IMPLEMENTASI *WEB SCRAPING* PADA ULASAN WISATA TUGU KHATULISTIWA KALIMANTAN BARAT MENGGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK* (CNN)

MARESHA WIDYA MULIADIASTI
H1091211033

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2026

IMPLEMENTASI *WEB SCRAPING* PADA ULASAN WISATA TUGU KHATULISTIWA KALIMANTAN BARAT MENGGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK* (CNN)

MARESHA WIDYA MULIADIASTI
H1091211033

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2026

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Implementasi *Web Scraping* Pada Ulasan Tugu Khatulistiwa Kalimantan Barat Menggunakan Metode *Convolutional Neural Network* (CNN)
Nama Mahasiswa	: Maresha Widya Muliadiasti
NIM	: H1091211033
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 10 Juli 2026
SK Pembimbing	: No.4391/DST/UN22.8/DT.00.09/2026/ Tanggal 17 Juli 2026
SK Penguji	: No.4429/DST/UN22.8/DT.00.09/2026/ Tanggal 20 Juli 2026

Dosen Pembimbing

Pembimbing I	Pembimbing II
Hendra Perdana, M.Sc.	Neva Satyahadewi, M.Sc.,CRA.,CRP.,CRMP.
NIP 198810102019031020	NIP 198212042005012001

Dosen Penguji

Ketua Penguji	Anggota Penguji
Nurfitri Imro'ah, M.Si.	Dr. Evy Sulistianingsih, M.Sc.
NIP 198907182019032021	NIP 198502172008122006
Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Hendra Perdana, M.Sc.	Neva Satyahadewi, M.Sc.,CRA.,CRP.,CRMP.
NIP 198810102019031020	NIP 198212042005012001

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Maresha Widya Muliadiasti

NIM : H1091211033

Program Studi/ Jurusan : Statistika/Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 28 Juli 2026



Maresha Widya Muliadiasti
H1091211033

Implementasi *Web Scraping* pada Ulasan Wisata Tugu Khatulistiwa Kalimantan Barat menggunakan Metode *Convolutional Neural Network* (CNN)

Abstrak

Pariwisata memiliki dampak positif bagi perekonomian suatu negara atau daerah yang memiliki objek wisata. Pontianak merupakan ibu kota Provinsi Kalimantan Barat yang memiliki salah satu objek wisata, yaitu wisata Tugu Khatulistiwa. Penelitian ini memanfaatkan metode *Convolutional Neural Network* (CNN) untuk melakukan eksplorasi opini terhadap wisatawan yang memberikan ulasan terhadap tempat wisata Tugu Khatulistiwa di *Google Maps Review*. Grafik *accuracy* dan *loss* menunjukkan bahwa data mengalami *overfitting* yang ditandai dengan adanya kenaikan sangat cepat pada data *training* di sekitar *epoch* ke-20, namun *loss* data *training* menurun mendekati nol. Sementara, *accuracy* pada data uji cenderung statis, dengan kenaikan hanya di sekitar 0,75 hingga 0,78 dan *loss* pada data uji terus meningkat seiring bertambahnya *epoch*. Evaluasi keseluruhan data dengan analisis menggunakan CNN cukup baik dengan *accuracy* sebesar 0,77. Nilai *precision, recall*, dan *f1-score* paling rendah terdapat pada kelas negatif (*precision* 0,25; *recall* 0,23; *f1-score* 0,24) dan nilai *precision, recall*, dan *f1-score* paling tinggi terdapat pada kelas positif (*precision* 0,82; *recall* 0,93; *f1-score* 0,87). Hasil prediksi dari CNN pada 806 data ulasan menghasilkan 534 ulasan positif, 223 ulasan netral, dan 49 ulasan negatif. Hal ini mengindikasikan bahwa pengunjung cenderung memberikan ulasan positif atau model lebih mengarahkan ulasan ke ulasan positif.

