

**PERINGKASAN OTOMATIS DONGENG NUSANTARA MENGGUNAKAN  
METODE LONG SHORT TERM MEMORY (LSTM)**

**SKRIPSI**

Oleh :  
**HAFIDZ NURROIS**  
**NIM. 210605110051**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**PERINGKASAN OTOMATIS DONGENG NUSANTARA MENGGUNAKAN  
METODE LONG SHORT TERM MEMORY (LSTM)**

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :  
**HAFIDZ NURROIS**  
**NIM. 210605110051**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**HALAMAN PERSETUJUAN**

**PERINGKASAN OTOMATIS DONGENG NUSANTARA MENGGUNAKAN  
METODE LONG SHORT TERM MEMORY (LSTM)**

**SKRIPSI**

Oleh :  
**HAFIDZ NURROIS**  
**NIM. 210605110051**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 05 Juni 2025

Pembimbing I,



Supriyono, M. Kom  
NIP. 19841010 201903 1 012

Pembimbing II,



Dr. Ir. Fachrul Kurniawan, M.MT., IPU  
NIP. 19771020 200912 1 001

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Ir. Fachrul Kurniawan, M.MT., IPU  
NIP. 19771020 200912 1 001

**HALAMAN PENGESAHAN**

**PERINGKASAN OTOMATIS DONGENG NUSANTARA MENGGUNAKAN  
METODE *LONG SHORT TERM MEMORY* (LSTM)**

**SKRIPSI**

Oleh :  
**HAFIDZ NURROIS**  
**NIM. 210605110051**

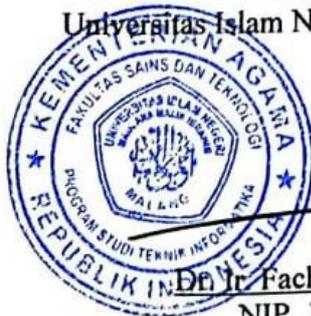
Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Tanggal: 23 Juni 2025

**Susunan Dewan Penguji**

Ketua Penguji	: <u>Ahmad Fahmi Karami, M.Kom</u> NIP. 19870909 202012 1 001	(  )
Anggota Penguji I	: <u>Nurizal Dwi Priandani, M.Kom</u> NIP. 19920830 202203 1 001	(  )
Anggota Penguji II	: <u>Supriyono, M.Kom</u> NIP. 19841010 201903 1 012	(  )
Anggota Penguji III	: <u>Dr. Ir. Fachrul Kurniawan, M.MT., IPU</u> NIP. 19771020 200912 1 001	(  )

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Ir. Fachrul Kurniawan, M.MT., IPU  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama : Hafidz Nurrois

NIM : 210605110051

Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Peringkasan Otomatis Dongeng Nusantara Menggunakan Metode *Long Short Term Memory* (LSTM)

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 17 Juni 2025

Yang membuat pernyataan



Hafidz Nurrois

NIM. 210605110051

## **MOTTO**

*~ Perlakukanlah Seseorang Sebagai mana kamu ingin diperlakukan ~*

*~ "Tidak perlu menjelaskan tentang dirimu kepada siapa pun, karena yang menyukaimu tidak butuh itu. Dan yang membencimu tidak akan percaya itu."*

## **HALAMAN PERSEMBAHAN**

Dengan segala keterbatasan dan rasa syukur yang mendalam, kupersembahkan karya ini untuk kedua orang tuaku tercinta. Terima kasih atas setiap doa yang terucap dalam diam, setiap lelah yang tak pernah terlihat, dan setiap air mata yang mungkin tak pernah ku tahu. Tanpa dukungan dan pengorbanan kalian, langkah ini tak akan pernah sampai di titik ini. Kupersembahkan juga untuk keluarga yang menjadi tempatku kembali, sahabat-sahabat yang selalu hadir di kala langkah mulai goyah, dan semua orang yang diam-diam menjadi penyemangat di perjalanan ini.

Ucapan terima kasih yang sebesar-besarnya juga saya sampaikan kepada seluruh dosen yang telah dengan tulus membagikan ilmu dan pengalaman selama masa perkuliahan. Terutama kepada dosen pembimbing yang dengan sabar mengarahkan dan memberikan motivasi ketika saya merasa lelah dan kehilangan arah. Semoga skripsi ini menjadi tanda kecil dari rasa hormat dan terima kasih saya atas semua perjuangan yang telah kita lewati bersama.

## KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, taufik, dan inayah-Nya yang telah mengiringi setiap langkah dalam proses penyusunan skripsi ini. Tanpa pertolongan dari-Nya, penyusunan karya ilmiah ini tidak akan mungkin dapat terselesaikan dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Proses penyusunan ini tidak lepas dari peran banyak pihak yang telah memberikan bantuan dalam berbagai bentuk, baik berupa semangat, bimbingan, maupun kontribusi ilmu pengetahuan.

Penulis menyadari bahwa capaian ini bukanlah hasil dari kemampuan pribadi semata, melainkan buah dari kolaborasi, dukungan, serta doa dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dan mendampingi selama proses penyusunan skripsi ini, baik secara langsung maupun tidak langsung. Ucapan terima kasih secara khusus penulis tujukan kepada:

1. Prof. Dr. H.M. Zainuddin, MA., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., IPU, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang
4. Supriyono, M.Kom, selaku pembimbing I, yang telah dengan sabar

membimbing, memberikan arahan, serta meluangkan waktu dalam setiap proses penyusunan skripsi ini dari awal hingga akhir. Nasihat dan masukan beliau sangat berarti.

5. Dr. Ir. Fachrul Kurniawan, M.MT., IPU, selaku pembimbing II, yang telah memberikan banyak saran selama proses penyusunan skripsi. Bimbingan beliau yang tegas namun membangun menjadi pendorong penting dalam penyelesaian skripsi ini.
6. Ahmad Fahmi Karami, M.Kom dan Nurizal Dwi Priandani, M.Kom, selaku dosen penguji, atas kritik, pertanyaan, serta masukan yang sangat berguna dalam memperbaiki dan menyempurnakan isi skripsi ini.
7. Segenap civitas akademika Program Studi Teknik Informatika, khususnya kepada para dosen yang telah berbagi ilmu, wawasan, dan inspirasi selama masa perkuliahan. Setiap pengajaran dan pengalaman menjadi fondasi penting dalam perjalanan akademik penulis.
8. Kedua orang tua tercinta, Bapak Ahmad Maki dan Ibu Nikmatul Khoiriyah, atas doa, dukungan moral dan materil, serta kasih sayang yang tiada henti. Semangat dan pengorbanan mereka adalah sumber kekuatan terbesar penulis dalam menyelesaikan studi ini.
9. Teman-teman seperjuangan, baik di dalam maupun di luar kampus, yang telah menjadi tempat berbagi semangat, pengalaman, dan tawa selama proses penulisan skripsi maupun selama menempuh studi
10. Staf administrasi Fakultas Sains dan Teknologi, yang telah membantu dalam kelancaran proses akademik dan penyusunan berkas-berkas yang diperlukan

selama masa studi.

11. Sayyed Amir Hasan atas segala bantuan yang telah diberikan selama proses penyusunan skripsi ini. Kehadiran dan bantuannya sangat berarti bagi saya.
12. Seluruh pihak lainnya yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan dalam bentuk apa pun selama penyusunan skripsi ini.

Sebagaimana karya ilmiah pada umumnya, skripsi ini tentu tidak luput dari keterbatasan, baik dari segi isi maupun metodologi. Oleh karena itu, penulis sangat terbuka terhadap kritik dan saran yang membangun dari para akademisi dan pembaca untuk penyempurnaan di masa mendatang. Penulis berharap karya ini dapat memberikan kontribusi positif, meskipun kecil, bagi pengembangan ilmu pengetahuan dan menjadi referensi yang berguna dalam bidang yang relevan

Malang, 30 Juni 2025

Penulis

## DAFTAR ISI

<b>HALAMAN PENGAJUAN</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>iv</b>
<b>PERNYATAAN KEASLIAN TULISAN</b> .....	<b>v</b>
<b>MOTTO</b> .....	<b>vi</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>vii</b>
<b>KATA PENGANTAR</b> .....	<b>viii</b>
<b>DAFTAR ISI</b> .....	<b>xi</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiii</b>
<b>DAFTAR TABEL</b> .....	<b>xiv</b>
<b>ABSTRAK</b> .....	<b>xv</b>
<b>ABSTRACT</b> .....	<b>xvi</b>
<b>مستخلص البحث</b> .....	<b>xvii</b>
<b>BAB I PENDAHULUAN</b> .....	<b>18</b>
1.1 Latar Belakang .....	18
1.2 Identifikasi Masalah .....	21
1.3 Tujuan Penelitian .....	21
1.4 Batasan Masalah .....	22
1.5 Manfaat Penelitian .....	23
<b>BAB II STUDI PUSTAKA</b> .....	<b>24</b>
2.1 Penelitian Terkait .....	24
2.2 Dongeng Nusantara.....	29
2.3 Ringkasan Otomatis .....	32
2.4 <i>Natural Language Processing (NLP)</i> .....	34
2.5 <i>Deep Learning</i> .....	36
2.6 <i>Recurrent Neural Network (RNN)</i> .....	38
2.7 <i>Long Short Term Memory (LSTM)</i> .....	40
2.8 <i>Recall-Oriented Understudy for Gisting (ROUGE)</i> .....	42
2.9 Python .....	44
2.10 Platform Youtube .....	45
<b>BAB III METODE PENELITIAN</b> .....	<b>47</b>
3.1 Desain Sistem.....	47
3.2 Pengumpulan Data .....	50
3.3 <i>Preprocessing Data</i> .....	50
3.3.1 Penghapusan tanda baca .....	51
3.3.2 <i>Case folding</i> .....	52
3.3.3 <i>Tokenizing</i> .....	52
3.3.4 <i>Stopword Removal</i> .....	53
3.3.5 <i>Stemming</i> .....	53
3.4 Implementasi metode <i>Long Short-Term Memory (LSTM)</i> .....	55
3.4.1 <i>Splitting Data</i> .....	60

3.4.2	<i>Sequence to Sequence</i> .....	61
3.4.3	Lapisan <i>Encoder</i> .....	63
3.4.4	Lapisan <i>Decoder</i> .....	63
3.4.5	<i>Attention Mechanism</i> .....	64
3.4.6	<i>Dense Layer</i> .....	65
3.5	Evaluasi.....	65
3.6	Skenario Pengujian .....	66
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>68</b>
4.1	Persiapan Data .....	68
4.2	Tahap Uji Coba .....	69
4.3	Konfigurasi Parameter LSTM.....	70
4.4	Skema Pengujian.....	74
4.5	Hasil Pengujian .....	75
4.5.1	Hasil Skenario 1: Menggunakan <i>Stemming</i> dan <i>Stopword Removal</i> .....	76
4.5.2	Hasil Skenario 2: Menggunakan <i>Stemming</i> .....	81
4.5.3	Hasil Skenario 3: Menggunakan <i>Stopword Removal</i> .....	85
4.5.4	Hasil Skenario 4: Tanpa <i>Stemming</i> dan <i>Stopword Removal</i> .....	91
4.6	Perbandingan Hasil .....	96
4.7	Hasil Implementasi Sistem .....	97
4.8	Pembahasan.....	98
4.9	Integrasi Islam.....	102
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>106</b>
5.1	Kesimpulan .....	106
5.2	Saran .....	107
<b>DAFTAR PUSTAKA</b>		
<b>LAMPIRAN</b>		

