

**SISTEM KOREKSI EJAAN BAHASA INDONESIA PADA  
BERITA *ONLINE* MENGGUNAKAN METODE *LONG SHORT-  
TERM MEMORY***

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika



Oleh :

Atikah Syifa Warastri

NIM 09021182025027

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2024**

**LEMBAR PENGESAHAN SKRIPSI**

**SISTEM KOREKSI EJAAN BAHASA INDONESIA PADA  
BERITA *ONLINE* MENGGUNAKAN METODE *LONG SHORT-  
TERM MEMORY***

Oleh :

**Atikah Syifa Warastri**

**NIM : 09021182025027**

**Palembang, 12 Agustus 2024**

**Pembimbing I**



**Alvi Syahrini Utami, M.Kom.**  
**NIP. 197812222006042003**

**Pembimbing II**



**Kanda Januar Miraswan, M.T.**  
**NIP. 199001092019031012**

**Mengetahui,**

**Ketua Jurusan Teknik Informatika**



**Hadipurnawan Satria, Ph.D.**  
**NIP. 198004182020121001**

## TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Jum'at tanggal 02 Agustus 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Atikah Syifa Warastri

NIM : 09021182025023

Judul : Sistem Koreksi Ejaan Bahasa Indonesia Pada Berita *Online* Menggunakan Metode *Long Short-Term Memory*

dan dinyatakan LULUS.

1. Ketua

Yunita, M.Cs.

198306062015042002

2. Penguji 1

Dian Palupi Rini, M.Kom., Ph.D.

197802232006042002

3. Pembimbing 1

Alvi Syahrini Utami, M.Kom.

197812222006042003

4. Pembimbing 2

Kanda Januar Miraswan, M.T.

199001092019031012

Y

DR

AS

KJ



## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Atikah Syifa Warastri  
NIM : 09021182025027  
Program Studi : Teknik Informatika  
Judul : Sistem Koreksi Ejaan Bahasa Indonesia Pada Berita *Online*  
Menggunakan Metode *Long Short-Term Memory*

Hasil Pengecekan Software iThenticate/Turnitin : 14%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya akan bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 8 Agustus 2024



Atikah Syifa Warastri

NIM 09021182025027

## **MOTTO DAN PERSEMBAHAN**

Motto:

*“If you can't beat the fear, just do it scared.”*

- Glennon Doyle Melton

Kupersembahkan Karya Tulis ini kepada:

- Allah SWT
- Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

## ABSTRACT

When writing news articles, ease of understanding and comprehension by readers is something that must be considered by the writer. One of the crucial things in writing a work is spelling errors or what are called typos. To avoid misinterpretation by readers, an automatic spelling correction system is needed in order to detect and correct spelling errors. One of the approaches used is the Long Short-Term Memory method using word2vec or TF-IDF to produce the probability of each unique word from the training data. In the experiment, the highest accuracy value was obtained at 0.3273 using the word2vec configuration. The test results show a Root Mean Square Error value of 0.997 with the distance between the probability of the predicted results and the ground truth ranging from 0.913 to 0.999.

**Keywords:** Typographical Error, Long Short-Term Memory, Word2Vec, TF-IDF, Root Mean Square Error

Palembang, 12 August 2024

Supervisor I



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

Supervisor II



Kanda Januar Miraswan, M.T.  
NIP. 199001092019031012

Approve,  
Head of Informatics Engineering Department



Hadiputra Satria, Ph.D.  
NIP. 198004182020121001

## ABSTRAK

Dalam pembuatan artikel berita, kemudahan untuk dimengerti dan dipahami oleh pembaca merupakan hal yang wajib dipertimbangkan oleh penulis. Salah satu hal krusial dalam penulisan suatu karya adalah kesalahan ejaan atau disebut dengan *typographical error*. Untuk menghindari kesalahan interpretasi oleh pembaca, maka dibutuhkan sistem koreksi ejaan otomatis yang dapat mendeteksi dan mengoreksi kesalahan eja. Salah satu pendekatan yang digunakan ialah metode *Long Short-Term Memory* yang dikembangkan dengan menggunakan *word2vec* atau TF-IDF untuk menghasilkan probabilitas tiap kata unik dari data latih. Pada percobaan, didapat nilai akurasi tertinggi sebesar 0,3273 dengan menggunakan konfigurasi *word2vec*. Hasil pengujian menampilkan nilai *Root Mean Square Error* sebesar 0,997 dengan jarak antara probabilitas hasil prediksi terhadap *ground truth* berkisar dari 0,913 hingga 0,999.