Kata Kunci: Pariwisata, Analisis Sentimen, *Google Maps Review*, Tugu Khatulistiwa, *Convolutional Neural Network.*

Implementasi *Web Scraping* pada Ulasan Wisata Tugu Khatulistiwa Kalimantan Barat menggunakan Metode *Convolutional Neural Network* (CNN)

Abstract

Tourism has a positive impact on the economy of a country or region that possesses tourist attractions. Pontianak is the capital city of West Kalimantan Province and is home to one of its prominent tourist attractions, namely the Equator Monument. This study applies the Convolutional Neural Network (CNN) method to explore tourists' opinions based on reviews of the Equator Park tourist destination collected from Google Maps Reviews. The accuracy and loss graphs indicate that the data experience overfitting, which is characterized by a very rapid increase in training accuracy around the 20th epoch, while the training loss decreases to near zero. Meanwhile, the test data accuracy tends to static, showing an increase only in the range of 0.75 to 0.78, and the test data loss continues to increase as the number of epochs grows. Overall, the evaluation results using CNN analysis show fairly good performance, achieving an accuracy of 0.77. The lowest precision, recall, and f1-score values are found in the negative class (precision 0.25; recall 0.23; f1-score 0.24), while the highest precision, recall, and f1-score values are found in the positive class (precision 0.82; recall 0.93; f1-score 0.87). The CNN prediction results on 806 review data consist of 534 positive reviews, 223 neutral reviews, and 49 negative reviews. This indicates that visitors tend to provide positive reviews or that the model is more inclined to classify reviews as positive.

Keywords: *Tourism, Sentiment Analysis, Google Maps Review, Tugu Khatulistiwa, Convolutional Neural Network.*

PRAKATA

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa karena atas segala berkat, rahmat, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi. Skripsi yang berjudul **"Implementasi *Web Scraping* pada Ulasan Wisata Tugu Khatulistiwa Kalimantan Barat menggunakan Metode *Convolutional Neural Network* (CNN)"** merupakan salah satu syarat untuk memperoleh gelar Sarjana Statistika pada Program Studi Statistika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.

Penulisan skripsi ini tidak terlepas dari dukungan dan bimbingan dari beberapa pihak, oleh karena itu penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua saya, adik saya, dan keluarga saya yang saya cintai yang telah memberikan dorongan, nasehat, serta dukungan moril dan materil.
2. Bapak Yudhi, S.Si.,M.Si., selaku dosen Pembimbing Akademik yang telah memberikan bimbingan, arahan, dukungan, serta motivasi untuk penulis selama perkuliahan.
3. Bapak Hendra Perdana, M.Sc., selaku Dosen Pembimbing I yang telah memberikan arahan dan motivasi untuk menyelesaikan penulisan skripsi ini.
4. Ibu Neva Satyahadewi, M.Sc., CRA., CRP., CRMP., selaku Dosen Pembimbing II yang juga telah memberikan arahan serta motivasi untuk menyelesaikan penulisan skripsi ini.
5. Ibu Nurfitri Imro'ah, M.Si., selaku Dosen Penguji I yang telah memberikan saran dan masukan untuk perbaikan skripsi ini sehingga menjadi lebih baik.
6. Ibu Dr. Evy Sulistianingsih, M.Sc., selaku Dosen Penguji II yang telah memberikan saran dan masukan untuk perbaikan skripsi ini sehingga menjadi lebih baik.
7. Teman-teman Statistika Angkatan 2021 yang telah membersamai ketika perkuliahan serta memberikan semangat dalam penulisan skripsi ini.
8. Semua orang yang saya kenal yang memberikan dukungan, nasehat, motivasi, dan doa yang tidak dapat saya sebutkan satu-satu.

Penulis berharap dapat memperoleh saran dan masukan untuk kesempurnaan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkannya. Akhir kata penulis mengucapkan terima kasih kepada semua pihak yang telah berperan dari awal hingga akhir dalam penulisan skripsi ini.