## DAFTAR GAMBAR

Gambar 2.1 Sel Pada LSTM .....	41
Gambar 3.1 Alur penelitian.....	47
Gambar 3.2 Desain perancangan sistem .....	48
Gambar 3.3 Flowchart Preprocessing .....	51
Gambar 3.5 Arsitektur LSTM.....	59
Gambar 3.6 Model Seq2Seq .....	62
Gambar 3.7 Alur <i>Attention Mechanism</i> .....	64
Gambar 4.1 Grafik Hasil Skenario 1 .....	77
Gambar 4.2 Grafik Hasil Skenario 2.....	83
Gambar 4.3 Grafik Hasil Skenario 3.....	87
Gambar 4.4 Grafik Hasil Skenario 4.....	92
Gambar 4.5 Hasil Implementasi Sistem.....	98

## DAFTAR TABEL

Tabel 2.1 Penelitian terkait .....	26
Tabel 3.1 Contoh preprocessing .....	54
Tabel 4.1 Tabel yang belum diringkas dan sudah diringkas .....	68
Tabel 4.2 Tabel Skenario pengujian.....	75
Tabel 4.3 Contoh Hasil <i>Preprocessing Stemming</i> dan stopword removal.....	76
Tabel 4.4 Hasil Evaluasi Matrik Skenario 1 .....	79
Tabel 4.5 Tabel Perbandingan Ringkasan manual dan mesin.....	79
Tabel 4.6 Contoh Hasil <i>Preprocessing Stemming</i> .....	81
Tabel 4.7 Hasil Metrik Evaluasi Skenario 2 .....	84
Tabel 4.8 Perbandingan Ringkasan Manual dan mesin (LSTM).....	84
Tabel 4.9 Hasil <i>Preprocessing Stopword Removal</i> .....	86
Tabel 4.10 Hasil Evaluasi Metrik Skenario 3 .....	89
Tabel 4.11 Hasil perbandingan Ringkasan Manual dan Mesin.....	89
Tabel 4.12 Contoh Teks Asli Tanpa <i>Preprocessing</i> .....	91
Tabel 4.13 Hasil Evaluasi Metrik Skenario 4 .....	94
Tabel 4.14 Tabel Perbandingan Ringkasan Manual dan Mesin.....	95
Tabel 4.15 Tabel hasil matrik dari semua skenario.....	96
Tabel 4.16 Penelitian sebelumnya mengenai peringkasan teks LSTM .....	101

## ABSTRAK

Nurrois, Hafidz, 210605110051, 2025, **Peringkasan Otomatis Dongeng Nusantara Menggunakan Metode *Long Short Term Memory (LSTM)***, Skripsi, Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Pembimbing : (I) Supriyono, M.Kom (II) Dr. Ir. Fachrul Kurniawan, M.MT., IPU

**Kata Kunci:** Dongeng Nusantara, Long Short-Term Memory, Peringkasan Teks Otomatis, Preprocessing

Dongeng merupakan bagian penting dari warisan budaya Nusantara yang sarat akan nilai moral dan pembelajaran. Namun, penyampaian dongeng dalam bentuk narasi panjang sering kali menyulitkan pembaca dalam menangkap inti cerita secara cepat. Oleh karena itu, dibutuhkan sistem peringkasan otomatis yang mampu menyederhanakan isi dongeng tanpa menghilangkan makna pentingnya. Penelitian ini bertujuan untuk menghasilkan ringkasan otomatis dari dongeng Nusantara menggunakan metode *Long Short-Term Memory* secara abstraktif, serta mengevaluasi pengaruh teknik *preprocessing* dengan cara *stemming* dan *stopword removal* terhadap akurasi model. Data dongeng diperoleh dari platform YouTube dan ditranskripsi ke dalam bentuk teks. Evaluasi dilakukan menggunakan metrik ROUGE-1, ROUGE-2, dan ROUGE-L dengan membandingkan hasil ringkasan mesin terhadap ringkasan ahli. Hasil terbaik diperoleh pada skenario dengan *stopword removal* tanpa *stemming*, dengan skor ROUGE-1 sebesar 0.606, ROUGE-2 0.3291, dan ROUGE-L sebesar 0.4613. Dibandingkan dengan penelitian terdahulu, model pada penelitian ini menunjukkan kinerja yang lebih baik. Temuan ini menegaskan pentingnya *preprocessing* dalam meningkatkan akurasi sistem peringkasan otomatis.

## ABSTRACT

Nurrois, Hafidz, 210605110051, 2025, **Automatic Summarization of Indonesian Folktales Using the Long Short Term Memory (LSTM) Method**, Thesis. Department of Informatics Engineering Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Supriyono, M.Kom (II) Dr. Ir. Fachrul Kurniawan, M.MT., IPU

**Kata Kunci:** Automatic Text Summarization, Indonesian Folktales, Long Short-Term Memory, Preprocessing

Folktales are an important part of the cultural heritage of the archipelago that is full of moral values and lessons. However, the delivery of folktales in the form of long narratives often makes it difficult for readers to grasp the essence of the story quickly. Therefore, an automatic summarization system is needed that is able to realize the contents of folktales without losing their important meaning. This study aims to produce an automatic summary of Nusantara folktales using the Long Short-Term Memory method abstractly, and to channel the influence of preprocessing techniques by stemming and stopword removal on model accuracy. Fairy tale data was obtained from the YouTube platform and transcribed into text form. The evaluation was carried out using the ROUGE-1, ROUGE-2, and ROUGE-L metrics by comparing the results of the machine summary to the expert summary. The best results were obtained in the scenario with stopword removal without stemming, with a ROUGE-1 score of 0.606, ROUGE-2 0.3291, and ROUGE-L of 0.4613. Compared to previous studies, the model in this study showed better performance. These findings emphasize the importance of preprocessing in improving the accuracy of automatic summarization systems.

## مستخلص البحث

نور الرئيس ، حافظ، . ٢٠٢٥. التلخيص الآلي للحكايات الشعبية الإندونيسية باستخدام **Long Short Term Memory (LSTM)**، أطروحة. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (أ) سوپريونو، الماجستير (ب) د. فخر الكورنيا وان، الماجستير

الكلمات المفتاحية: الحكايات الشعبية الإندونيسية، التلخيص الآلي للنصوص، المعالجة المسبقة.

عدّ الحكايات الشعبية جزءًا هامًا من التراث الثقافي للأرخبيل، وهي زاخرة بالقيم والعبر الأخلاقية. ومع ذلك، فإنّ تقديم الحكايات الشعبية في شكل سرديات طويلة غالبًا ما يُصعّب على القراء استيعاب جوهر القصة بسرعة. لذلك، هناك حاجة إلى نظام تلخيص تلقائي قادر على إدراك محتويات الحكايات الشعبية دون فقدان معناها المهم. تهدف هذه الدراسة إلى إنتاج ملخص تلقائي لحكايات نوسانتارا الشعبية باستخدام طريقة الذاكرة طويلة المدى قصيرة المدى بشكل تجريدي، وتوجيه تأثير تقنيات المعالجة المسبقة عن طريق إزالة الكلمات المتوقفة والمجدعية على دقة النموذج. تم الحصول على بيانات القصص الخيالية من منصة YouTube ونسخها إلى شكل نصي. تم إجراء التقييم باستخدام مقاييس ROUGE-1 و ROUGE-2 و ROUGE-L من خلال مقارنة نتائج الملخص الآلي بملخص الخبير. تم الحصول على أفضل النتائج في السيناريو مع إزالة الكلمات المتوقفة دون استخدام الجذر، مع درجة ROUGE-1 0.606 و ROUGE-2 0.3291 و ROUGE-L 0.4613. وبالمقارنة مع الدراسات السابقة، أظهر النموذج في هذه الدراسة أداءً أفضل. تؤكد هذه النتائج على أهمية المعالجة المسبقة في تحسين دقة أنظمة التلخيص التلقائي.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Dongeng merupakan salah satu bentuk karya sastra yang telah menjadi bagian penting dalam kehidupan masyarakat Indonesia. Dongeng adalah salah satu warisan budaya tradisi lisan dari Indonesia. Bahkan, Pramoedya Ananta Toer menyatakan bahwa dongeng merupakan salah satu bentuk medium yang paling indah dalam tradisi lisan di Nusantara (Rakhman et al., 2021). Meskipun cerita dalam dongeng tidak selalu menggambarkan kejadian yang sebenarnya tetapi nilai-nilai moral dan pelajaran hidup yang terkandung di dalamnya sangat bermanfaat bagi kehidupan manusia (Chamalah et al., 2021). Priyono Kusumo menyatakan bahwa dongeng adalah cerita imajinatif yang mungkin tampak tidak logis, tetapi mengandung informasi yang dapat memberikan manfaat bagi para pembacanya (Anggraeni & Rafiyanti, 2022).

Seiring dengan perkembangan zaman, dongeng kini tidak hanya hadir dalam bentuk tradisional yang disampaikan secara lisan, tetapi juga melalui berbagai media digital, seperti video di platform YouTube (Dwi Maharani & Nita priyanti, 2023). Namun, tantangan yang muncul adalah durasi dan panjangnya alur cerita dongeng yang terkadang tidak efisien dalam menyampaikan inti pesan kepada pembaca atau pendengar, terutama dalam konteks pendidikan. Oleh karena itu, diperlukan solusi berupa peringkasan teks yang dapat membantu menyederhanakan cerita dongeng tanpa menghilangkan makna dan pesan moralnya. Peringkasan teks

otomatis menjadi alternatif yang diandalkan untuk merangkum cerita dengan cepat dan efektif (Yuliska & Syaliman, 2020).

Peringkasan teks otomatis adalah metode yang digunakan untuk menghasilkan ringkasan singkat dengan tetap mempertahankan informasi penting dan makna keseluruhan (Bahari & Dewi, 2024). Peringkasan teks dapat dibedakan menjadi dua jenis: ekstraktif dan abstraktif. Peringkasan ekstraktif menciptakan ringkasan dengan memilih dan menggabungkan kalimat-kalimat yang dianggap paling representatif dari inti dokumen. Sementara itu, peringkasan abstraktif menghasilkan ringkasan dengan terlebih dahulu membangun representasi semantik internal, yang kemudian digunakan sebagai dasar untuk menciptakan ringkasan, dengan tujuan untuk mendekati kualitas ringkasan yang dibuat oleh manusia (Akbar, 2021).

Dalam peringkasan otomatis, tahap awal yang dilakukan adalah preprocessing data. Preprocessing merupakan proses untuk menghilangkan data yang tidak terstruktur dari data yang diperoleh. Oleh karena itu, *preprocessing* memiliki peran penting karena dapat memengaruhi akurasi dalam pembuatan ringkasan (Uysal & Gunal, 2014). Beberapa langkah yang termasuk dalam *preprocessing* antara lain *stopword removal* dan *stemming*. *Stopword removal* adalah kata-kata umum dalam bahasa yang tidak memberikan kontribusi makna dalam sebuah kalimat, sedangkan *stemming* adalah proses penghilangan imbuhan pada kata sehingga kata tersebut kembali ke bentuk dasarnya (Santosa et al., 2022).