**Kata Kunci:** *Typographical Error, Long Short-Term Memory, Word2Vec, TF-IDF, Root Mean Square Error*

Palembang, 12 Agustus 2024

Pembimbing I



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

Pembimbing II



Kanda Januar Miraswan, M.T.  
NIP. 199001092019031012

Mengetahui,  
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.  
NIP. 1958064182020121001

## KATA PENGANTAR

Puji Syukur kepada Allah SWT yang telah memberikan nikmat berupa kesehatan, kesempatan, serta karunia-Nya yang telah menyertai penulis sehingga dapat menyelesaikan penelitian ini dengan baik. Penelitian yang berjudul “Sistem Koreksi Ejaan Bahasa Indonesia Pada Berita *Online* Menggunakan Metode *Long Short-Term Memory*” ini dilaksanakan guna memenuhi syarat menyelesaikan pendidikan program Strata-1 (S1) di Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam menyelesaikan penelitian ini, penulis menerima berbagai dukungan dan bimbingan baik secara langsung maupun tidak langsung. Oleh karena itu, penulis bermaksud untuk menyampaikan terima kasih kepada:

1. Allah SWT yang telah memberkahi rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini.
2. Alm. Drs. Megah Mulya, M.T., selaku ayah dan panutan penulis yang telah mendidik penulis untuk pantang menyerah dan bekerja keras. Des Alwine Zayanti, M.Si., selaku ibu dan idola penulis yang sudah memberi dukungan materi dan moral yang sedemikian rupa besarnya sehingga memotivasi penulis untuk menyelesaikan penelitian ini.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. M. Fachrurrozi, S.Si., M.T., selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.



5. Ibu Alvi Syahrini Utami, M.Kom., selaku Dosen Pembimbing I yang telah memberi bimbingan, arahan dan masukan selama penelitian ini dilaksanakan.
6. Bapak Kanda Januar Miraswan, M.T., selaku Dosen Pembimbing II yang telah memberi bimbingan, arahan dan masukan selama penelitian ini dilaksanakan.
7. Saudara dan saudari penulis, Muhammad Fauzan Arroyan dan Aurell Safira Izzati, yang telah mendukung dan menghibur penulis saat penelitian.
8. Teman-teman penulis, Liza Hertias Ningrum, Diah Shinta Resmy, Munawaroh Syahfitri, Afifah Putri Mevtillina, Rahma Nabila Putri, teman-teman !ruangan, dan teman-teman DDB Telkom yang tidak bisa disebutkan satu per satu, sebagai penyemangat, pendukung, penghibur sekaligus teman seperjuangan penulis selama melakukan penelitian.
9. Seluruh staf Jurusan Teknik Informatika dan Fakultas Ilmu Komputer Universitas Sriwijaya.
10. Seluruh pihak yang tidak dapat disebutkan satu per satu yang telah memberi dukungan selama perkuliahan dan penelitian.

Penulis menyadari bahwa masih terdapat kekurangan dari berbagai aspek. Oleh karena itu, penulis mengharapkan kritik dan saran demi kemajuan penelitian yang akan datang. Semoga penelitian ini bermanfaat bagi pengembangan wawasan.

Palembang, 12 Agustus 2024

Penulis



Atikah Syifa Warastri

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
TANDA LULUS UJIAN KOMPREHENSIF.....	iii
HALAMAN PERNYATAAN .....	iv
MOTTO DAN PERSEMBAHAN .....	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xiv
BAB I PENDAHULUAN .....	I-1
1.1    Pendahuluan .....	I-1
1.2    Latar Belakang .....	I-1
1.3    Rumusan Masalah .....	I-4
1.4    Tujuan Penelitian.....	I-4
1.5    Manfaat Penelitian.....	I-5
1.6    Batasan Masalah.....	I-5
1.7    Sistematika Penulisan.....	I-5
1.8    Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR .....	II-1
2.1    Pendahuluan .....	II-1
2.2    Landasan Teori .....	II-1
2.2.1    Sistem Koreksi Ejaan .....	II-1
2.2.2 <i>Preprocessing</i> .....	II-2
2.2.3 <i>Long Short-Term Memory</i> .....	II-3
2.2.4    Tanh.....	II-6
2.2.5    Softmax .....	II-7
2.2.6 <i>Hadamard Product</i> .....	II-7
2.2.7    Word2vec .....	II-8
2.2.8 <i>Root Mean Squared Error</i> .....	II-9

