

EXTRACTIVE SUMMARIZATION BERITA BERBAHASA INDONESIA MENGGUNAKAN LSTM

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



Oleh :

Alif Akbar Muhammad Isra
NIM 09021382025134

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI

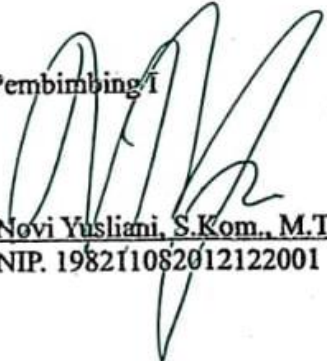
EXTRACTIVE SUMMARIZATION BERITA BERBAHASA
INDONESIA MENGGUNAKAN LSTM

Oleh :

Alif Akbar Muhammad Isra

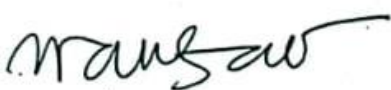
NIM : 09021382025134

Pembimbing I


Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Palembang, Agustus 2024

Pembimbing II


M. Naufal Rachmatullah, S.Kom., M.T.
NIP. 199212012022031008

Mengetahui,

Ketua Jurusan Teknik Informatika



TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari rabu tanggal 31 Juli 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Alif Akbar Muhammad Isra

NIM : 09021382025134

Judul : Extractive Summarization Berita Berbahasa Indonesia Menggunakan LSTM

Dan dinyatakan **LULUS**.

1. Ketua Penguji

Desty Rodiah, S.Kom., M.T.

198912212020122011



2. Penguji 1

Rizki Kurniati, S.Kom., M.T.

199107122019032016



3. Pembimbing 1

Novi Yusliani, S.Kom., M.T.

198211082012122001



4. Pembimbing 2

M. Naufal Rachmatullah, S.Kom., M.T

NIP. 199212012022031008



.....

Mengetahui,
Ketua Jurusan Teknik Informatika

Hadipurnawar Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Alif Akbar Muhammad Isra
NIM : 09021382025134
Program Studi : Teknik Informatika
Judul Skripsi : Extractive Summarization Berita Berbahasa Indonesia
Menggunakan LSTM

Hasil pengecekan Software *iThenticate/Turnitin* : 18 %

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

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapa pun.



Palembang, Agustus 2024



Alif Akbar Muhammad Isra
NIM 09021382025134

MOTTO :

"If the mind is willing, the flesh could go on and on without
many things."

— *Sun Tzu, The Art of War*—

Kupersembahkan karya tulis ini kepada:

- Allah SWT
- Orang tua ku tercinta
- Kakak-kakakku tersayang
- Dosen pembimbing dan penguji
- Sahabat dan teman seperjuangan
- Almamaterku

INDONESIAN NEWS EXTRACTIVE SUMMARIZATION USING LSTM

Alif Akbar Muhammad Isra
09021382025134


ABSTRACT

In the internet era, the vast amount of available information makes it challenging for readers to find relevant news and requires time to read entire documents to extract desired information. Therefore, summaries are needed to help readers extract and represent the most crucial information from news articles efficiently and effectively. This research focuses on developing an Extractive Summarization model for Indonesian news texts using Long Short-Term Memory (LSTM) networks due to their ability to maintain memory of word relationships. The designed model achieved a ROUGE-1 Precision of 0.3287, Recall of 0.6045, and F1-score of 0.4207 as well as a ROUGE-2 Precision of 0.2159, Recall of 0.4351, and F1-score of 0.2841, with a Loss of 0.09 and Validation Loss of 0.0898. The dataset comprised 18,952 data points, with 14,469 for training, 2,241 for validation, and 2,242 for testing.