Diharapkan dengan adanya sistem peringkasan teks otomatis ini, pembaca dapat lebih efisien dalam memahami inti cerita dari dongeng yang dibaca, sehingga

mereka dapat menangkap informasi penting dengan lebih cepat. Allah *subhanahu wa ta'ala* berfirman dalam QS. Al-Ankabut 43 yang berbunyi:

وَتِلْكَ الْأَمْثَالُ نَضْرِبُهَا لِلنَّاسِ وَمَا يَعْقِلُهَا إِلَّا الْعُلَمَاءُ

“Perumpamaan-perumpamaan itu Kami buat untuk manusia. Namun, tidak ada yang memahaminya, kecuali orang-orang yang berilmu.” (QS. Al-Ankabut:43)

Menurut tafsir Al-Muyassar Perumpamaan-perumpamaan ini Kami adakan bagi sekalian manusia, agar mereka dapat mengambil manfaat darinya dan mengambil pelajaran darinya. Dan tidak ada yang memahaminya kecuali orang yang mengenal Allah, ayat-ayat dan syariatNya. Dalam konteks dongeng, sistem peringkasan teks otomatis dapat membantu pembaca untuk memahami pesan dan nilai-nilai yang terkandung dalam cerita. Dengan demikian, pengetahuan yang diperoleh dari dongeng dapat menginspirasi dan mendidik pembaca, sejalan dengan pengajaran Al-Qur'an bahwa hikmah harus dapat diambil dan dipahami oleh mereka yang berusaha untuk belajar.

Dalam penelitian ini, peringkasan dongeng Nusantara akan dilakukan menggunakan metode *Long Short-Term Memory* secara abstraktif, LSTM memiliki mekanisme yang lebih fokus pada memori dibandingkan dengan langkah pelatihan sebelumnya untuk membangun hubungan antar kata, sehingga sangat cocok untuk pemrosesan teks seperti pengenalan suara, terjemahan, dan peringkasan (Yosia Wibowo et al., 2024). *Long Short Term Memory* adalah pengembangan dari salah satu algoritma *deep learning*, yaitu *Recurrent Neural Network* (RNN), yang dirancang untuk mengatasi salah satu kelemahan RNN, yakni dalam pengelolaan informasi selama periode yang panjang (Zahara et al., 2019). Diharapkan dongeng-

dongeng yang tersedia di platform digital dapat diringkas secara lebih efisien menghasilkan ringkasan teks sesuai harapan. Penelitian ini bertujuan untuk mengembangkan solusi yang tidak hanya mempermudah pemahaman cerita, tetapi juga mempertahankan nilai-nilai budaya dan moral yang terkandung dalam dongeng Nusantara.

Kebaruan dari penelitian ini terletak pada objek penerapan metode *Long Short-Term Memory* (LSTM) dalam peringkasan teks dongeng Nusantara secara otomatis, yang belum banyak dilakukan dalam penelitian-penelitian sebelumnya.

## **1.2 Identifikasi Masalah**

Berdasarkan uraian latar belakang yang telah disampaikan, penelitian ini difokuskan untuk mencari jawaban atas pertanyaan-pertanyaan berikut.

- a. Bagaimana pengaruh penggunaan *preprocessing stemming dan stopword removal* terhadap hasil peringkasan teks otomatis menggunakan LSTM?
- b. Bagaimana kinerja metode *Long Short-Term Memory* (LSTM) dalam menghasilkan ringkasan otomatis teks dongeng Nusantara?

## **1.3 Tujuan Penelitian**

Berdasarkan permasalahan yang telah diuraikan pada bagian latar belakang, tujuan dari penelitian ini adalah sebagai berikut:

- a. Mengetahui kinerja metode *Long Short-Term Memory* (LSTM) dalam melakukan peringkasan teks dongeng Nusantara secara otomatis.

- b. Mengetahui pengaruh penerapan teknik *preprocessing* (*stemming* dan *stopword removal*) terhadap hasil ringkasan teks otomatis menggunakan LSTM.

#### 1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan yang perlu ditegaskan guna memastikan fokus penelitian agar tetap terarah, antara lain:

- a. Penelitian ini dibatasi pada peringkasan teks dongeng Nusantara yang diperoleh dari platform YouTube dengan kata kunci "Dongeng Nusantara" dan dongeng nusantara yang ada di website.
- b. Metode peringkasan yang digunakan adalah *Long Short-Term Memory* (LSTM) dengan pendekatan abstraktif, tanpa membandingkan dengan metode peringkasan lainnya seperti ekstraktif atau metode machine learning yang berbeda.
- c. Dataset yang digunakan dalam penelitian ini dibatasi pada dongeng-dongeng yang tersedia dalam bentuk teks yang dihasilkan dari transkripsi video YouTube ataupun dari website, tanpa melakukan analisis pada jenis media lain seperti audio atau visual.
- d. Penelitian ini difokuskan pada evaluasi akurasi dan kualitas peringkasan teks, tanpa menganalisis aspek lain seperti kecepatan atau efisiensi komputasi sistem LSTM yang digunakan.

## 1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan mampu memberikan kontribusi yang bermanfaat baik dalam aspek teori maupun penerapan praktis, dengan rincian sebagai berikut:

- a. Memberikan kontribusi dalam pengembangan sistem peringkasan otomatis yang membantu siswa dan guru dalam memahami inti cerita dan nilai-nilai moral dari dongeng Nusantara secara lebih cepat dan efisien.
- b. Pelestarian dongeng Nusantara dengan menyajikan ringkasan yang lebih mudah diakses dan dipahami oleh generasi muda melalui platform digital.
- c. Menjadi referensi untuk pengembangan lebih lanjut di bidang *Natural Language Processing* (NLP), khususnya dalam penerapan metode *Long Short-Term Memory* (LSTM) untuk peringkasan teks berbasis bahasa Indonesia.
- d. Memudahkan pembaca untuk memahami esensi cerita dongeng Nusantara secara singkat, sehingga dapat menghemat waktu sambil tetap mendapatkan pesan moral dari cerita tersebut.

## BAB II

### STUDI PUSTAKA

#### 2.1 Penelitian Terkait

Penelitian yang dilakukan oleh (Alfhi Saputra, 2021) menggunakan metode *Long Short-Term Memory* (LSTM) untuk peringkasan teks secara otomatis dalam Bahasa Indonesia pada artikel berita daring. Hasil penelitian menunjukkan bahwa LSTM memberikan performa yang baik dalam pendekatan abstraktif dengan nilai evaluasi terbaik pada ROUGE-1 sebesar 0.13846 pada F1-score rata-rata, yang jauh lebih tinggi dibandingkan metode lain yang umumnya menggunakan pendekatan ekstraktif. Penelitian ini juga mengidentifikasi bahwa penggunaan *stopwords removal* memiliki pengaruh besar terhadap hasil performa sistem, meskipun *stemming* juga memberikan pengaruh, namun tidak signifikan.

Penelitian yang dilakukan oleh (Alfin et al., 2024) menunjukkan bahwa penggunaan metode RNN dengan variasi model CBOW menghasilkan kinerja yang lebih unggul dalam aspek presisi dan efisiensi peringkasan teks dibandingkan dengan model Skip-gram. Berdasarkan evaluasi menggunakan metrik Rouge-1, model CBOW memperoleh nilai recall, presisi, dan F-measure yang lebih tinggi, menjadikannya lebih cocok untuk aplikasi peringkasan otomatis yang membutuhkan ringkasan yang akurat dan relevan. Dengan kemampuan untuk menangkap inti dari berbagai sumber berita secara efisien, model CBOW sangat berguna dalam situasi di mana pengguna perlu memahami informasi utama dari berbagai dokumen secara cepat. Algoritma RNN dengan CBOW mampu

menyajikan ringkasan yang informatif dan akurat, memberikan solusi efektif terhadap tantangan akses informasi di tengah arus berita yang terus bertambah (Alfin et al., 2024).

Ditemukan dua penelitian terkait yang menyoroti perbandingan metode peringkasan teks otomatis pada teks berbahasa Indonesia dan Inggris. Penelitian pertama berjudul "Peringkasan Teks Otomatis Abstraktif Menggunakan Transformer pada Teks Bahasa Indonesia" (Bahari & Dewi, 2024). Studi ini menggunakan model Transformer untuk peringkasan teks otomatis abstraktif dan menunjukkan bahwa Transformer, dengan mekanisme attention yang unik, mengatasi kendala model recurrent seperti RNN atau LSTM dalam mempertahankan konteks input secara penuh. Evaluasi menggunakan dataset IndoSum menghasilkan nilai ROUGE-1 sebesar 0,61 dan ROUGE-2 sebesar 0,51, yang mengindikasikan bahwa Transformer mampu memberikan hasil ringkasan yang mendekati ringkasan manusia dengan efisiensi tinggi dalam menangani berbagai berita dalam jumlah besar (Bahari & Dewi, 2024).

Selain itu, penelitian kedua oleh (Abdillah, 2024) membandingkan efektivitas metode ekstraktif TextRank dan metode abstraktif LSTM dalam merangkum teks berita berbahasa Inggris. Hasil penelitian menunjukkan bahwa LSTM lebih unggul dalam hal recall, presisi, dan F1-score pada skenario dengan 90% data training. Meskipun demikian, model LSTM masih memiliki akurasi di bawah 0,5, menunjukkan bahwa optimasi lebih lanjut diperlukan untuk meningkatkan performa ringkasan yang dihasilkan. Penelitian ini menekankan bahwa TextRank lebih sesuai untuk tugas-tugas sederhana, sedangkan LSTM menawarkan potensi

yang lebih besar dalam menghasilkan ringkasan yang lebih mendalam dan kontekstual.

Berdasarkan beberapa penelitian yang membahas metode *Long Short-Term Memory* (LSTM), belum ada yang meneliti penerapannya dalam peringkasan otomatis teks dongeng nusantara. Penelitian sebelumnya yang dilakukan oleh (Fenardo et al., 2021). telah mengembangkan peringkasan otomatis untuk cerita anak Indonesia menggunakan metode *Maximum Marginal Relevance* (MMR), dengan mengambil cerita dari beberapa situs web, seperti [didongeng.blogspot.com](http://didongeng.blogspot.com), [bobo.grid.id](http://bobo.grid.id), dan [dongengceritarakyat.com](http://dongengceritarakyat.com). Hasil penelitian tersebut menunjukkan bahwa metode MMR mampu mencapai tingkat akurasi rata-rata sebesar 83,67%, dengan waktu pemrosesan sekitar 5 hingga 16 detik per dokumen, yang cukup efisien untuk menghasilkan ringkasan secara cepat. Namun, penelitian ini berbeda karena menggunakan pendekatan LSTM untuk menyederhanakan teks dongeng nusantara. Algoritma LSTM diharapkan dapat meningkatkan kualitas ringkasan dengan menangkap pola yang lebih kompleks dalam teks melalui pembelajaran mendalam, sehingga menghasilkan ringkasan yang lebih relevan dan akurat.

Tabel 2.1 Penelitian terkait

No	Peneliti	Judul	Metode	Hasil
1.	(Halimah et al., 2022)	Peringkasan Teks Otomatis ( <i>Automated Text Summarization</i> ) Pada Artikel Berbahasa Indonesia Menggunakan Algoritma Lexrank	LexRank	Pada tingkat kompresi 50%, nilai f-measure untuk metrik ROUGE-1, ROUGE-2, dan ROUGE-L tercatat masing-masing sebesar 67,53%, 59,10%, dan 67,05%. Sedangkan pada kompresi 30%, f-measure untuk ROUGE-1, ROUGE-2, dan ROUGE-L berturut-turut mencapai 55,82%, 45,51%, dan 54,76%.