Pontianak, 28 Juli 2026

Maresha Widya Muliadiasti

DAFTAR ISI

LEMBAR PENGESAHAN ..... **ii**
PERNYATAAN INTEGRITAS AKADEMIK ..... **iii**
Abstrak ..... **iv**
Abstract ..... **v**
PRAKATA ..... **vi**
DAFTAR ISI ..... **viii**
DAFTAR TABEL ..... **x**
DAFTAR GAMBAR ..... **xi**
DAFTAR SIMBOL ..... **xiii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Batasan Masalah 3
1.5 Tinjauan Pustaka 3
1.6 Metodologi Penelitian 4
BAB II LANDASAN TEORI **7**
2.1 *Google Maps* 7
2.2 *Data Mining* 7
2.3 *Machine Learning* 8
2.4 Jaringan Syaraf Tiruan 8
2.5 *Deep Learning* 9
2.6 Analisis Sentimen 9
2.7 *Natural Language Processing* 10
BAB III *WEB SCRAPING* DAN ANALISIS SENTIMEN MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK* **11**
3.1 *Web Scraping* 11
3.2 *Preprocessing Data* 12
3.3 *Word2Vec* 12
3.4 *Convolutional Neural Network* 14
3.5 Parameter Pengujian 17

3.5.1 *Recall* 17
3.5.2 *Precision* 17
3.5.3 Akurasi 17
3.5.4 *F1-Score* 17
BAB IV HASIL DAN PEMBAHASAN 19
4.1. Pengambilan Data Ulasan dengan Metode *Web Scraping* 19
4.2. *Preprocessing* Data 22
4.2.1 *Translate* Data 22
4.2.2 *Cleaning* Data 24
4.2.3 *Spelling Normalization* 25
4.2.4 *Case Folding* 26
4.2.5 Tokenisasi 27
4.2.6 *Stopword Removal* 29
4.2.7 *Stemming* 30
4.3 Analisis Sentimen dengan Metode *Convolutional Neural Network (CNN)* .31
4.3.1 Pembentukan *Word Embedding* Menggunakan *Word2Vec* 32
4.3.2 Simulasi Perhitungan *Skip-Gram* 34
4.3.3 Tokenisasi dan *Padding* 38
4.3.4 Pembagian Data Latih, Validasi, dan Data Uji 39
4.3.5 Analisis *Convolutional Neural Network* (CNN) 41
4.3.6 Evaluasi Arsitektur *Convolutional Neural Network* 44
4.4 Interpretasi Distribusi Hasil Prediksi CNN 49
BAB V KESIMPULAN 54
5.1 Kesimpulan 54
5.2 Saran 55
DAFTAR PUSTAKA 56
LAMPIRAN 58

DAFTAR TABEL

Tabel 4.1 Sebagian Data Hasil *Web Scraping*.....21
Tabel 4.2 Hasil Proses *Translate* Data.....23
Tabel 4.3 Hasil Proses *Cleaning* Data.....24
Tabel 4.4 Hasil Proses *Spelling Normalization*.....26
Tabel 4.5 Data Hasil Proses *Case Folding*.....27
Tabel 4.6 Data Hasil Proses Tokenisasi.....28
Tabel 4.7 Data Hasil Proses *Stopword Removal*.....29
Tabel 4.8 Hasil Proses *Stemming*.....30
Tabel 4.9 Pasangan Kata *Word2Vec* Metode Skip-gram.....34
Tabel 4.10 Nilai Vektor *Output Word2Vec*.....35
Tabel 4.11 Perhitungan *Dot Product Word2Vec*.....36
Tabel 4.12 Skor kata dan Bentuk Eksponensial.....36
Tabel 4.13 Nilai Probabilitas dan *Loss* Kata.....38
Tabel 4.14 Ringkasan Arsitektur CNN.....43
Tabel 4.15 Hasil Evaluasi CNN.....48
Tabel 4.16 Distribusi Hasil Prediksi Sentimen menggunakan CNN.....49
Tabel 4.17 Perbandingan Hasil Klasifikasi *TextBlob* dan CNN.....51