Keywords: Automatic Text Summarization, Extractive Summarization, LSTM, Natural Language Processing, Indonesian News Text, ROUGE Metrics

Palembang, August 2024

Supervisor I


Novi Yusliani, S.Kom., M.T.
NIP. 198211082072122001

Supervisor II


M. Naufal Rachmatullah, S.Kom., M.T.
NIP. 199212012022031008

Approved,

Head of Informatics Department



EXTRACTIVE SUMMARIZATION BERITA BERBAHASA INDONESIA MENGGUNAKAN LSTM


Alif Akbar Muhammad Isra
09021382025134

ABSTRAK

Di era internet, banyaknya informasi yang tersedia menyebabkan sulit bagi pembaca untuk menemukan berita yang relevan dan membutuhkan waktu untuk membaca seluruh dokumen guna mendapatkan informasi yang diinginkan. Oleh karena itu, ringkasan dibutuhkan agar pembaca dapat mengekstrak dan merepresentasikan informasi paling krusial dari berita dengan efektif dan efisien. Penelitian ini berfokus pada pengembangan model Extractive Summarization untuk teks berita berbahasa Indonesia menggunakan jaringan Long Short-Term Memory (LSTM) karena kemampuannya dalam mempertahankan memori hubungan antar kata. Model yang dirancang mencapai ROUGE-1 Precision sebesar 0,3287, Recall sebesar 0,6045, dan F1-skor sebesar 0,4207, serta ROUGE-2 Precision sebesar 0,2159, Recall sebesar 0,4351, dan F1-skor sebesar 0,2841, dengan nilai Loss sebesar 0,09 dan Validation Loss sebesar 0,0898. Dataset terdiri dari 18.952 data, dengan 14.469 untuk pelatihan, 2.241 untuk validasi, dan 2.242 untuk pengujian.


Kata Kunci: *Automatic Text Summarization, Extractive Summarization, LSTM, Natural Language Processing, Teks Berita Berbahasa Indonesia, Metrik ROUGE*

Pembimbing I


Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Palembang, Agustus 2024

Pembimbing II


M. Naufal Rachmatullah, S.Kom., M.T.
NIP. 199212012022031008

Mengetahui,

Ketua Jurusan Teknik Informatika



KATA PENGANTAR

Puji dan syukur penyusun panjatkan kepada Tuhan yang Maha Esa karena dengan rahmat dan karunia-Nya, penyusun diberi kesempatan untuk menyusun dan menyelesaikan Tugas Akhir yang berjudul “Extractive Sumarization Berita Berbahasa Indonesia Menggunakan LSTM” sebagai salah satu syarat untuk menyelesaikan Pendidikan Strata-1 Teknik Informatika pada Fakultas Ilmu Komputer, Universitas Sriwijaya.

Untuk selanjutnya, penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, yaitu:

1. Bapak Prof. DR. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Ibu Novi Yusliani, S.Kom., M.T. selaku selaku dosen pembimbing I dan Bapak M. Naufal Rachmatullah, S.Kom., M.T. selaku dosen pembimbing II yang telah mengarahkan dan memberi *feedback* dalam menyusun tugas akhir.
4. Ibu Desty Rodiah, S.Kom., M.T. selaku dosen Ketua penguji dan Ibu Rizki Kurniati, S.Kom., M.T. selaku dosen penguji I yang telah memberikan koreksi dan masukan untuk tugas akhir.
5. Bapak Osvari Arsalan, S.Kom., M.T. selaku dosen pembimbing akademik selama kuliah.
6. Seluruh dosen program studi Teknik Informatika Universitas Sriwijaya yang memberikan banyak ilmu, pengalaman, hingga motivasi.
7. Seluruh staff dan pegawai yang telah membantu administrasi dalam perkuliahan.
8. Seluruh pihak lainnya yang tidak dapat disebutkan satu per satu.

Penyusun menyadari bahwa tugas akhir ini masih memiliki kekurangan. Oleh karena itu, penyusun mohon maaf apabila terdapat kesalahan maupun kata yang kurang berkenan. Penyusun juga terbuka terhadap kritik dan saran yang membangun. Semoga tugas akhir ini dapat bermanfaat.