No	Peneliti	Judul	Metode	Hasil
2.	(W. Pratama et al., 2022)	Peringkasan Otomatis Makalah Menggunakan Maximum Marginal Relevance	Maximum Marginal Relevance	Penelitian ini menghasilkan ringkasan terbaik dengan rata-rata nilai 0,68 pada pengujian ROUGE-1.
3.	(Idhafi et al., 2023)	Peringkasan teks otomatis pada artikel berbahasa Indonesia menggunakan metode maximum marginal relevance	maximum marginal relevance	Pengujian pada tingkat kompresi 50% menunjukkan F1-score masing-masing sebesar 71,86% untuk ROUGE-1, 64,18% untuk ROUGE-2, dan 71,56% untuk ROUGE-L. Sementara itu, pada tingkat kompresi 30%, F1-score yang diperoleh adalah 62,95% untuk ROUGE-1, 53,61% untuk ROUGE-2, dan 62,47% untuk ROUGE-L.
4.	(Alfin et al., 2024)	Peringkasan Multi Dokumen Berbahasa Indonesia Menggunakan Metode Recurrent Neural Network	Recurrent Neural Network	Hasil penelitian menunjukkan nilai recall, presisi, dan F-measure yang signifikan. Pada model CBOW, nilai recall, presisi, dan F-measure masing-masing sebesar 0,487, 0,704, dan 0,550. Sedangkan pada model Skip-gram, nilai recall, presisi, dan F-measure tercatat sebesar 0,414, 0,687, dan 0,504.
5.	(Alfhi Saputra, 2021)	Peringkasan Teks Otomatis Bahasa Indonesia secara Abstraktif Menggunakan Metode Long Short-Term Memory	Long Short-Term Memory	Pengujian dengan metode LSTM menghasilkan nilai terbaik pada evaluasi menggunakan ROUGE-1, yakni sebesar 0,13846.
6.	(Fauzi, 2022)	Penerapan Algoritma Text Mining dan Lexrank dalam Meringkas Teks Secara Otomatis	Text Mining dan Lexrank	Hasil ringkasan LexRank menunjukkan bobot tertinggi secara berurutan, yaitu $D2 = 1,433$ , $D10 = 1,289$ , $D3 = 1,253$ , hingga $D8 = 0,673$ .
7.	(Setiawan et al., 2022)	Sentiment Summarization Learning Evaluation Using LSTM ( <i>Long Short Term Memory</i> ) Algorithm	LSTM	Dalam pengujian menggunakan confusion matrix, sistem mencapai nilai akurasi sebesar 0,902 dan f-measure sebesar 0,921. Pada pengujian menggunakan Recall-Oriented

No	Peneliti	Judul	Metode	Hasil
				Understudy for Gisting Evaluation (ROUGE), evaluasi positif memperoleh skor 0,16, sedangkan evaluasi negatif mendapatkan skor 0,2.
8.	(AL-Hafiidh et al., 2022)	Peringkasan Teks Otomatis Pada Portal Berita Olahraga Menggunakan Metode Maximum Marginal Relevance	Maximum Marginal Relevance	Hasil dari penelitian ini menunjukkan bahwa dengan menggunakan nilai lambda 0,7, diperoleh rata-rata presisi sebesar 57,7%, recall sebesar 48,5%, dan F-measure sebesar 50,3%.
9.	(Ayu Syahfitri et al., 2022)	Penerapan Algoritma Maximum Marginal Relevance Dalam Peringkasan Teks Secara Otomatis	Maximum Marginal Relevance	Hasil ringkasan dari sistem peringkasan teks otomatis pada penelitian ini berupa kalimat utama yang memiliki kesamaan dengan query dan dipilih berdasarkan urutan bobot.
10.	(Rusbandi et al., 2021)	Otomatisasi Peringkasan Teks Pada Dokumen Hukum Menggunakan Metode Latent Semantic Analysis	Latent Semantic Analysis	Hasil yang diperoleh dari evaluasi peringkasan teks otomatis menggunakan metode Latent Semantic Analysis menunjukkan bahwa untuk compression rate 75%, nilai precision, recall, f-measure, dan accuracy masing-masing adalah 53%, 27%, 35%, dan 71%. Sementara itu, untuk compression rate 50%, nilai tersebut meningkat menjadi 54%, 56%, 55%, dan 75%. Sedangkan untuk compression rate 25%, hasilnya adalah 51%, 79%, 61%, dan 75%.
11.	(Setiawan et al., 2022)	Sentiment Summarization Evaluasi Pembelajaran Menggunakan Algoritma Lstm (Long Short Term Memory)	LSTM	Dalam pengujian menggunakan confusion matrix, sistem memperoleh nilai akurasi sebesar 0,902 dan nilai f-measure sebesar 0,921. Sementara itu, pada pengujian menggunakan metode Recall-Oriented Understudy for Gisting Evaluation (ROGUE),

No	Peneliti	Judul	Metode	Hasil
				evaluasi terhadap ringkasan positif menghasilkan skor 0,16, sedangkan evaluasi terhadap ringkasan negatif memperoleh skor 0,2.
12.	(Yosia Wibowo et al., 2024)	Implementasi <i>Long Short-Term Memory</i> Dalam Analisis Sentimen Pengguna Aplikasi Twitter Yang Mengandung Ujaran Kebencian	Long Short Term Memory	Sistem menghasilkan tingkat akurasi sebesar 83% dengan rata-rata makro (macro average) sebesar 65%.
13.	(Mawaridi, 2024)	Sistem Peringkasan Teks Berita Berbahasa Indonesia Menggunakan Latent Dirichlet Allocation Dan Maximum Marginal Relevance	Maximum Marginal Relevance	ROUGE-1 terbaik sebesar 0.488 untuk tingkat kompresi 50% dan 0.462 untuk tingkat kompresi 30%
14.	(Ahsan, 2023)	Peringkasan Teks Multi Dokumen Berita Berbahasa Indonesia Menggunakan Fasttext Dan K-Means Clustering	Fasttext Dan K-Means Clustering	Variasi 1 mendapat nilai rata-rata precision 0,674, rata-rata recall 0,620, dan rata-rata f-measure 0,637

## 2.2 Dongeng Nusantara

Dongeng Nusantara adalah kumpulan cerita rakyat yang diwariskan secara turun-temurun dari generasi ke generasi, dengan kandungan nilai-nilai moral dan budaya yang sangat mendalam dan berakar kuat dalam kehidupan masyarakat Indonesia (Fadhila et al., 2023). Dongeng tidak hanya berfungsi sebagai hiburan semata, tetapi juga memiliki peran strategis sebagai sarana pendidikan karakter bagi anak-anak sejak usia dini. Melalui alur cerita yang sederhana namun sarat makna, dongeng mengajarkan nilai-nilai penting seperti kejujuran, tanggung jawab, empati, dan kepedulian sosial. Misalnya, cerita “Putri Bunga Melur” merupakan contoh konkret yang menggambarkan pentingnya perhatian, kasih sayang, dan rasa peduli

terhadap orang lain di sekitar kita. Dongeng-dongeng ini memiliki kekuatan untuk menyentuh hati dan membentuk perilaku anak menjadi lebih baik, karena nilai-nilai yang disampaikan dikemas dalam bentuk cerita yang menarik dan mudah dipahami.

Dongeng juga mengandung nilai-nilai sastra secara intrinsik, yang berkontribusi besar dalam perkembangan estetika dan imajinasi anak-anak. Nilai-nilai tersebut mencakup fungsi hiburan yang menyenangkan, kemampuan untuk merangsang daya imajinasi dan kreativitas, memperkaya pengalaman emosional dan intelektual, membentuk wawasan anak menjadi lebih manusiawi, serta memperkenalkan berbagai ragam pengalaman universal yang berlaku dalam kehidupan masyarakat secara luas. Selain itu, dongeng berperan penting dalam mewariskan tradisi kesusastraan kepada generasi muda, yang pada gilirannya akan memperkuat identitas budaya bangsa. Sementara itu, dari sisi ekstrinsik, karya sastra termasuk dongeng memiliki pengaruh yang signifikan terhadap perkembangan sosial anak, pembentukan kepribadian, peningkatan kemampuan berpikir kritis dan reflektif, serta penguatan keterampilan berbahasa anak secara menyeluruh (Ayu Latifah & Suprayitno, 2021).

Menurut Kamus Besar Bahasa Indonesia (KBBI Daaring, 2022), dongeng diartikan sebagai kisah atau cerita yang bersifat fiktif, artinya tidak benar-benar terjadi di dunia nyata, dan umumnya berkaitan dengan peristiwa atau kejadian yang terjadi di masa lampau yang memiliki karakteristik unik, luar biasa, dan penuh keajaiban. Selain itu, istilah “dongeng” juga dapat merujuk pada ucapan, kabar, atau informasi yang tidak sesuai dengan kenyataan atau fakta yang sebenarnya. Penyampaian dongeng dapat dilakukan melalui berbagai media, baik secara lisan

dalam bentuk tutur cerita langsung maupun secara tulisan melalui buku atau media digital, dengan tujuan utama menyampaikan pesan tertentu kepada pendengarnya, baik secara eksplisit maupun implisit. Secara esensial, dongeng tidak hanya berfungsi sebagai sarana hiburan yang menyenangkan, tetapi juga memiliki nilai edukatif yang tinggi dan berperan penting dalam menanamkan nilai-nilai budaya, moral, serta cara pandang terhadap kehidupan. Oleh karena itu, kegiatan mendongeng dianggap sebagai bagian dari praktik budaya yang alami, organik, dan sangat dianjurkan untuk dilakukan sejak usia dini. Hal ini terbukti dengan dimasukkannya materi dongeng ke dalam kompetensi dasar dalam mata pelajaran Bahasa Indonesia pada kurikulum pendidikan nasional, yang menunjukkan pentingnya dongeng dalam sistem pembelajaran formal (Husniyah, 2022).

Dalam konteks yang lebih luas dan menyeluruh, dongeng Nusantara juga berfungsi sebagai media untuk memperkenalkan anak-anak pada kekayaan kearifan lokal dari berbagai daerah di Indonesia, yang mencerminkan keragaman budaya bangsa. Contohnya adalah kisah “Timun Mas” dari daerah Jawa yang mengajarkan keberanian dan kecerdikan, serta cerita “Batu Menangis” dari Kalimantan yang mengandung pesan tentang pentingnya menghormati orang tua. Dongeng-dongeng semacam ini tidak hanya berfungsi sebagai hiburan yang menarik, tetapi juga menjadi media efektif untuk menyampaikan pesan-pesan moral yang relevan dengan kehidupan sehari-hari serta mampu membentuk karakter anak sejak dini dengan cara yang menyenangkan (Murtiningsih, 2022). Pengaruh dongeng sebagai media pembelajaran terbukti sangat efektif dalam membangun karakter positif pada anak-anak karena mampu menjangkau sisi emosional dan kognitif mereka. Sebagai

salah satu alat edukasi yang kontekstual dan fleksibel, dongeng Nusantara membawa nilai-nilai moral dan budaya ke dalam lingkungan belajar yang lebih interaktif, komunikatif, dan menyenangkan. Dengan pendekatan ini, anak-anak akan lebih mudah menginternalisasi pesan-pesan positif yang terkandung dalam cerita rakyat, yang kemudian dapat diaplikasikan dalam kehidupan sehari-hari mereka (Dimas et al., 2023).