2.2.9	<i>Accuracy dan Loss</i> .....	II-9
2.2.10	<i>Rational Unified Process</i> .....	II-10
2.3	Penelitian Lain yang Relevan.....	II-12
2.4	Kesimpulan.....	II-18
<b>BAB III METODOLOGI PENELITIAN</b> .....		III-1
3.1	Pendahuluan .....	III-1
3.2	Pengumpulan Data .....	III-1
3.2.1	Jenis Data .....	III-1
3.2.2	Sumber Data.....	III-1
3.2.3	Metode Pengumpulan Data.....	III-2
3.3	Tahapan Penelitian .....	III-2
3.3.1	Kriteria Pengujian .....	III-6
3.3.2	Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-6
3.3.3	Pengujian Penelitian.....	III-7
3.4	Metode Pengembangan Perangkat Lunak .....	III-7
3.4.1	Fase Insepsi .....	III-7
3.4.2	Fase Elaborasi .....	III-8
3.4.3	Fase Konstruksi.....	III-8
3.4.4	Fase Transisi .....	III-9
3.5	Manajemen Proyek Penelitian.....	III-9
3.5.1	Penjadwalan Waktu Penelitian.....	III-9
3.6	Kesimpulan.....	III-13
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK</b> .....		IV-1
4.1.	Pendahuluan .....	IV-1
4.2.	Fase Insepsi .....	IV-1
4.2.1.	Pemodelan Bisnis .....	IV-1
4.2.2.	Kebutuhan Sistem .....	IV-1
4.2.3.	Analisis dan Desain.....	IV-2
4.3.	Fase Elaborasi.....	IV-11
4.3.1.	Pemodelan Bisnis .....	IV-11
4.3.2.	Perancangan Data.....	IV-11
4.3.3.	Kebutuhan Sistem .....	IV-11

4.3.4.	Analisis dan Desain.....	IV-12
4.4.	Fase Konstruksi .....	IV-13
4.4.1.	Implementasi <i>Interface</i> .....	IV-13
4.5.	Fase Transisi.....	IV-13
4.5.1	Pemodelan Bisnis .....	IV-13
4.5.2	Rencana Pengujian .....	IV-14
4.5.3	Implementasi .....	IV-14
4.6.	Kesimpulan.....	IV-15
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan .....	V-1
5.2	Data Hasil Penelitian .....	V-1
5.2.1	Konfigurasi Percobaan .....	V-1
5.2.2	Data Hasil Konfigurasi.....	V-2
5.3	Analisis Hasil Penelitian .....	V-6
5.4	Kesimpulan.....	V-8
BAB VI KESIMPULAN DAN SARAN .....		VI-1
6.1	Pendahuluan .....	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran .....	VI-1
DAFTAR PUSTAKA .....		iii

## DAFTAR TABEL

Tabel II-1 Kata Hasil Koreksi Pengujian Data 1 .....	II-16
Tabel II-2 Hasil Pengujian .....	II-17
Tabel III-1 Rancangan Tabel Hasil Analisis <i>Error Correction</i> .....	III-7
Tabel III-2 <i>Work Breakdown Structure</i> .....	III-9
Tabel IV-1 Kebutuhan Fungsional Perangkat Lunak.....	IV-1
Tabel IV-2 Kebutuhan Non-Fungsional Perangkat Lunak .....	IV-2
Tabel IV-3 Hasil Proses <i>Noise Removal</i> .....	IV-4
Tabel IV-4 Hasil Proses <i>Case Folding</i> .....	IV-4
Tabel IV-5 Hasil Proses <i>Stemming</i> .....	IV-5
Tabel IV-6 Hasil Proses <i>Tokenization</i> .....	IV-6
Tabel IV-7 Definisi Aktor.....	IV-8
Tabel IV-8 Definisi <i>Use Case</i> .....	IV-8
Tabel IV-9 Skenario <i>Use Case</i> Deteksi Typo .....	IV-9
Tabel IV-10 Skenario <i>Use Case</i> Koreksi Typo.....	IV-10
Tabel IV-11 Rencana Pengujian <i>Use Case</i> Deteksi <i>Error</i> dan Koreksi <i>Error</i> .....	IV-14
Tabel IV-12 Pengujian <i>Use Case</i> Deteksi <i>Error</i> dan Koreksi <i>Error</i> .....	IV-14
Tabel V-1 Konfigurasi Penelitian .....	V-1
Tabel V-2 Sampel Data Uji.....	V-3
Tabel V-3 Pengujian Deteksi dan Koreksi Kesalahan Ejaan .....	V-6
Tabel V-4 Hasil Perhitungan Nilai RMSE.....	V-6