Palembang, Agustus 2024



Alif Akbar Muhammad Isra

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN.....	iv
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-3
1.6 Batasan Penelitian	I-4
1.7 Sistematika Penulisan	I-4
1.8 Kesimpulan	I-4
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Extractive Summarization.....	II-1
2.2.2 LSTM	II-3
2.2.3 Preprocessing	II-7
2.2.4 Word Embedding	II-9
2.2.5 ROUGE	II-12
2.2.6 Rational Unified Process (RUP)	II-13
2.3 Penelitian Relevan	II-14
2.4 Kesimpulan	II-15
BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1
3.3 Tahapan Penelitian	III-3
3.3.1 Studi Literatur	III-3
3.3.2 Mengumpulkan Data	III-3
3.3.3 Kerangka Kerja	III-4
3.3.4 Merancang Arsitektur Model	III-5

3.3.5 Menentukan Kriteria Pengujian	III-6
3.3.6 Alat Bantu Pengujian.....	III-7
3.3.7 Format Data Pengujian	III-7
3.3.8 Pengujian Penelitian	III-8
3.3.9 Analisis Hasil Pengujian Penelitian	III-9
3.4 Metode Pengembangan Perangkat Lunak	III-10
3.4.1 Fase Insepsi	III-10
3.4.2 Fase Elaborasi	III-10
3.4.3 Fase Konstruksi	III-11
3.4.4 Fase Transisi.....	III-11
3.5 Manajemen Proyek Penelitian.....	III-11
3.6 Kesimpulan	III-15
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 Rational Unified Process (RUP)	IV-1
4.2.1 Fase Insepsi	IV-1
4.2.2 Fase Elaborasi	IV-10
4.2.3 Fase Konstruksi	IV-19
4.2.4 Fase Transisi	IV-24
4.3 Kesimpulan	IV-26
BAB V HASIL DAN ANALISA PENELITIAN	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan	V-1
5.2.1 Konfigurasi Percobaan	V-1
5.2.2 Hasil pengujian.....	V-6
5.3 Kesimpulan	V-19
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran.....	VI-1
DAFTAR PUSTAKA	xiii
LAMPIRAN.....	xvi

DAFTAR TABEL

	Halaman
Tabel II-1. Contoh Casefolding.....	II-8
Tabel II-2. Contoh Tokenisasi.....	II-8
Tabel II-3. Contoh <i>Padded Sequence</i>	II-9
Tabel III-1. Dataset.....	III-1
Tabel III-2. Data Sampel.....	III-2
Tabel III-3. Format Skenario Pengujian	III-8
Tabel III-4. Format Data Hasil Pengujian <i>Loss</i> dan <i>Validation Loss</i>	III-8
Tabel III-5. Format Data Hasil Pengujian evaluasi Rouge-N	III-9
Tabel III-6. Work Breakdown Structure (WBS).....	III-12
Tabel IV-1. Kebutuhan Fungsional	IV-3
Tabel IV-2. Kebutuhan Non Fungsional	IV-3
Tabel IV-3. Definisi Aktor.....	IV-6
Tabel IV-4. Definisi <i>Use Case</i>	IV-6
Tabel IV-5. Skenario <i>Use Case</i> Melakukan Pelatihan Data	IV-7
Tabel IV-6. Definisi <i>Use Case</i> Membuat Ringkasan	IV-8
Tabel IV-7. Definisi <i>Use Case</i> Mengevaluasi Hasil Ringkasan	IV-8
Tabel IV-8. <i>Library</i> yang digunakan	IV-20
Tabel IV-9. Implementasi Kelas	IV-20
Tabel IV-10. Rancangan Pengujian <i>Use Case</i>	IV-25
Tabel IV-11. Pengujian <i>Use Case</i>	IV-25
Tabel V-1. Dataset	V-1
Tabel V-2. Konfigurasi Parameter Tetap.....	V-4
Tabel V-3. Skenario Pengujian	V-5
Tabel V-4. Hasil Pengujian <i>Loss</i> dan <i>Validation Loss</i>	V-6
Tabel V-5. Jumlah kalimat dalam Ringkasan Dataset Test	V-11
Tabel V-6. Data <i>Samples</i>	V-11
Tabel V-7. Contoh Artikel dengan Ringkasan Referensi Jumlah Kalimat 2	V-13
Tabel V-8. Contoh-Artikel dengan Ringkasan Referensi Jumlah Kalimat 3	V-14
Tabel V-9. Contoh Artikel dengan Ringkasan Referensi Jumlah Kalimat 4	V-15
Tabel V-10. Contoh Artikel dengan Ringkasan Referensi Jumlah Kalimat 5.....	V-16
Tabel V-11. Contoh Generasi Ringkasan Bagian 1.	V-16
Tabel V-12. Contoh Generasi Ringkasan Bagian 2	V-17
Tabel V-13. Contoh Generasi Ringkasan Bagian 3.....	V-17
Tabel V-14. Contoh Generasi Ringkasan Bagian 4.....	V-16
Tabel V-15. Hasil Pengujian Rouge-N dengan Meringkas 30 % Jumlah kalimat Teks Masukan.....	V-18
Tabel V-16. Hasil Pengujian Rouge-N Jumlah Kalimat Sama dengan Ringkasan Referensi.....	V-19