### **2.3 Ringkasan Otomatis**

Peringkasan otomatis merupakan sebuah metode dalam bidang pemrosesan *Natural Language Processing* yang bertujuan untuk mengekstraksi atau meringkas informasi inti dari satu atau lebih dokumen teks secara otomatis dengan bantuan teknologi dan algoritma tertentu. Metode ini sangat berguna untuk mempermudah proses pembacaan dan pemahaman informasi, terutama dalam konteks dokumen yang panjang dan kompleks. Dengan peringkasan otomatis, isi dokumen dapat disajikan dalam bentuk yang lebih ringkas, efisien, dan mudah dipahami, namun tetap mempertahankan poin-poin penting dan esensial dari isi aslinya (Halimah et al., 2022). Teknologi ini sangat relevan di era informasi saat ini, di mana pengguna dituntut untuk menyerap informasi dalam waktu yang singkat.

Menurut (AL-Hafiidh et al., 2022) peringkasan teks merupakan suatu kegiatan yang bertujuan untuk menyaring dan merangkum informasi utama dari sebuah sumber bacaan dengan cara sistematis, yakni dengan menghilangkan bagian-bagian yang kurang relevan. Proses ini tidak hanya melibatkan pemangkasan isi secara acak, melainkan membutuhkan analisis terhadap struktur dan makna dari teks untuk mempertahankan ide pokok yang ingin disampaikan oleh

penulis asli. Dengan kata lain, peringkasan bertujuan untuk menyajikan inti sari atau gagasan utama dalam bentuk yang lebih padat dan efektif, namun tetap menjaga konteks, makna, dan pesan yang terkandung di dalam teks secara utuh.

Di tengah perkembangan pesat era digital dan transformasi informasi, kecepatan serta kemudahan dalam mengakses informasi telah menjadi kebutuhan utama dalam kehidupan sehari-hari, baik dalam dunia pendidikan, bisnis, maupun media. Dalam situasi seperti ini, keberadaan teknologi peringkasan teks otomatis menjadi semakin krusial karena mampu membantu pengguna untuk mendapatkan inti informasi dari teks yang panjang secara lebih efisien, praktis, dan hemat waktu. Tujuan utama dari teknologi ini adalah menyajikan pokok-pokok isi dari sebuah dokumen atau artikel tanpa membutuhkan campur tangan manusia secara langsung dalam proses penyuntingan atau perangkuman teks (St Tuhpatussania & Erniwati, 2025). Dengan adanya teknologi ini, informasi inti dapat dikomunikasikan secara ringkas dan akurat, tanpa mengubah substansi atau makna asli dari dokumen sumber.

Peringkasan otomatis secara umum terbagi menjadi dua jenis utama, yaitu peringkasan ekstraktif dan peringkasan abstraktif. Peringkasan ekstraktif bekerja dengan cara memilih dan mengekstrak kalimat-kalimat penting yang secara langsung diambil dari teks asli, tanpa melakukan modifikasi terhadap struktur atau susunan kalimat tersebut. Metode ini cenderung lebih mudah diterapkan karena tidak memerlukan pemahaman semantik yang mendalam. Sebaliknya, peringkasan abstraktif bersifat lebih kompleks karena menghasilkan kalimat-kalimat baru yang dirangkai berdasarkan pemahaman dan interpretasi terhadap isi atau makna dari

dokumen secara keseluruhan. Peringkasan jenis ini berusaha menyampaikan kembali isi dokumen dengan kata-kata baru, sehingga memungkinkan interpretasi yang lebih fleksibel dan menyerupai ringkasan buatan manusia (Idhafi et al., 2023). Kedua pendekatan ini memiliki kelebihan dan kekurangan masing-masing, dan dapat digunakan sesuai dengan kebutuhan serta tujuan pengguna dalam memahami isi dokumen.

#### **2.4 *Natural Language Processing (NLP)***

*Natural Language Processing (NLP)* merupakan sebuah metode dalam bidang kecerdasan buatan yang digunakan untuk mengubah input teks atau bahasa alami menjadi informasi yang lebih terstruktur, seperti kata-kata kunci, entitas penting, atau makna utama berdasarkan respons dan kebutuhan pengguna. Bidang ini merupakan salah satu cabang dari ilmu komputer yang berkolaborasi secara erat dengan ilmu linguistik, karena melibatkan pemahaman mendalam terhadap struktur, makna, dan penggunaan bahasa manusia dalam berbagai konteks (Migunani & Kevin Aditama, 2020). NLP menjadi jembatan antara komputer dan manusia dalam berkomunikasi secara efektif melalui bahasa alami, baik lisan maupun tulisan, yang secara teknis kompleks karena bahasa manusia sarat dengan nuansa, ambiguitas, dan konteks budaya.

Tujuan utama dari *Natural Language Processing* adalah untuk mengubah bahasa manusia yang tidak terstruktur menjadi data yang dapat diproses dan dianalisis oleh sistem komputer. Dengan menggunakan berbagai pendekatan seperti machine learning, deep learning, dan algoritma linguistik, NLP memungkinkan mesin untuk memahami, menginterpretasi, menganalisis, dan menghasilkan bahasa

manusia dalam bentuk yang lebih sistematis. Sebagai contoh ilustratif, ketika mesin menerima sebuah teks panjang sebagai input, sistem akan memanfaatkan serangkaian algoritma untuk mengidentifikasi struktur kalimat, mengenali entitas penting seperti nama, tempat, atau waktu, serta mengekstraksi makna dari setiap frasa atau kalimat yang ada. Proses ini dilakukan secara bertahap dan melibatkan banyak tahapan seperti tokenisasi, stemming, parsing, dan klasifikasi semantik. Namun demikian, meskipun teknologi NLP telah mengalami kemajuan yang pesat, masih terdapat sejumlah tantangan yang signifikan dalam implementasinya.

Salah satu tantangan utama adalah kesulitan mesin dalam memahami arti atau makna kalimat secara tepat dan akurat. Hal ini disebabkan oleh kompleksitas bahasa manusia yang sangat tinggi, termasuk penggunaan sinonim, metafora, idiom, makna ganda (*ambiguitas*), serta konteks sosial dan budaya yang melekat pada bahasa. Bahasa tidak bersifat statis, melainkan dinamis dan dapat mengalami perubahan makna dari waktu ke waktu, tergantung pada perkembangan zaman dan penggunaan masyarakat. Oleh karena itu, meskipun sistem NLP dirancang untuk memahami bahasa secara otomatis, sering kali diperlukan pembaruan data dan penyesuaian model untuk memastikan bahwa interpretasi mesin tetap relevan dan sesuai dengan makna yang dimaksud oleh pengguna (Mubarok, 2021).

Dalam penerapan NLP untuk menyelesaikan suatu permasalahan, digunakan berbagai teknik, di antaranya:

a. *Sentence segmentation*

*Sentence segmentation* adalah salah satu teknik dalam NLP yang berfungsi untuk mengidentifikasi unit proses berupa kumpulan kata. Tujuan dari teknik ini adalah untuk mendeteksi atau menentukan batas akhir dari sebuah kalimat.

b. *Tokenization*

Teknik ini merupakan salah satu dasar dalam NLP, yang berfungsi dengan cara memecah kalimat menjadi token-token, di mana setiap token merepresentasikan sebuah kata.

c. *Stopword removal*

*Stopword removal* adalah salah satu teknik dalam NLP yang bertujuan untuk menghilangkan kata-kata yang tidak membawa informasi penting, sehingga dapat meningkatkan akurasi hasil pemrosesan.

## 2.5 *Deep Learning*

*Deep learning* merupakan salah satu cabang lanjutan dari *machine learning* yang menggunakan algoritma kompleks yang dirancang untuk meniru struktur dan cara kerja otak manusia. Pendekatan ini dikenal dengan istilah jaringan saraf tiruan atau *Artificial Neural Network (ANN)*, yang memungkinkan mesin untuk meniru cara manusia dalam memproses informasi melalui jaringan neuron buatan yang saling terhubung dan bekerja secara parallel (Rochmawati et al., 2021). Setiap “*neuron*” dalam jaringan ini berperan dalam mengolah dan mentransmisikan data melalui koneksi yang disebut bobot (*weights*), yang akan diperbarui secara bertahap selama proses pelatihan agar hasil prediksi menjadi lebih akurat. Dengan struktur jaringan yang mendalam (*deep*), model *deep learning* mampu menangkap pola-pola

kompleks dalam data yang tidak dapat dikenali oleh algoritma machine learning tradisional. *Artificial Neural Network* (ANN) meniru sistem kerja otak manusia dengan cara membentuk jaringan kompleks dari sejumlah besar neuron buatan yang saling berinteraksi. Jaringan ini terdiri dari beberapa lapisan: *input layer*, *hidden layer*, dan *output layer*, yang masing-masing memiliki fungsi spesifik dalam mentransformasikan input menjadi output. Proses pembelajaran dalam jaringan ini terjadi melalui teknik propagasi balik (*backpropagation*), di mana bobot antar-neuron disesuaikan berdasarkan kesalahan prediksi yang dihasilkan, agar kinerja model semakin membaik. ANN memungkinkan komputer untuk memahami hubungan *non-linear* dalam data, yang penting dalam tugas-tugas seperti klasifikasi, regresi, dan pengenalan pola.

*Deep Learning*, yang juga dikenal dengan istilah *deep structured learning*, *hierarchical learning*, atau *deep neural learning*, merupakan pendekatan pembelajaran yang menggunakan berbagai transformasi non-linear secara bertingkat (*layered*), sehingga memungkinkan model untuk membangun representasi data dalam level abstraksi yang lebih tinggi. Setiap lapisan dalam jaringan ini secara bertahap mengekstraksi fitur-fitur dari data mentah menjadi informasi yang bermakna. *Deep learning* dapat dianggap sebagai kombinasi antara pendekatan *machine learning* tradisional dan konsep kecerdasan buatan (*Artificial Intelligence*) melalui pemanfaatan jaringan saraf tiruan yang lebih dalam dan kompleks (Nugroho et al., 2020). Hal ini membuat *deep learning* sangat efektif dalam mengolah data tidak terstruktur seperti gambar, suara, teks, dan video.

Dalam praktiknya, *deep learning* memungkinkan komputer untuk secara otomatis belajar mengklasifikasikan gambar, mendeteksi objek, mengenali suara, menerjemahkan bahasa, hingga melakukan prediksi berdasarkan pola-pola dalam data. Proses pembelajaran ini membutuhkan daya komputasi yang sangat tinggi karena melibatkan banyak lapisan dan parameter yang harus dioptimalkan. Oleh karena itu, selain menggunakan CPU dan RAM untuk komputasi dasar, *deep learning* juga secara intensif memanfaatkan *Graphics Processing Unit* (GPU) guna mempercepat proses pengolahan data dalam skala besar dan meningkatkan efisiensi pelatihan model (Pratiwi et al., 2021). Kombinasi antara algoritma canggih dan infrastruktur komputasi ini menjadikan *deep learning* sebagai salah satu teknologi inti dalam pengembangan sistem cerdas modern saat ini, seperti dalam bidang pengenalan wajah, kendaraan otonom, chatbot, serta diagnosis medis berbasis citra.

