen yang meliputi pelabelan data dan Eksplorasi Data Analisis untuk menggali *insight* dari data menggunakan bahasa pemrograman Python.

7. Membuat model *Deep Learning Neural Network* dan mengevaluasi model.
8. Mengimplementasikan model pada perangkat lunak berupa *User Interface*(UI).
9. Menulis dokumen skripsi.

## 1.6 Sistematika Pembahasan

### 1. Bab 1 Pendahuluan

Bab ini berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi dan sistematika pembahasan.

### 2. Bab 2 Landasan Teori

Bab ini berisi dasar teori yang digunakan yaitu tahapan-tahapan *Text Mining*, analisis sentimen, *Deep Learning*, jenis-jenis arsitekturnya seperti CNN dan *Bidirectional LSTM*, evaluasi model dan visualisasi data.

### 3. Bab 3 Alur Pengerjaan & Eksplorasi Tools

Bab ini berisi alur gambaran umum pengerjaan skripsi dan eksplorasi *tools* yang digunakan. Alur pengerjaan meliputi pengumpulan data, *Text Preprocessing*, pelabelan data, visualisasi data, pembuatan dan evaluasi model serta pembuatan UI perangkat lunak.

### 4. Bab 4 Eksplorasi Data dan Pembuatan Model

Bab ini berisi proses pembuatan model dengan data yang telah dikumpulkan. Tahapan pada bab ini meliputi pengumpulan data, *Text Preprocessing*, pelabelan data, visualisasi data, pembuatan dan evaluasi model *Deep Learning Neural Network*.

### 5. Bab 5 Implementasi Perangkat Lunak

Bab ini berisi pengimplementasian model dan eksplorasi data ke dalam perangkat lunak berbasis *website*

### 6. Bab 6 Kesimpulan dan Saran

Bab ini berisi kesimpulan dan saran dari penelitian ini.