DAFTAR GAMBAR

	Halaman
Gambar II-1. Arsitektur <i>Extractive Summarization</i>	II-2
Gambar II-2. Komponen LSTM	II-4
Gambar II-3. <i>LSTM Architecture</i>	II-5
Gambar II-4. Ilustrasi <i>Word Embedding</i>	II-9
Gambar II-5. Arsitektur CBOW	II-11
Gambar II-6. Arsitektur <i>Skip Gram</i>	II-12
Gambar II-7. Diagram Proses RUP	II-13
Gambar III-1. Diagram Tahapan Penelitian	III-3
Gambar III-2. Pipeline dari model penelitian.....	III-4
Gambar III-3. Arsitektur Model.....	III-6
Gambar IV-1. Diagram <i>Use Case</i>	IV-5
Gambar IV-2. Rancangan <i>Interface</i> Bagian Membuat Ringkasan	IV-11
Gambar IV-3. Rancangan <i>Interface</i> Bagian Evaluasi Hasil Ringkasan.....	IV-12
Gambar IV-4. Diagram <i>Activity</i> Skenario 1 Melakukan Pelatihan Data	IV-14
Gambar IV-5. Diagram <i>Activity</i> Skenario 2 Membuat Ringkasan.....	IV-15
Gambar IV-6. Diagram <i>Activity</i> Skenario 3 Mengevaluasi Hasil Ringkasan	IV-16
Gambar IV-7. Diagram <i>Sequence</i> Skenario 1 Melakukan Pelatihan Data	IV-17
Gambar IV-8. Diagram <i>Sequence</i> Skenario 2 Membuat Ringkasan.	IV-17
Gambar IV-9. Diagram <i>Sequence</i> Skenario 3 Mengevaluasi Hasil Ringkasan	IV-18
Gambar IV-10. Diagram <i>Class</i>	IV-19
Gambar IV-11. <i>Interface</i> bagian 1	IV-22
Gambar IV-12. <i>Interface</i> bagian 2.....	IV-22
Gambar IV-13. <i>Interface</i> bagian 3.....	IV-22
Gambar IV-14. <i>Interface</i> bagian 4.....	IV-23
Gambar IV-15. <i>Interface</i> bagian 5.....	IV-23
Gambar IV-16. <i>Interface</i> bagian 6.....	IV-24
Gambar V-1. Pelatihan <i>Loss</i> Skenario M1	V-7
Gambar V-2. Pelatihan <i>Loss</i> Skenario M2... ..	V-7
Gambar V-3. Pelatihan <i>Loss</i> Skenario M3.....	V-8
Gambar V-4. Pelatihan <i>Loss</i> Skenario M4.....	V-8
Gambar V-5. Pelatihan <i>Loss</i> Skenario M5.....	V-9
Gambar V-6. Pelatihan <i>Loss</i> Skenario M6.....	V-9
Gambar V-7. Pelatihan <i>Loss</i> Skenario M7.....	V-10
Gambar V-8. Pelatihan <i>Loss</i> Skenario M8.....	V-10
Gambar V-9. Pelatihan <i>Loss</i> Skenario M9.....	V-11
Gambar V-10. Pelatihan <i>Loss</i> Skenario M10.....	V-11

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab pendahuluan membicarakan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, dan sistematika penulisan. Selain itu, Bab membicarakan secara umum mengenai keseluruhan penelitian. Dimulai dengan penjelasan mengenai objek penelitian, serta metode penyelesaian.