## **2.6 Recurrent Neural Network (RNN)**

RNN (*Recurrent Neural Network*) merupakan salah satu jenis arsitektur dalam pemodelan *Neural Network* yang dirancang secara khusus untuk menangani data yang memiliki urutan atau bersifat sekuensial, seperti teks, sinyal suara, atau data waktu (*time-series*) (Tarkus et al., 2020). Tidak seperti jaringan saraf tradisional yang memproses input secara independen, RNN dirancang untuk memiliki hubungan antar waktu, yaitu dengan memanfaatkan keluaran dari langkah sebelumnya sebagai bagian dari input untuk langkah selanjutnya. Hal ini memungkinkan RNN memahami konteks dari urutan data dan memberikan hasil prediksi yang lebih akurat dalam konteks yang bergantung pada waktu atau urutan.

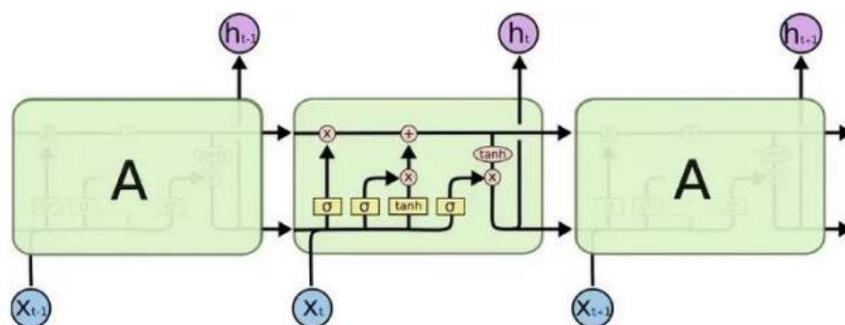
RNN (*Recurrent Neural Network*) adalah jenis arsitektur jaringan saraf tiruan yang memproses data secara berulang, di mana output dari satu elemen input akan memengaruhi proses selanjutnya. Ciri khas utama dari RNN adalah adanya *loop* atau mekanisme umpan balik dalam arsitekturnya, yang membuat jaringan ini mampu menyimpan informasi dari elemen sebelumnya dalam bentuk *hidden state*. Hal ini sangat penting untuk menangani tugas-tugas yang membutuhkan pemahaman konteks urutan, seperti dalam pemrosesan *Natural Language Processing*, terjemahan otomatis, atau pengenalan suara.

RNN dikategorikan sebagai bagian dari teknologi *Deep Learning*, karena arsitekturnya melibatkan pemrosesan data melalui berbagai lapisan non-linear secara bertingkat dan kompleks (Firmansyah et al., 2020). Kemajuan pesat dalam pengembangan dan penerapan *Recurrent Neural Network* telah membawa dampak besar dalam berbagai bidang, terutama dalam domain NLP, di mana pemahaman terhadap urutan kata sangat penting. RNN memiliki kemampuan istimewa untuk “mengingat” informasi dari langkah-langkah sebelumnya dalam sebuah sekuens input, yang disimpan dalam memori jangka pendeknya. Kemampuan ini memungkinkan RNN untuk memprediksi elemen berikutnya atau membuat keputusan berdasarkan konteks historis dari data yang telah diproses sebelumnya (Sari et al., 2022). Misalnya, dalam penerapan pada teks, RNN dapat mengantisipasi kata selanjutnya dalam kalimat berdasarkan kata-kata sebelumnya. Namun, meskipun memiliki keunggulan dalam menangani data sekuensial, RNN juga memiliki sejumlah keterbatasan yang signifikan.

Salah satu tantangan utamanya adalah masalah *vanishing gradient* (hilangnya gradien) dan *exploding gradient*, yang kerap terjadi saat proses pelatihan pada jaringan yang sangat dalam atau saat memproses urutan data yang sangat panjang. Hal ini disebabkan oleh sifat fungsi aktivasi seperti *tanh* atau *ReLU*, yang dapat menyebabkan nilai gradien menjadi sangat kecil atau sangat besar selama propagasi balik, sehingga mempersulit proses pembelajaran model (Alim, 2023). Akibatnya, RNN tradisional seringkali tidak efektif dalam mempertahankan informasi yang berasal dari input yang jauh di masa lalu. Untuk mengatasi keterbatasan tersebut, dikembangkanlah varian lain dari RNN seperti *Long Short-Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU), yang dirancang untuk mempertahankan informasi lebih lama dan mengontrol aliran informasi dalam jaringan secara lebih stabil.

## 2.7 *Long Short Term Memory* (LSTM)

*Long Short-Term Memory* (LSTM) adalah salah satu pengembangan dari arsitektur *Recurrent Neural Network* (RNN) yang pertama kali diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun 1997. LSTM dikembangkan sebagai solusi terhadap kelemahan utama RNN dalam menangani ketergantungan jangka panjang, yaitu ketika informasi penting dari waktu lampau sulit untuk dipertahankan dan digunakan kembali dalam prediksi pada urutan yang panjang. Pada RNN standar, informasi yang dibawa dari satu langkah waktu ke langkah berikutnya cenderung memudar seiring bertambahnya panjang urutan, sehingga membuat model kesulitan dalam mengenali pola-pola yang bersifat jangka Panjang (Arkadia et al., 2022).



Gambar 2.1 Sel Pada LSTM (Mubarok, 2021)

Gambar 2.1 menggambarkan ilustrasi dari isi sel yang ada pada LSTM. Gambar tersebut menunjukkan alur kerja *Long Short-Term Memory* (LSTM), sebuah jenis RNN yang dirancang untuk mengatasi masalah *long-term dependencies*. Setiap blok LSTM menerima input saat ini ( $x_t$ ) dan hidden state dari waktu sebelumnya ( $h_{t-1}$ ). Pertama, *forget gate* memutuskan informasi mana dari cell state sebelumnya ( $C_{t-1}$ ) yang harus dibuang. Selanjutnya, *input gate* menentukan informasi baru yang akan ditambahkan ke cell state, dengan bantuan fungsi aktivasi *sigmoid* dan *tanh*. *Cell state* kemudian diperbarui dengan menggabungkan informasi yang dipertahankan dan informasi baru. Terakhir, *output gate* menentukan nilai *hidden state* baru ( $h_t$ ) yang akan diteruskan ke langkah waktu berikutnya.

LSTM memiliki struktur sel yang kompleks, yang terdiri dari lapisan neuron yang digambarkan dengan persegi panjang, sedangkan elemen-elemen di dalamnya dilambangkan dengan lingkaran. Aliran data di dalam sel, antar sel, dan dari sel keluaran (*output h*) ditunjukkan dengan panah hitam. Dalam LSTM, terdapat tiga jenis gerbang, yaitu: *input gate* ( $x_t - 1$ ), *forget gate* ( $x_t$ ), dan *output gate* ( $x_t + 1$ ). *Input gate* berfungsi untuk menentukan nilai *input* yang akan diperbarui dalam

memori, dengan hasilnya diinisialisasi oleh  $h_{t-1}$ . *Forget gate* berperan untuk memutuskan informasi mana yang akan dihapus dari sel, yang dihitung menggunakan lapisan *sigmoid* ( $\sigma$ ) dan *tanh*. Hasil dari *forget gate* diinisialisasi dengan  $h_t$ . Sedangkan *output gate* bertugas untuk memutuskan hasil *output* yang akan dihasilkan, yang harus sesuai dengan hasil dari *input gate* dan *forget gate* yang telah diproses. Hasil dari *output gate* ini ditulis dengan variabel  $h_t$  (IRHAMI, 2023).

LSTM mampu mencegah masalah *vanishing gradient*, membuatnya sangat cocok untuk memproses data *time series* dan memberikan performa prediksi yang lebih akurat dibandingkan metode lain pada data historis (Moch Farryz Rizkilloh & Sri Widiyanesti, 2022).

## **2.8 Recall-Oriented Understudy for Gisting (ROUGE)**

ROUGE merupakan singkatan dari *Recall-Oriented Understudy of Gisting Evaluation*, yaitu suatu metode evaluasi yang terdiri atas sejumlah metrik yang dirancang untuk mengukur kualitas teks yang dihasilkan secara otomatis. Metode ini banyak digunakan dalam penelitian bidang pemrosesan bahasa alami, terutama dalam menilai performa sistem peringkasan otomatis, baik dalam konteks akademik maupun industri (Fadlilah et al., 2024). ROUGE menjadi penting karena mampu memberikan ukuran kuantitatif mengenai seberapa dekat hasil ringkasan sistem otomatis dengan ringkasan referensi buatan manusia.

Menurut (Somantri, 2022) ROUGE adalah metode evaluasi yang telah digunakan secara luas sebagai standar dalam pengujian hasil peringkasan teks berbahasa Indonesia. Dalam praktiknya, ROUGE menilai kualitas suatu ringkasan

dengan cara membandingkan hasil ringkasan otomatis yang dihasilkan oleh sistem dengan ringkasan manual yang dibuat oleh manusia sebagai acuan. Semakin tinggi tingkat kemiripan atau kesesuaian antara keduanya, maka semakin baik pula kualitas dari sistem peringkasan yang diuji. Karena keandalannya dalam melakukan evaluasi berbasis teks, ROUGE telah menjadi metode baku yang digunakan secara luas dalam berbagai penelitian ilmiah terkait pengembangan teknologi peringkasan.

Dalam penggunaannya, terdapat beberapa varian dari metrik ROUGE, namun dua yang paling umum digunakan adalah ROUGE-N dan ROUGE-L. ROUGE-N adalah metrik yang mengevaluasi kualitas ringkasan berdasarkan kesesuaian *n-gram*, yaitu urutan kata berukuran *n*, antara ringkasan sistem dan ringkasan referensi. ROUGE-N umumnya digunakan dalam bentuk ROUGE-1 (*unigram*) dan ROUGE-2 (*bigram*), yang masing-masing mengukur kecocokan kata dan pasangan kata (Caesarardhi et al., 2023).

Sementara itu, ROUGE-L berfokus pada pencocokan *longest common subsequence* (LCS), yaitu urutan kata terpanjang yang muncul dalam kedua ringkasan tanpa harus berurutan secara ketat namun mempertahankan urutan relatif kata tersebut (Hartawan et al., 2024). ROUGE-L memberikan nilai lebih terhadap struktur dan konteks kalimat karena mempertimbangkan kesamaan urutan kata secara menyeluruh. Dengan demikian, kombinasi dari ROUGE-N dan ROUGE-L memberikan evaluasi yang *komprehensif* dalam menilai hasil peringkasan teks otomatis, baik dari sisi kata maupun struktur kalimat. Hal ini menjadikan ROUGE sebagai salah satu alat ukur yang paling tepercaya dalam menilai performa sistem ringkasan berbasis NLP.

## 2.9 Python

Python adalah bahasa pemrograman tingkat tinggi yang dikembangkan oleh Guido Van Rossum dan pertama kali dirilis secara publik pada tahun 1991. Sejak peluncurannya, Python terus mengalami perkembangan pesat hingga menjadi salah satu bahasa pemrograman paling populer di dunia. Dalam beberapa tahun terakhir, Python telah berhasil menarik perhatian para pengembang dari berbagai bidang karena kemudahan penggunaannya, struktur sintaks yang bersih, serta komunitas yang sangat aktif. Keunggulan utama dari Python terletak pada fleksibilitas dan sifatnya yang multifungsi. Python dapat digunakan di berbagai bidang pengembangan, mulai dari aplikasi web, automasi sistem, hingga ilmu data, *Machine Learning*, dan *Deep Learning* (Alfarizi et al., 2023).