## DAFTAR GAMBAR

Gambar II-1 Struktur LSTM Pada Tiap <i>Timestep</i> .....	II-5
Gambar II-2 Fungsi Aktivasi Tanh .....	II-7
Gambar II-3 <i>Rational Unified Process (RUP)</i> .....	II-11
Gambar II-4 Grafik Pengujian Jenis <i>Typographical Error Insertion</i> Berdasarkan Jumlah Kata <i>Typographical Error</i> Dalam Satu Dokumen .....	II-14
Gambar II-5 Grafik Pengujian Jenis <i>Typographical Error Substitution</i> Berdasarkan Jumlah Kata <i>Typographical Error</i> Dalam Satu Dokumen .....	II-14
Gambar II-6 Hasil Scan Data Uji .....	II-15
Gambar II-7 Saran Kata Terdekat yang Direkomendasikan Algoritma <i>Damerau-Levenshtein Distance</i> .....	II-17
Gambar II-8 Saran Kata yang Salah .....	II-18
Gambar III-1 Alur Penelitian .....	III-2
Gambar III-2 Alur Pengumpulan Data.....	III-3
Gambar III-4 Alur Persiapan Data .....	III-4
Gambar III-5 Alur Koreksi <i>Error</i> .....	III-5
Gambar IV-1 Alur Tahapan Pra-Pengolahan Teks .....	IV-3
Gambar IV-2 Alur Proses Deteksi Kesalahan Ejaan.....	IV-7
Gambar IV-3 Diagram <i>Use Case</i> .....	IV-8
Gambar IV-4 Desain Antarmuka Aplikasi.....	IV-12
Gambar IV-5 Implementasi Antarmuka Aplikasi .....	IV-13

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab pendahuluan ini menerangkan gagasan utama yang mendasari rencana riset. Gagasan utama adapun mencakup latar belakang, rumusan permasalahan, tujuan serta manfaat dari implementasi eksperimen termasuk pembatasan permasalahan.

### **1.2 Latar Belakang**

Perkembangan pesat era kemajuan teknologi informasi serta komunikasi tercermin dalam kehadirannya *platform* daring. Kemajuan berdampak atas ketenaran *platform* secara keseluruhan, elemen terbesar yaitu *platform* berwujud daring. Era saat ini ditandai oleh laju perubahan yang sangat cepat dan serba instan, hal ini juga berpengaruh pada perkembangan dimana hal tersebut mempermudah semuanya secara cepat. Hadirnya peranan internet, sebagai sumber Informasi beraneka ragam, secara pasti menggeser dominasi *platform* berwujud percetakan. Demikian juga diperlihatkan banyaknya beraneka pemberitaan di media internet menghadirkan informasi secara aktual serta tidak menyertakan pembiayaan tambahan.

Terlepas dari kemudahan akses berita secara *online*, kemudahan untuk dimengerti dan dipahami oleh pembaca juga merupakan hal yang wajib dipertimbangkan oleh penulis. Berita harus ditulis secara sistematis, jelas, logis, dan komunikatif sesuai dengan konsep dan konteks. Informasi yang ada dalam

berita diharapkan dapat memberikan manfaat untuk orang lain bahkan bisa sampai ke masyarakat luas (Oktaria, 2017).

Elemen sebagaimana dikatakan paling vital saat penyusunan dokumentatif ataupun susunan kata yaitu salah dalam ejaan atau juga dikatakan sebagai *typographical error*. Kesalahan ejaan sangat umum dalam teks buatan manusia baik teks formal maupun teks informal dengan penyebab yang dapat bervariasi dari kesalahan tangan atau pikiran yang tidak disengaja hingga kurangnya pengetahuan ejaan penulis. Untuk pembaca, pendeteksian dan koreksi ejaan teks mungkin sepele karena manusia secara alami dapat mendeteksi pola dengan mudah yang membuat mereka dapat membaca dan memahami teks meskipun teks tersebut memiliki kesalahan ejaan, sementara mesin tidak dapat melakukannya tanpa diinstruksikan atau dilatih (Zaky dan Romadhony, 2019).