1.2 Latar Belakang

Di era internet yang memudahkan penyebaran informasi menyebabkan banyak portal berita yang bermunculan. Akibat dari banyaknya berita yang tersedia, pembaca sulit mencari berita yang sesuai dan membutuhkan waktu untuk membaca semua isi dari dokumen untuk mendapatkan informasi yang ingin diketahui. Oleh karena itu, dibutuhkan ringkasan agar pembaca dapat mengekstrak dan merepresentasikan informasi paling krusial dari berita dengan efektif dan efisien.

Ringkasan dibutuhkan dalam mengekstrak informasi secara sekilas dari semua bagian berita. Pembuatan ringkasan dapat dilakukan dengan mudah apabila pembaca sudah terbiasa membaca teks atau apabila teks tersebut memiliki isi yang tidak terlalu panjang. Hasil ringkasan dari setiap orang berbeda. Hal ini disebabkan, pembaca menuangkan isi pikirannya secara subjektif sedangkan untuk menghasilkan ringkasan yang memenuhi standar dan berkualitas tinggi perlu ada acuan dan pertimbangan secara objektif. Selain itu pembuatan ringkasan juga memerlukan waktu sehingga diperlukan sebuah sistem yang dapat meringkas teks secara efektif dan efisien.

Text summarization adalah salah satu fokus dalam *Natural Language Processing* (NLP) yang membahas mengenai pembuatan peringkasan teks. *Automatic Text summarization* berfokus untuk mereduksi semua isi teks menjadi teks yang hanya berisi hal-hal penting. Salah satu metode *Text summarization* adalah *Extractive Summarization*.

Extractive Summarization adalah salah satu Teknik *Automatic Text summarization* (ATS) yang dirancang untuk mengekstraksi informasi penting dari sebuah dokumen dengan memilih sebagian kecil kata, frasa, atau kalimat yang paling akurat menggambarkan inti dari konten di teks (Mendoza et al., 2014). Salah satu metode *Extractive Summarization* yang dapat digunakan adalah prediksi menggunakan LSTM.

Berdasarkan penelitian terdahulu metode LSTM telah terbukti menjadi teknik yang sangat efektif dalam melakukan ekstraksi ringkasan secara *extractive* dengan menggunakan prediksi. LSTM, sebagai salah satu jenis metode *Recurrent Neural Network* (RNN), digunakan untuk menemukan hubungan antara kata-kata yang muncul dalam teks. Khususnya, LSTM memiliki keunggulan dalam mempertahankan memori terhadap hubungan antar kata-kata, sehingga sangat cocok untuk digunakan dalam proses *speech-to-text* dan *summarization* (Nallapati et al., 2016). Thakare dan Voditel (2022) melakukan penelitian dengan menggunakan dataset artikel berita CNN untuk melatih model LSTM dalam melakukan prediksi. Penelitian tersebut berhasil mencapai F1-Skor rata-rata sebesar 0,835, Sementara itu, Anjanadevi dan Nagakishore (2021) menggunakan dataset dari Arxiv untuk membandingkan performa antara LSTM dan GRU. Hasil analisis menunjukkan bahwa model LSTM unggul dalam berbagai metrik evaluasi, seperti akurasi, F1-skor, *Recall* dan presisi.

Temuan ini menandakan bahwa model LSTM lebih efektif dalam mengolah dan menganalisis data yang dikumpulkan jika dibandingkan dengan model GRU, Penelitian lain yang dilakukan oleh Khan et al., (2023) menggunakan dataset dari Ubuntu Discussion Online Forum dan TripAdvisor Online Forum. Evaluasi terhadap F1-skor, *Recall*, dan presisi menunjukkan bahwa LSTM mengungguli CBOW dalam melakukan *summarization*.