Menurut Agus Suharto (Suharto, 2023) Python merupakan bahasa pemrograman dinamis dan tingkat tinggi yang termasuk dalam kategori bahasa pemrograman *interpreter*, yaitu jenis bahasa yang mengonversi *source code* menjadi *machine code* secara langsung saat program dijalankan, tanpa perlu melalui proses kompilasi terlebih dahulu. Hal ini menjadikan Python sangat efisien untuk pengembangan cepat (*rapid development*), serta cocok digunakan oleh pemula hingga profesional karena memudahkan proses debugging dan modifikasi kode.

Python juga dikenal dengan berbagai fitur unggulan yang menjadikannya menonjol dibanding bahasa lain. Menurut (B. Kurniawan & Romzi, 2022). beberapa kelebihan Python yang signifikan meliputi:

- a. Memiliki koleksi pustaka (*library*) yang luas, dengan modul-modul siap pakai untuk berbagai keperluan, termasuk NumPy, Pandas, TensorFlow, dan Scikit-learn.
- b. Memiliki struktur sintaks yang sederhana, jelas, dan mudah dipahami, sehingga dapat meningkatkan produktivitas pengembang.
- c. Mendukung pemrograman berorientasi objek (OOP), yang memudahkan dalam pengembangan perangkat lunak berskala besar.
- d. Menyediakan pengelolaan memori otomatis, yang mengurangi beban pengembang dalam menangani alokasi dan dealokasi memori secara manual.
- e. Bersifat *modular*, memungkinkan pengembang untuk menulis dan mengelola kode dalam bentuk modul-modul terpisah yang dapat digunakan kembali di berbagai proyek.

Dengan kombinasi antara kemudahan penulisan kode, dokumentasi yang melimpah, serta kompatibilitas tinggi terhadap berbagai platform dan sistem operasi, Python telah menjadi pilihan utama bagi banyak institusi pendidikan, perusahaan teknologi, serta para peneliti dalam mengembangkan solusi teknologi masa kini.

## 2.10 Platform Youtube

YouTube merupakan platform berbagi video terbesar dan terpopuler saat ini, menarik beragam pengguna dari berbagai kelompok usia, mulai dari anak-anak hingga orang dewasa. Pengguna YouTube dapat mengunggah video, mencari konten, menonton video, berdiskusi, atau mengikuti sesi tanya jawab seputar video, serta membagikan klip video secara gratis. Dengan jutaan orang mengakses

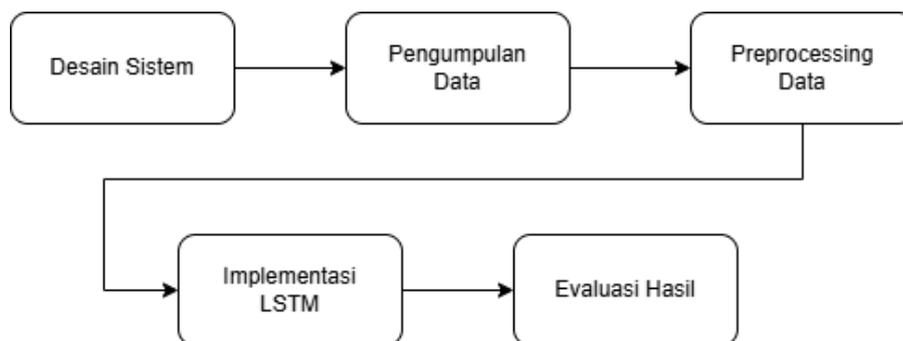
YouTube setiap hari, platform ini menjadi sarana yang sangat efektif untuk menyebarkan informasi dan teknologi komunikasi (Sitanggang et al., 2024).

YouTube diluncurkan pada tahun 2005 oleh tiga mantan karyawan PayPal, yaitu Chad Hurley, Steve Chen, dan Jawed Karim. Berkantor pusat di San Bruno, California, YouTube mengalami perkembangan pesat pada tahun 2006 dan berhasil menjadi salah satu media paling populer. Platform ini menawarkan berbagai jenis konten, mulai dari musik dan tutorial hingga vlog pribadi, yang dapat diakses dengan mudah oleh pengguna di berbagai perangkat, seperti smartphone dan smart TV (Qori Qordofa & As'ad, 2022). YouTube berperan sebagai salah satu situs dengan jumlah kunjungan tertinggi, memiliki potensi besar dalam mendukung peningkatan kualitas pendidikan di Indonesia. Dengan sekitar dua miliar pengguna aktif per bulan, YouTube mencatat lebih dari satu miliar penayangan harian pada konten-konten pembelajaran dan edukasi. Berdasarkan laporan dari DataIndonesia.id dan riset CNBC Indonesia, platform YouTube menduduki peringkat kedua sebagai situs yang paling banyak dikunjungi, baik di Indonesia maupun di tingkat global (F. D. Pratama, 2024)

## BAB III

### METODE PENELITIAN

Pada bab ini akan dijelaskan mengenai metodologi yang digunakan dalam penelitian ini untuk membuat sistem ringkasan otomatis dongeng Nusantara menggunakan metode *Long Short Term Memory* (LSTM). Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem yang dapat menghasilkan ringkasan teks dari cerita rakyat Nusantara secara otomatis. Untuk menilai kualitas ringkasan yang dihasilkan, hasil *output* sistem akan diuji oleh ahli bahasa yang berkompeten, untuk mengevaluasi kelengkapan dan kesesuaian makna dalam konteks bahasa Indonesia. Bab ini juga akan menjelaskan secara rinci tahapan-tahapan yang dilakukan dalam proses penelitian, Adapun blok diagram alur penelitian dapat dilihat pada gambar 3.1 sebagai berikut.

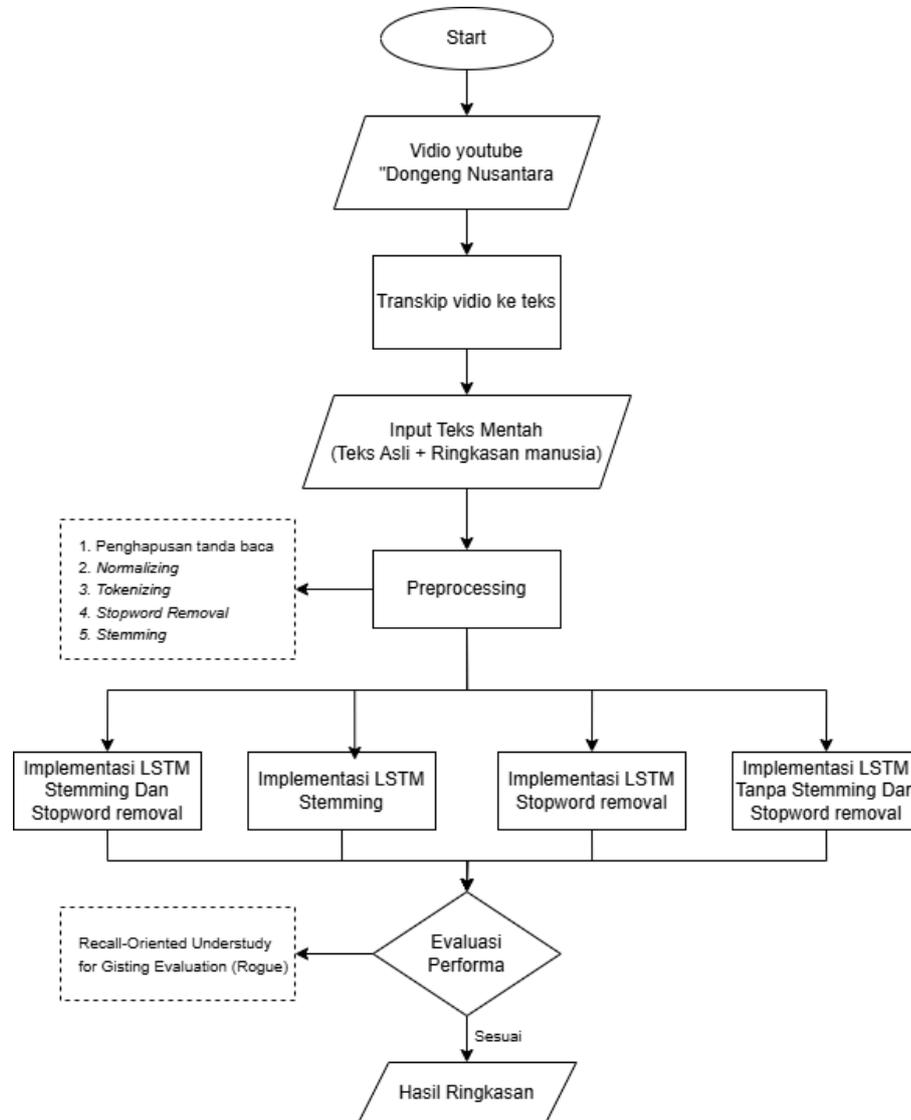


Gambar 3.1 Alur penelitian

### 3.1 Desain Sistem

Desain sistem dalam penelitian ini menggunakan metode *Long Short-Term Memory* (LSTM), yang merupakan jenis *Recurrent Neural Network* (RNN) yang mampu menangani data sekuensial dengan hubungan jangka panjang. LSTM

dipilih karena kemampuannya dalam mempertahankan informasi penting dari data sebelumnya. Alur lengkap sistem ini dapat dilihat pada Gambar 3.2.



Gambar 3.2 Desain perancangan sistem

Desain perancangan sistem pada penelitian ini menggambarkan alur proses peringkasan otomatis dongeng Nusantera menggunakan metode *Long Short-Term Memory* (LSTM). Sistem dimulai dari tahap pengambilan data yang bersumber dari video YouTube dengan topik khusus dongeng Nusantera. Tidak semua video diambil, melainkan hanya video yang sesuai dengan kriteria, yaitu yang benar-

benar memuat cerita dongeng Nusantara. Video yang telah dipilih kemudian ditranskripsi menjadi teks agar dapat diolah lebih lanjut oleh sistem.

Setelah proses transkripsi selesai, data yang digunakan terdiri dari dua bagian, yaitu teks asli hasil transkripsi dan ringkasan manual yang disusun oleh ahli bahasa. Ringkasan manual ini berfungsi sebagai pembanding dalam proses evaluasi model. Selanjutnya, kedua data tersebut menjadi input teks mentah yang akan diproses pada tahap berikutnya.

Proses berikutnya adalah *preprocessing*, yaitu tahap penting untuk mempersiapkan data sebelum dimasukkan ke dalam model LSTM. Pada tahap ini, diterapkan beberapa proses dasar *preprocessing* yang selalu digunakan pada seluruh data, yaitu penghapusan tanda baca, *case folding*, dan *tokenizing*. Dua tahapan *preprocessing* tambahan, yaitu *stopword removal* dan *stemming*, digunakan sebagai bagian dari skema pengujian dan tidak selalu diterapkan pada semua data. *Stopword removal* bertujuan untuk menghapus kata-kata yang tidak memiliki makna penting seperti “dan”, “yang”, “di”, sedangkan *stemming* berfungsi untuk mengembalikan kata ke bentuk dasarnya. Kedua teknik ini digunakan dalam beberapa skenario untuk menguji pengaruhnya terhadap kinerja model.

Setelah melalui *preprocessing*, data kemudian dimasukkan ke dalam empat skema implementasi LSTM yang berbeda, yaitu, Implementasi LSTM *Stemming* *stopword removal*, LSTM *Stemming*, LSTM *Stopword removal*, dan LSTM tanpa *Stemming* dan *Stopword Removal*. Keempat skema ini bertujuan untuk mengetahui

bagaimana masing-masing metode preprocessing mempengaruhi hasil peringkasan otomatis yang dihasilkan oleh LSTM.