Terdapat beberapa jenis *typographical error* yaitu *insertion*, *deletion*, *substitution* serta *transformation*. Dalam mendeteksi *typographical error* pada teks dibutuhkan suatu aplikasi disebut sistem koreksi ejaan atau *spelling checker*. Dimaknai dalam sebuah alur pemrosesan untuk menginvestigasi kalimat/kata apabila demikian didapatkan sebuah kekeliruan pada pengejaan serta menghadirkan opsional dalam wujud kata/kalimat umum hal tersebut disebutkan sebagai prosedur *spelling checker* (Soleh dan Purwarianti, 2011).

Beberapa teknik telah digunakan pada penelitian terdahulu yaitu N-Gram serta *Levenshtein Distances* (Fahma, Cholissodin dan Perdana, 2018), metode



N-gram dan *Boyer Moore* (Suriyanto, et al., 2020) dan metode *Damerau-Levenshtein Distance* (Christianti, et al., 2020).

Pada penelitian oleh Fahma, Cholissodin dan Perdana (2018) terdapat kekurangan pada metode N-gram dan *Levenshtein Distance* yang menghasilkan jumlah opsi kata kandidat yang ditampilkan oleh sistem terkadang menjadi terlalu banyak. Sedangkan untuk penelitian oleh Suriyanto, et al. (2020) yang menggunakan metode N-gram dan *Boyer Moore*, masih terdapat beraneka ragam kata sebagaimana dikelola secara salah atas aturan sebenarnya disebabkan oleh kalkulasi teknik N-Gram mendapatkan luaran sebagaimana disebutkan melampaui pada kata-kata lainnya. Penelitian yang dilakukan oleh Christianti, et al. (2020), menggunakan metode *Damerau-Levenshtein Distance* memberi hasil analisis koreksi otomatis memberikan saran yang salah. Hal ini dikarenakan saran kata pertama tidak selalu merupakan kata yang tepat untuk mengoreksi ejaan dalam koreksi otomatis.

Teknik *Long Short-Term Memory* (LSTM) merujuk kepada semacam pembelajaran mesin berdasarkan metodologi *Recurrent Neural Network* dimana memperkirakan kondisi mesin saat ini dengan memanfaatkan percikan mesin pengolah data berskala besar. *Long Short Term Memory* (LSTM) yakni peningkatan pada komputasi *Recurrent Neural Network* (RNN) dimana melibatkan penyesuaian RNN dengan menstimulasi sel memori atau unit memori kemudian mampu menyimpan data sebagaimana dimajukan lewat perantara LSTM untuk jangka waktu lama (Manaswi, 2018). RNN dan LSTM berpotensi untuk memiliki performa yang lebih baik karena fungsinya yang

menyusun bobot kata sehingga kata yang penting dapat disorot (Zhang et al, 2021).

Pada penelitian ini, metode Long Short-Term Memory akan digunakan dalam pengembangan sistem yang diharapkan mampu mendeteksi sekaligus mengoreksi kesalahan ejaan (*typographical error*).

### **1.3 Rumusan Masalah**

Berdasarkan latar belakang di atas, maka dapat dirumuskan masalah sebagai berikut:

1. Bagaimana mengidentifikasi dan mengoreksi kesalahan ejaan pada berita *online* dalam Bahasa Indonesia?
2. Bagaimana kinerja dari metode *Long Short-Term Memory* dalam melakukan pengoreksian ejaan Bahasa Indonesia pada berita *online*?

### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

1. Memahami cara mengidentifikasi dan mengoreksi kesalahan ejaan pada berita *online* dalam Bahasa Indonesia.
2. Mengembangkan perangkat lunak sistem pengoreksi ejaan Bahasa Indonesia menggunakan metode *Long Short-Term Memory*
3. Mengetahui kinerja dari metode *Long Short-Term Memory* dalam melakukan pengoreksian ejaan Bahasa Indonesia pada berita *online*.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini ialah:

1. Bagi Masyarakat, penelitian dapat memberi luaran dalam mengklasifikasikan salah ejaan pada berita *online* berbahasa Indonesia.
2. Bagi Peneliti Lain, penelitian dapat digunakan sebagai referensi untuk penelitian yang akan datang.
3. Bagi Institusi, penelitian dapat memberikan kontribusi berupa penambah wawasan dan bisa dijadikan referensi bagi kalangan mahasiswa.