Penelitian ini menggunakan metode LSTM sebagai dasar Pembangunan model *Extractive Summarization* pada teks berita Berbahasa Indonesia. Metode LSTM digunakan karena memiliki keunggulan dalam mempertahankan memori terhadap hubungan antar kata-kata, sehingga sangat cocok untuk digunakan dalam proses summarization.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini sebagai berikut :

1. Bagaimana mengembangkan model *Extractive Summarization* berita Berbahasa Indonesia dengan menggunakan LSTM ?
2. Bagaimana kinerja LSTM dalam melakukan *Extractive Summarization* pada teks berita berbahasa Indonesia berdasarkan skor ROUGE-1 dan ROUGE-2 ?

1.4 Tujuan Penelitian

Tujuan dari penelitian adalah untuk mengembangkan suatu sistem yang dapat mengekstraksi ringkasan dari teks berita berbahasa Indonesia secara otomatis dengan menggunakan metode LSTM. Tujuan spesifik dari penelitian ini adalah:

1. Menghasilkan model LSTM untuk melakukan *Extractive Summarization* pada teks berita berbahasa Indonesia.
2. Mengetahui kinerja LSTM dalam melakukan *Extractive Summarization* pada teks berita berbahasa Indonesia berdasarkan skor ROUGE-1 dan ROUGE-2.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Menghasilkan model yang dapat mempercepat proses meringkas berita.
2. Hasil penelitian dapat digunakan sebagai referensi berdasarkan untuk penelitian di masa yang akan datang.

1.6 Batasan Penelitian

Batasan pada penelitian ini sebagai berikut :

1. Dataset hanya bersumber dari dataset Indosum yang datanya terakhir diperbarui tahun 2019 pada laman Kaggle dengan panjang kalimat artikel.

1.7 Sistematika Penulisan

Sistematika Penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Dalam bab ini dibahas mengenai latar belakang, rumusan masalah, tujuan, dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Dalam bab ini dibahas mengenai dasar teori, definisi dari istilah terkait, tujuan, dan manfaat penelitian, seperti penjelasan teori mengenai LSTM, arsitektur LSTM dan sebagainya.

BAB III METODE PENELITIAN

Bab ini menjelaskan tahapan yang dilakukan untuk menyelesaikan penelitian ini. Seperti *Preprocessing* data, pembuatan arsitektur model, dan evaluasi data secara detail.

1.8 Kesimpulan

Bab pendahuluan telah membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, dan sistematika penulisan .

DAFTAR PUSTAKA

- Anjanadevi, B., & Nagakishore, S. (2021). Optimized Text Summarization using Customized Recurrent Neural Network Model. In *Turkish Journal of Computer and Mathematics Education* (Vol. 12, Issue 14).
- Anwar, A. (2014). *A Review of RUP (Rational Unified Process)*.
<https://api.semanticscholar.org/CorpusID:203560301>
- Aries, A., Zegour, D. eddine, & Hidouci, W. K. (2019). *Automatic text summarization: What has been done and what has to be done*.
<http://arxiv.org/abs/1904.00688>
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural Machine Translation by Jointly Learning to Align and Translate*. <http://arxiv.org/abs/1409.0473>
- Bengio, S., Vinyals, O., Jaitly, N., & Shazeer, N. (2015). *Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks*.
<http://arxiv.org/abs/1506.03099>
- Dwarampudi, M., & Reddy, N. V. S. (2019). *Effects of padding on LSTMs and CNNs*. <http://arxiv.org/abs/1903.07288>
- Gerhana, Y. A., Atmadja, A. R., Zulfikar, W. B., & Ashanti, N. (2017). The implementation of K-nearest neighbor algorithm in case-based reasoning model for forming automatic answer identity and searching answer similarity of algorithm case. *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, 1–5.
<https://doi.org/10.1109/CITSM.2017.8089233>
- Graves, A. (2013). *Generating Sequences With Recurrent Neural Networks*.
<http://arxiv.org/abs/1308.0850>
- Grefenstette, G. (1999). Tokenization. In H. van Halteren (Ed.), *Syntactic Wordclass Tagging* (pp. 117–133). Springer Netherlands.
https://doi.org/10.1007/978-94-015-9273-4_9
- Gunawan, D., Pasaribu, A., Rahmat, R. F., & Budiarto, R. (2017). Automatic Text Summarization for Indonesian Language Using TextTeaser. *IOP Conference Series: Materials Science and Engineering*, 190(1), 12048.
<https://doi.org/10.1088/1757-899X/190/1/012048>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jang, B., Kim, I., & Kim, J. W. (2019). Word2vec convolutional neural networks for classification of news articles and tweets. *PLoS ONE*, 14(8).
<https://doi.org/10.1371/journal.pone.0220976>