DAFTAR GAMBAR

Gambar 1.1 *Flowchart* Penelitian.....6
Gambar 3. 1 Arsitektur Model *Skip-gram*.....13
Gambar 3. 2 Arsitektur *Convolutional Neural Network* (CNN).....15
Gambar 4.1 Menu Pencarian *Website Google Maps*.....20
Gambar 4.2 Proses Scraping Data menggunakan *Instant Data Scraper*.....20
Gambar 4.3 Hasil Sebagian *Scraping* Ulasan.....21
Gambar 4.4 Kode Python Analisis Sentimen dengan CNN.....32
Gambar 4.5 Proses *Word Embedding* menggunakan *Word2Vec*.....33
Gambar 4.6 Tokenisasi dan *Padding*.....39
Gambar 4.7 Pembagian Data Latih dan Data Uji.....40
Gambar 4.8 Menentukan Arsitektur *Convolutional Neural Network*.....41
Gambar 4.9 Perhitungan Data Training dan Data Uji : a.) *Accuracy*; b.) *Loss*.....44
Gambar 4.10 *Confusion Matrix* Model CNN untuk Tiga Kelas Sentimen.....46

DAFTAR LAMPIRAN

Lampiran 1 *Syntax* Analisis CNN menggunakan *Google Colab* 58
Lampiran 2 Contoh *Review* 68

DAFTAR SIMBOL

x_i : Vektor *word embedding* ke-i

n : Jumlah total elemen dalam satu kalimat

$s_1, s_2, \dots, s_n$: *Word embedding*

$\oplus$: Operator *concatentation*

b : Bias

c_i : Nilai fitur hasil konvolusi pada posisi ke-i

h : Ukuran filter atau *window* untuk jumlah keseluruhan kata yang dianalisis

f : Fungsi aktivasi

w : Bobot filter konvolusi

$S_{i:i+l-1}$: Gabungan vektor *word embedding* pada posisi ke-i sampai $i + l - 1$

$\hat{y}$: Probabilitas hasil prediksi kelas klasifikasi

x : Vektor hasil *fully connected layer*

x^T : *Transpose* dari vektor x

W_j : Bobot kelas ke-j

b_j : Bias kelas ke-j

K : Jumlah total kelas

j : Indeks kelas

k : Indeks penjumlahan kelas

$arg\ max$: Operator untuk memilih nilai maksimum

exp : Fungsi eksponensial

w_t : Kata target pada posisi ke-t

w_j : Kata konteks pada posisi ke-j

v_{w_t} : Vektor input kata target pada posisi ke-t

v'_{w_j} : Vektor *output* kata konteks

$\mathcal{L}$: Fungsi *hidden layer skip-gram*

L : Nilai *loss*

S : Matriks *word embedding*

$\hat{c}$: Fungsi *max-pooling*

W : Matriks bobot *fully connected layer*

y_i : Label aktual

$\hat{y}_i$: Probabilitas prediksi pada kelas ke-i

A : Kalimat pada *Word2vec*

T : Jumlah kata pada kalimat *Word2vec*

t : Indeks kata target *Word2vec*

$P(\cdot)$: Fungsi Probabilitas *Word2vec*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pariwisata adalah sektor paling efektif untuk meningkatkan devisa Indonesia yang terdiri dari Sumber Daya Manusia (SDM), letak geografis yang terdiri dari luas wilayah serta keragaman daya alam, budaya, kuliner, dan kekayaan yang ada di Indonesia yang menjadi daya tarik wisatawan domestik maupun mancanegara (Rahma, 2020). Pariwisata memberikan dampak positif bagi perekonomian suatu negara atau daerah yang memiliki objek wisata yang dapat berdampak pada segi pendapatan maupun dampak pada lapangan pekerjaan yang tercipta dengan adanya pengembangan objek tempat wisata (Saropah et al., 2022).

Kalimantan Barat adalah salah satu daerah yang memiliki kekayaan etnis, budaya, dan bahasa. Sungai Kapuas merupakan suatu ciri khas yang menarik karena merupakan sungai terpanjang di Indonesia. Meskipun Kalimantan Barat tidak menjadikan sektor pariwisata sebagai sumber perekonomian utama, Kota Pontianak merupakan ibukota Provinsi memiliki objek wisata, kekayaan kuliner, rumah adat, serta hal menarik lainnya. Sehingga Kota Pontianak memiliki daya tarik wisata yang dapat meningkatkan daya tarik wisatawan untuk berkunjung (Hardilia et al., 2024).