### 1.6 Batasan Masalah

Adapun Batasan masalah pada riset ialah:

1. Deteksi kesalahan ejaan dilakukan menggunakan metode *Dictionary Lookup* yang mengacu pada Kamus Besar Bahasa Indonesia.
2. Kata yang tidak terdapat pada korpus Kamus Besar Bahasa Indonesia diasumsikan sebagai kesalahan ejaan.

### 1.7 Sistematika Penulisan

Sistematika penulisan proposal skripsi ini adalah sebagai berikut:

#### **BAB I. PENDAHULUAN**

Bagian ini memperjelas landasan, pembahasan topik, gagasan utama serta kebermanfaatan penyelidikan, batasan permasalahan/ruang lingkup, tata cara penelitian, termasuk sistematika penyusunannya.

## **BAB II. KAJIAN LITERATUR**

Elemen ini akan memperjelas landasan teori riset/eksperimen, yakni sistem koreksi ejaan, *preprocessing*, *Long Short-Term Memory*, dan *Rational Unified Process*.

## **BAB III. METODOLOGI PENELITIAN**

Elemen ini membahas beraneka ragam tahapan riset sebagaimana yang akan diimplementasikan. Setiap rencana tahap eksplorasi digambarkan secara luas di inti sari strukturnya. Bagian inip diakhiri dengan sebuah proyek dimana menguraikan pendekatan eksekutif dalam memandu ujian.

## **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Elemen ini akan mendemonstrasikan metode sebagaimana diungkapkan paling sering dimanfaatkan dalam pendayagunaan pemrograman Kerangka Revisi Ejaan Bahasa Indonesia untuk berita online dengan menggunakan *Long Short Term Memory*.

## **BAB V. HASIL DAN ANALISIS PENELITIAN**

Elemen ini menghadirkan temuan eksperimen/riset merujuk pada tahapan sebagaimana sudah diselesaikan peneliti. Kajian yang diberikan merupakan landasan bagi tujuan dimana hal tersebut mendasari riset ini.

## **BAB VI. KESIMPULAN DAN SARAN**

Bagian ini menawarkan kesimpulan serta konseptualisasi dimaksudkan agar berguna dalam analisis selanjutnya.

### **1.8 Kesimpulan**

Dari pendahuluan terkait, diuraikan secara umum pokok-pokok pikiran yang melandasi penelitian yang akan dilakukan, meliputi latar belakang dan rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

## DAFTAR PUSTAKA

- Akbar, R., Santoso, R. dan Warsito, B. 2022. Prediksi Tingkat Temperatur Kota Semarang Menggunakan Metode *Long Short-Term Memory* (LSTM), Jurnal Gaussian Volume 11 Number 4, (<https://ejournal3.undip.ac.id/index.php/gaussian/>, diakses 3 September 2023).
- Arinata, G. B., Satwika, I. P. dan Utami, N. W. 2022. Analisis Sentimen Pada EDOM STMIK Primakara Menggunakan Metode Long Short Term Memory, Jurnal Teknologi Informasi dan Komputer Volume 8 Number 4, (<https://jurnal.undhirabali.ac.id/index.php/jutik/article/>, diakses 23 April 2023).
- Fahma, A. I., Cholissodin, I., dan Perdana, R. S. 2017. Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer Volume 2 Number 1, (<https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/690>, diakses 29 Agustus 2023).
- Handayani, S. F., Pratiwi, R. W., Dairoh dan Af'idah, D. I. 2021. Analisis Sentimen Pada Data Ulasan Twitter Dengan Menggunakan Long Short Term Memory. Jurnal Teknologi Rekayasa Volume 7 Number 1, (<https://jtera.polteksmi.ac.id/index.php/jtera/article/>, diakses 24 April 2023).
- Hanifa, F. R. N. 2019. Coreference Resolution Bahasa Indonesia Pada Kata Ganti Kepemilikan Menggunakan Recurrent Neural Network. Skripsi Program Teknik Informatika FAKULTAS TEKNIK DAN ILMU KOMPUTER Universitas Komputer Indonesia.
- Hemmati-Sarapardeh, A., Larestani, A., Nait Amar, M., & Hajirezaie, S. (2020). Introduction. Applications of Artificial Intelligence Techniques in the Petroleum Industry, 1–22. (<https://www.sciencedirect.com/science/article/pii/B978012818680000011>, diakses 24 Juli 2024).
- Kambezidis, H. D. (2012). The Solar Resource. *Comprehensive Renewable Energy*, 27–84. (<https://www.sciencedirect.com/science/article/abs/pii/B9780080878720003024>, diakses 24 Juli 2024).