- Khan, M. Y., Qayoom, A., Nizami, M. S., Siddiqui, M. S., Wasi, S., & Raazi, S. M. K.-R. (2021). Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques. *Complex.*, 2021, 2553199:1-2553199:18. <https://api.semanticscholar.org/CorpusID:239045248>
- Khattak, F. K., Jeblee, S., Pou-Prom, C., Abdalla, M., Meaney, C., & Rudzicz, F. (2019). A survey of word embeddings for clinical text. *Journal of Biomedical Informatics*, 100, 100057. <https://doi.org/https://doi.org/10.1016/j.yjbinx.2019.100057>
- Koto, F. (n.d.). *A Publicly Available Indonesian Corpora for Automatic Abstractive and Extractive Chat Summarization*. <http://www.statista.com/>
- Kumari, N., Sharma, N., & Singh, P. (2022). Performance of Optimizers in Text Summarization for News Articles. *Procedia Computer Science*, 218, 2430–2437. <https://doi.org/10.1016/j.procs.2023.01.218>
- Kurniawan, K., & Louvan, S. (2019). IndoSum: A New Benchmark Dataset for Indonesian Text Summarization. *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*, 215–220. <https://doi.org/10.1109/IALP.2018.8629109>
- Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries.
- Mani, I., & Maybury, M. T. (1999). *Advances in Automatic Text Summarization*. MIT Press.
- Mendoza, M., Bonilla, S., Noguera, C., Cobos, C., & León, E. (2014). Extractive single-document summarization based on genetic operators and guided local search. *Expert Systems with Applications*, 41(9), 4158–4169. <https://doi.org/10.1016/J.ESWA.2013.12.042>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. <http://arxiv.org/abs/1301.3781>
- Nallapati, R., Zhai, F., & Zhou, B. (2016). *SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents*. <http://arxiv.org/abs/1611.04230>
- Nawang Sari, R. P., Kusumaningrum, R., & Wibowo, A. (2019). Word2Vec for Indonesian Sentiment Analysis towards Hotel Reviews: An Evaluation Study. *Procedia Computer Science*, 157, 360–366. <https://doi.org/https://doi.org/10.1016/j.procs.2019.08.178>

- Sirohi, N. K., Bansal, Dr. M., & Rajan, Dr. S. N. R. (2021). Text Summarization Approaches Using Machine Learning & LSTM. *Revista Gestão Inovação e Tecnologias*, 11(4), 5010–5026.
<https://doi.org/10.47059/revistageintec.v11i4.2526>
- Thakare, A., & Voditel, P. (2022). Extractive Text Summarization Using LSTM-Based Encoder-Decoder Classification. *ECS Transactions*, 107, 11665–11672. <https://doi.org/10.1149/10701.11665ecst>
- Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A Review on the Long Short-Term Memory Model. *Artificial Intelligence Review*, 53.
<https://doi.org/10.1007/s10462-020-09838-1>
- Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C.-C. J. (2019). *Evaluating Word Embedding Models: Methods and Experimental Results*.
<https://doi.org/10.1017/ATSIP.2019.12>
- Wijayanti, R., Khodra, M. L., & Widyanoro, D. H. (2021). Single Document Summarization Using BertSum and Pointer Generator Network. *International Journal on Electrical Engineering and Informatics*, 13(4).
<https://doi.org/10.15676/ijeei.2021.13.4.10>
- Zamzam, M. A. (2020). SISTEM AUTOMATIC TEXT SUMMARIZATION MENGGUNAKAN ALGORITMA TEXTRANK. *MATICS*, 12(2), 111–116.
<https://doi.org/10.18860/mat.v12i2.8372>