Kota Pontianak adalah Kota Khatulistiwa di mana Kota Pontianak tepat dilalui oleh Garis Khatulistiwa. Sebagai salah satu kota tujuan wisata di Provinsi Kalimantan Barat, Kota Pontianak memiliki daya tarik wisata prioritas yaitu Tugu Khatulistiwa, objek tersebut berasosiasi langsung dengan potensi Sungai Kapuas yang memberikan nuansa daya tarik wisata sungai secara natural bersandingan dengan karakteristik keunikan multi budaya, kawasan ekuator yang membelah zona utara maupun selatan bumi dan satu-satunya zona khatulistiwa berada tepat di Kota provinsi yaitu Kota Pontianak (Andrasmoro dan Sariani, 2023). Tugu Khatulistiwa memiliki sejarah yang berpotensi besar untuk dikembangkan menjadi daya tarik

wisata edukasi di Pontianak yang bertujuan untuk meningkatkan pengetahuan dalam bidang tertentu yang menjadi fokus kunjungan (Hardilia et al., 2024).

Perkembangan media digital dapat memberi dampak positif pada industri pariwisata. Wisatawan dapat memberikan ulasan atau pengalaman wisata mereka dengan mudah melalui berbagai macam *platform online*. *Google Maps* merupakan salah satu layanan navigasi dan pemetaan lokasi yang dikembangkan oleh Google Inc. Selain dapat digunakan sebagai alat navigasi, *Google Maps* dapat digunakan sebagai *platform online* untuk mempublikasikan ulasan pengunjung mengenai wisata, yang mencakup opini pribadi mengenai pengalaman dari pengunjung wisata.

Pengunjung dapat mengandalkan pandangan dan pengalaman wisatawan yang pernah berkunjung ke suatu destinasi wisata yang terpublikasi di *Google Maps* untuk mendapatkan wawasan yang lebih baik sebelum mengunjungi suatu tempat dan dapat menjadi penentu keputusan untuk berkunjung ke suatu tempat wisata. Ulasan pengunjung memberikan dampak langsung pada keputusan berwisata pengunjung dan reputasi wisata. Sehingga, ulasan yang diberikan pengunjung memberikan fungsi penting dalam analisis kualitas wisata. Dengan perkembangan penelitian mengenai analisis sentimen pada sektor pariwisata memberikan kontribusi penting terhadap optimalisasi sistem manajemen wisata di Indonesia (Hamdana et al., 2025)

Penelitian ini memanfaatkan metode *Convolutional Neural Network (CNN)* untuk melakukan eksplorasi opini terhadap wisatawan yang memberikan ulasan terhadap tempat wisata Tugu Khatulistiwa di *Google Maps Review*. Analisis sentimen merupakan teknik *data mining* berupa opini publik yang menganalisis pendapat seseorang, penilaian, sikap, dan emosi terhadap entitas, individu, isu, peristiwa, topik, dan lainnya. Analisis sentimen akan memproses hasil pencarian untuk item tertentu, menghasilkan daftar produk, dan kemudian menggabungkan masing-masing pendapat.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana implementasi *web scraping* dalam pengumpulan data ulasan wisata Tugu Khatulistiwa Kalimantan Barat pada *Google Maps Review*?
2. Bagaimana implementasi metode *Convolutional Neural Network* (CNN) dalam memprediksi sentimen positif, negatif, dan netral pada ulasan wisata Tugu Khatulistiwa Kalimantan Barat?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dikemukakan, tujuan dalam penelitian ini adalah:

1. Mengimplementasikan *web scraping* untuk mengambil data ulasan wisata Tugu Khatulistiwa Kalimantan Barat pada *Google Maps Review.*
2. Mengimplementasikan metode *Convolutional Neural Network* (CNN) dalam memprediksi sentimen positif, negatif, dan netral pada ulasan Tugu Khatulistiwa Kalimantan Barat.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan merupakan data ulasan wisata Tugu Khatulistiwa Kalimantan Barat pada *Google Maps Review*.
2. Ulasan yang diambil merupakan ulasan yang menggunakan Bahasa Inggris.