- Kurniawan, M., Kusriani, K., dan Arief, M. 2022. Part of Speech Tagging Pada Teks Bahasa Indonesia dengan BiLSTM + CNN + CRF dan ELMo. Jurnal Eksplora Informatika Volume 11 Number 1, (<https://doi.org/10.30864/eksplora.v11i1.506>, diakses 23 April 2023).
- Manaswi, N. K. 2018. Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras. Apress Berkeley, CA.
- Nurrohmat, M. A. dan Azhari, S. N. 2019. Sentiment Analysis of Novel Review Using Long Short-Term Memory Method. Indonesian Journal of Computing and Cybernetics Systems Volume 13 Number 3, (<https://jurnal.ugm.ac.id/ijccs/article/view/>, diakses 24 Agustus 2023).
- Normawati, D. dan Prayogi, S. A. 2021. Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. Jurnal Sains Komputer & Informatika Volume 5 Number 2, (<http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/>, diakses 24 Agustus 2023).
- Oktaria, D., Andayani, & Saddhono K. (2017). Penguasaan Kalimat Efektif sebagai Kunci Peningkatan Keterampilan Menulis Eksposisi. Jurnal Pendidikan Bahasa dan Sastra Indonesia Metalingua Volume 15 Number 2, (<https://journal.trunojoyo.ac.id/metalingua>, diakses 29 Agustus 2023).
- Pardede, J. dan Helgandamar, F. 2015. Aplikasi Peringkasan Dokumen Menggunakan Algoritma Interactive Graph-Based dan Similarity. Jurusan Teknik Informatika Institut Teknologi Nasional.
- Pessoa, M., and Anwar, A. 2014. A Review of RUP (Rational Unified Process). International Journal of Software Engineering Volume 5 Number 2, (<https://www.cscjournals.org/>, diakses 21 April 2023).
- Simanjuntak, M. S., Sujaini, H. dan Safriadi, N. 2018. *Spelling Corrector* Bahasa Indonesia dengan Kombinasi Metode Peter Norvig dan N-Gram. Jurnal Edukasi dan Penelitian Informatika Volume 4 Number 1, (<https://jurnal.untan.ac.id/index.php/jepin/article/>, diakses 25 Agustus 2023).
- Suleiman, D., and Awajan, A. A. 2019. Using Part of Speech Tagging for Improving Word2vec Model. 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), (doi:10.1109/ictcs.2019.8923081, diakses 24 Agustus 2023).

- Suprpto, Yuwono, K. T., Sukardiyono, T. dan Dewanto, A. 2008. Buku Bahasa Pemrograman Untuk SMK. Direktorat Pembinaan Sekolah Menengah Kejuruan, Indonesia.
- Surianto, D., Triyanto, D. and Ristian, U. Penerapan Algoritma Boyer Moore Dan Metode N-Gram Pada Aplikasi Penyunting Naskah Teks Bahasa Indonesia Berbasis Web. Coding : Jurnal Komputer dan Aplikasi Volume 8 Number 3, (<https://jurnal.untan.ac.id/index.php/jcskommipa/article/view/>, diakses 30 Agustus 2023).
- Zaky, D., and Romadhony, A. 2019. An LSTM-based Spell Checker for Indonesian Text. International Conference of Advanced Informatics: Concepts, Theory and Applications pp. 1-6, (<https://doi.org/10.1109/ICAICTA.2019.8904218>, diakses 22 April 2023).
- Zhang, Y., Wang, J., Zhang, X. 2021. Conciseness is better: Recurrent attention LSTM model for document-level sentiment analysis. Neurocomputing Volume 462, (<https://www.sciencedirect.com/science/article/>, diakses 24 April 2023).