1.5 Tinjauan Pustaka

Wianto (2018) mengimplementasikan Convolutional Neural Network pada analisis sentimen pada studi kasus politik. Data diperoleh dari *threats* Twitter dengan data awal sebanyak 64.245.029 *threats* dengan hasil data yang telah melalui data *preprocessing* yang akan dilatih dengan menggunakan CNN sebanyak 9.948 data dengan label positif, netral, dan negatif. Hasil analisis dibagi menjadi 3

tahapan, tahapan A terdiri dari dua kelas (positif dan negatif) dengan nilai akurasi sebesar 0,833 dan *f1-score* sebesar 0,833, tahapan B terdiri dari tiga kelas (positif, negatif, dan netral) dengan nilai akurasi 0,709 dan *f1-score* sebesar 0,703, dan tahapan C dengan 3 kelas per topik politik dengan melakukan klasifikasi opini terhadap topik politik seperti tokoh, partai, dan kebijakan memiliki nilai akurasi sebesar 0,658 dan *f1-score* sebesar 0,656. Sehingga, CNN lebih unggul dari SVM dan *naïve bayes* dan mampu mengenali pola bahasa tidak baku dan tidak formal di media sosial.

Yuliska (2021) mengimplementasikan *Convolutional Neural Network* pada analisis sentimen pada data saran mahasiswa terhadap kinerja departemen di perguruan tinggi. Data berasal dari saran mahasiswa terhadap kinerja departemen perguruan tinggi di Politeknik Caltex Riau dengan total saran sebanyak 8.400 data. Hasil analisis diperoleh bahwa *Simple* CNN dengan 2 kelas memiliki nilai akurasi sebesar 97% dan *f1-score* 97% dan *Simple* CNN dengan 3 kelas memiliki nilai akurasi sebesar 91% dan *f1-score* sebesar 91%, serta *DoubleMax* CNN dengan 2 kelas memiliki nilai akurasi sebesar 98% dan *f1-score* sebesar 98% dan *Simple* CNN dengan 3 kelas memiliki nilai akurasi sebesar 91% dan *f1-score* sebesar 90%.

Putra (2024) mengimplementasikan beberapa algoritma *deep learning* pada analisis sentimen pada tempat wisata di Kalimantan Tengah. Data diambil pada *Google Maps* sejak tahun 2017 sampai dengan 2024 sebanyak 19.315 data ulasan, kemudian telah dilakukan *preprocessing* data sehingga menghasilkan data sebanyak 9.722 data ulasan. Hasil analisis diperoleh metode BERT memiliki nilai akurasi sebesar 98,02% dan *f1-score* sebesar 98,82%, model LSTM memiliki akurasi 84% dan *f1-score* sebesar 91%, dan model CNN memiliki akurasi sebesar 86,70% dan *f1-score* sebesar 92,66%.

1.6 Metodologi Penelitian

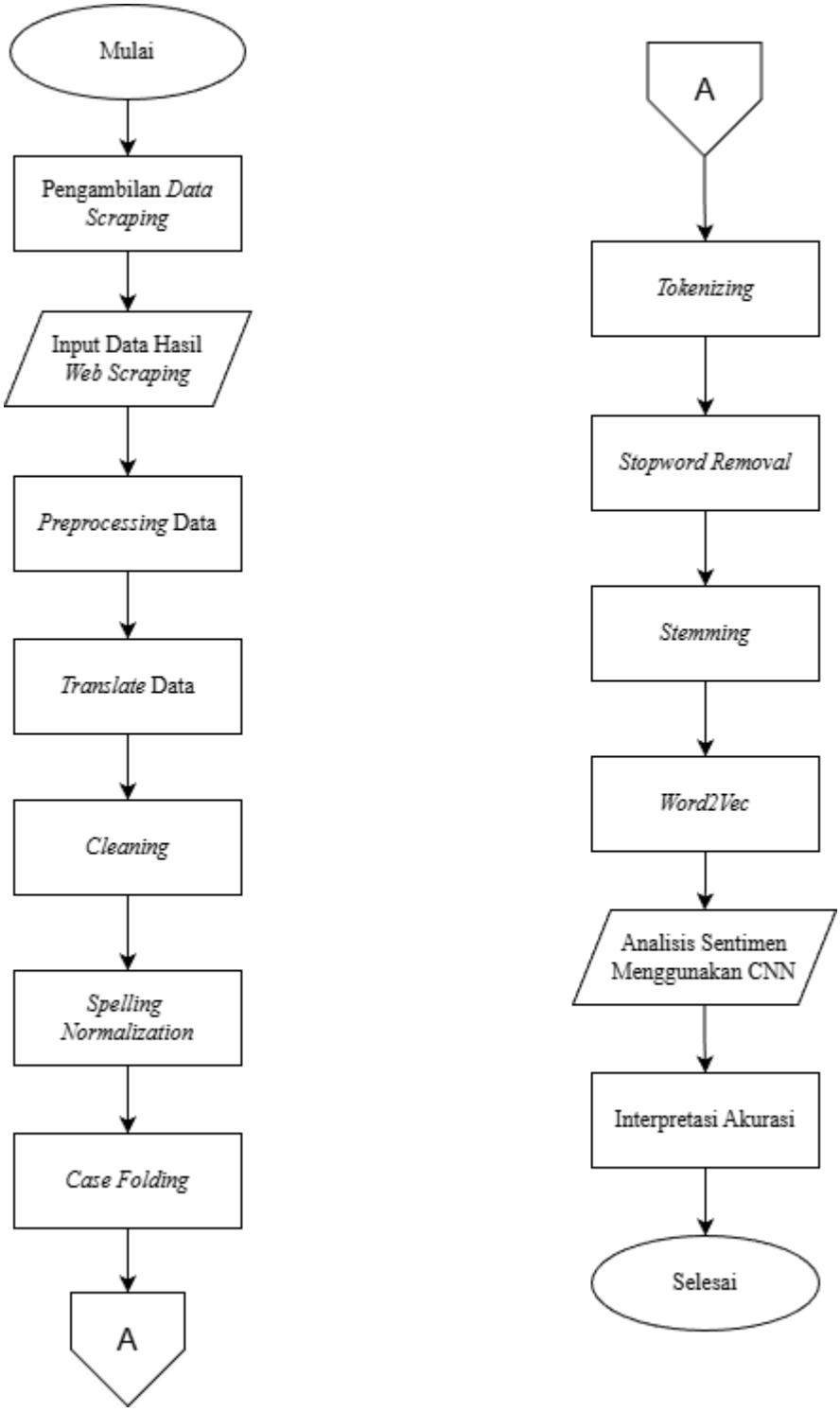
Metode yang digunakan pada penelitian ini diawali menggunakan studi literatur untuk mencari referensi yang relevan dengan permasalahan yang akan diteliti. Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari *Google Maps Review* Tugu Khatulistiwa, Kalimantan Barat. Data

diambil menggunakan *web scraping* pada tanggal 4 Juni 2025 berjumlah 806 data *review* Tugu Khatulistiwa, Kalimantan Barat dari *Google Maps*.

Dalam memulai proses metode penelitian diperlukan langkah-langkah berikut:

a. *Scraping* data, proses pengumpulan data dari Google Maps Review. Proses ini dilakukan menggunakan *browser Google Chrome* dengan bantuan ekstensi *Instant Data Scraper* yang telah diunduh terlebih dahulu
b. *Preprocessing* data*,* proses mengubah bentuk suatu data menjadi data yang lebih terstruktur dalam proses *data mining.* Tahapan *preprocessing* data dimulai dari *Translate data, cleaning, spelling normalization, case folding, tokenizing, stopword removal*, dan *stemming.*
c. *Word2Vec*, algoritma representasi kata sebuah kalimat yang memproses data *review* dari bentuk teks ke bentuk numerik seperti vektor.
d. Pelatihan model untuk membentuk model untuk klasifikasi *review* yang mengandung sentimen positif, negatif, dan netral menggunakan CNN.
e. Pengujian model untuk mengetahui kinerja model dalam klasifikasi sentimen positif, negatif, dan netral yang terdiri dari akurasi, *recall, precision, dan f-measure.*

Adapun alur dari penelitian divisualisasikan dalam bentuk *flowchart* pada Gambar 1.1 berikut:



Gambar 1.1 *Flowchart* Penelitian