Hasil ringkasan yang dihasilkan oleh sistem dievaluasi menggunakan metode *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE), yang membandingkan hasil ringkasan otomatis dengan ringkasan manual yang telah dibuat oleh ahli bahasa. Jika hasil evaluasi sudah sesuai dengan kriteria yang ditetapkan, maka sistem akan menghasilkan ringkasan akhir yang dianggap sebagai hasil terbaik dari proses peringkasan teks dongeng Nusantara secara otomatis.

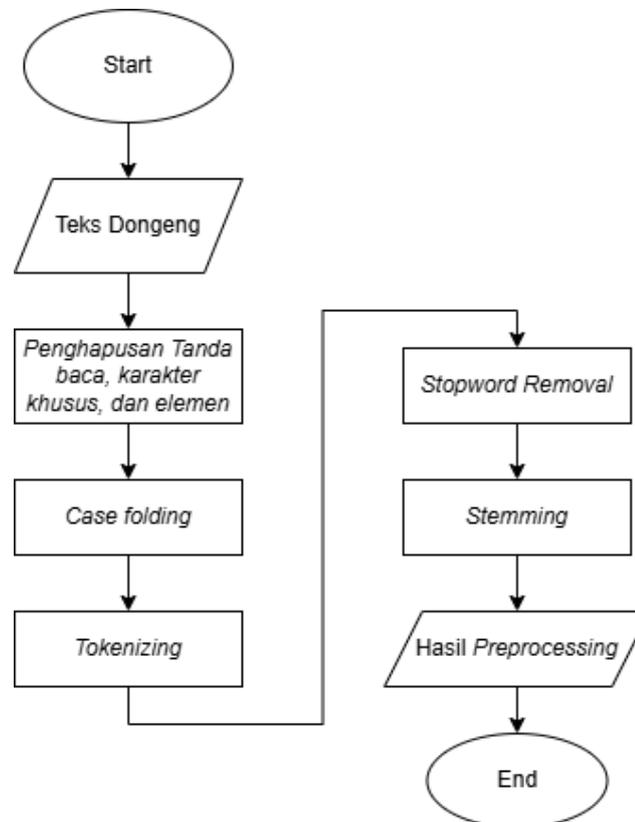
### **3.2 Pengumpulan Data**

Pengumpulan data dalam penelitian ini dilakukan dengan mengakses dongeng Nusantara yang tersedia di platform YouTube, yang merupakan salah satu sumber utama dalam menyediakan variasi cerita rakyat dari berbagai daerah di Indonesia. Data yang diambil berupa teks di website atau video dongeng dengan kata kunci pencarian “Dongeng Nusantara”, yang kemudian ditranskrip untuk memperoleh teks cerita secara keseluruhan. Proses ekstraksi teks dilakukan dengan menggunakan *tools* di website yaitu <https://anthiago.com>

### **3.3 Preprocessing Data**

*Preprocessing* data merupakan langkah esensial dalam mengubah teks mentah menjadi data terstruktur yang dapat digunakan untuk membangun sistem peringkasan teks otomatis. Dalam *Natural Language Processing* (NLP), tahapan ini biasanya mencakup proses seperti penghapusan kata-kata umum yang tidak memiliki makna signifikan (*stopword*), pemecahan teks menjadi kata atau token

(*tokenizing*), Dan mengubah kata-kata ke bentuk dasarnya (*Stemming*) Berikut ini adalah tahapan pra-pemrosesan data yang diimplementasikan pada penelitian ini.



Gambar 3.3 Flowchart Preprocessing

### 3.3.1 Penghapusan tanda baca

Tahap awal dalam pemrosesan teks adalah menghilangkan semua tanda baca, seperti titik, koma, tanda seru, dan tanda tanya. Langkah ini dilakukan untuk memastikan bahwa analisis model lebih terfokus pada kata-kata bermakna tanpa gangguan dari simbol yang tidak memberikan kontribusi langsung terhadap pemahaman konteks. Selain itu, tanda baca dapat menyebabkan variasi dalam tokenisasi, misalnya antara kata "*model?*" dan "*model*", yang dapat dianggap sebagai dua token berbeda. Dengan menghapus tanda baca, proses pemrosesan teks menjadi lebih konsisten dan membantu model *Long Short-Term Memory* (LSTM)

dalam memahami pola bahasa dengan lebih baik, sehingga dapat menghasilkan ringkasan yang lebih akurat.

### **3.3.2 Case folding**

Proses selanjutnya dalam *preprocessing* teks adalah *Case folding*, yang bertujuan untuk mengubah seluruh teks menjadi huruf kecil. Hal ini penting karena model *Natural Language Processing* (NLP) tidak perlu membedakan antara kata yang ditulis dengan huruf kapital dan yang tidak. Misalnya, kata "Kancil" dan "kancil" seharusnya dianggap sama, karena keduanya merujuk pada entitas yang sama. Tanpa langkah ini, model mungkin akan memperlakukan dua kata tersebut sebagai entitas berbeda hanya karena perbedaan kapitalisasi, yang dapat mengarah pada hasil yang kurang optimal dalam analisis teks. Pengaturan huruf kecil memastikan bahwa kata-kata yang memiliki bentuk kapitalisasi yang berbeda akan diperlakukan secara konsisten. Langkah ini sangat penting untuk mengurangi kerumitan yang tidak perlu dalam pemrosesan teks.

### **3.3.3 Tokenizing**

Setelah teks diubah menjadi huruf kecil, langkah berikutnya adalah *tokenizing*, yaitu pemisahan teks menjadi unit-unit yang lebih kecil yang disebut token. Token bisa berupa kata-kata atau kalimat. Biasanya, dalam banyak kasus, tokenisasi dilakukan berdasarkan kata. Tokenisasi membantu untuk memecah teks yang panjang menjadi bagian-bagian yang lebih mudah dikelola dan diproses oleh model. Selain itu, tokenisasi juga memungkinkan model untuk lebih mudah mengenali kata-kata individual, sehingga analisis dapat dilakukan dengan lebih

efisien. Dalam konteks peringkasan teks, tokenisasi juga membantu untuk mengidentifikasi kata-kata penting yang perlu diberi perhatian lebih dalam proses pembuatan ringkasan.

### **3.3.4 *Stopword Removal***

Selanjutnya, kita melakukan *stopword removal*, yaitu menghapus kata-kata yang dianggap tidak memberikan informasi penting dalam konteks pemahaman teks. *Stopwords* adalah kata-kata umum seperti "dan", "di", "dari", "adalah", dll, yang sering kali muncul dalam teks tetapi tidak membantu dalam analisis makna atau konten utama. Dalam analisis teks, *stopwords* dianggap sebagai kata yang tidak memiliki bobot informasi yang signifikan untuk memahami makna atau konteks keseluruhan dari kalimat tersebut. Menghapus *stopwords* membantu model untuk lebih fokus pada kata-kata yang memiliki kontribusi lebih besar terhadap pemahaman teks, yang akhirnya meningkatkan kualitas pemrosesan dan pengambilan informasi penting. Dalam penelitian ini, proses *Stopword removal* dilakukan menggunakan library *Sastrawi*, yang merupakan pustaka pemrosesan bahasa alami berbahasa Indonesia.

### **3.3.5 *Stemming***

*Stemming* adalah langkah penting yang bertujuan untuk mengubah kata-kata ke bentuk dasarnya. Misalnya, kata-kata seperti "berjalan" akan dikembalikan ke bentuk dasar "jalan". Proses ini membantu mengurangi kompleksitas model dengan menghilangkan variasi bentuk kata yang merujuk pada konsep yang sama. Dalam teks yang lebih panjang, kata-kata dalam bentuk yang berbeda sering kali memiliki

arti yang sama, sehingga meningkatkan efisiensi pemrosesan dan ketepatan hasil. Dalam penelitian ini, proses *stemming* juga dilakukan menggunakan library *Sastrawi*, yang merupakan pustaka pemrosesan bahasa alami berbahasa Indonesia.

Langkah-langkah *preprocessing* yang diterapkan dalam penelitian ini dapat dilihat secara rinci pada Tabel 3.2, yang mencakup tahapan penghapusan tanda baca, *Case folding*, *tokenizing*, *stopword removal*, dan *stemming*.

Tabel 3.1 contoh preprocessing

No	Langkah	Hasil
1.	Paragraf Asli	"Di sebuah hutan yang lebat, hiduplah seekor kancil yang cerdas. Setiap hari, kancil mencari makan sambil berhati-hati terhadap hewan buas yang berkeliaran. Suatu hari, ia melihat seekor harimau yang sedang terjebak dalam perangkap. Harimau itu memohon agar dilepaskan. Dengan rasa iba, kancil membebaskannya. Namun, setelah bebas, harimau itu justru ingin memakan kancil."
2.	Penghapusan Tanda baca	Di sebuah hutan yang lebat hiduplah seekor kancil yang cerdas Setiap hari kancil mencari makan sambil berhati-hati terhadap hewan buas yang berkeliaran Suatu hari ia melihat seekor harimau yang sedang terjebak dalam perangkap Harimau itu memohon agar dilepaskan Dengan rasa iba kancil membebaskannya Namun setelah bebas harimau itu justru ingin memakan kancil
3.	<i>Case folding</i>	di sebuah hutan yang lebat hiduplah seekor kancil yang cerdas setiap hari kancil mencari makan sambil berhati-hati terhadap hewan buas yang berkeliaran suatu hari ia melihat seekor harimau yang sedang terjebak dalam perangkap harimau itu memohon agar dilepaskan dengan rasa iba kancil membebaskannya namun setelah bebas harimau itu justru ingin memakan kancil
4.	<i>Tokenizing</i>	['di', 'sebuah', 'hutan', 'yang', 'lebat', 'hiduplah', 'seekor', 'kancil', 'yang', 'cerdik', 'setiap', 'hari', 'kancil', 'mencari', 'makan', 'sambil', 'berhati-hati', 'terhadap', 'hewan', 'buas', 'yang', 'berkeliaran', 'suatu', 'hari', 'ia', 'melihat', 'seekor', 'harimau', 'yang', 'sedang', 'terjebak', 'dalam', 'perangkap', 'harimau', 'itu', 'memohon', 'agar', 'dilepaskan', 'dengan', 'rasa', 'iba', 'kancil', 'membebaskannya', 'namun', 'setelah', 'bebas', 'harimau', 'itu', 'justru', 'ingin', 'memakan', 'kancil']
5.	<i>Stopword Removal</i>	['hutan', 'lebat', 'hiduplah', 'seekor', 'kancil', 'cerdik', 'setiap', 'hari', 'kancil', 'mencari', 'makan', 'berhati-hati', 'hewan', 'buas', 'berkeliaran', 'suatu', 'hari', 'melihat', 'seekor', 'harimau', 'sedang', 'terjebak', 'perangkap', 'harimau', 'memohon', 'dilepaskan', 'rasa', 'iba', 'kancil', 'membebaskannya', 'bebas', 'harimau', 'justru', 'ingin', 'memakan', 'kancil']
6.	<i>Stemming</i>	['hutan', 'lebat', 'hidup', 'ekor', 'kancil', 'cerdik', 'tiap', 'hari', 'kancil', 'cari', 'makan', 'hati-hati', 'hewan', 'buas', 'keliar', 'suatu', 'hari', 'lihat', 'ekor', 'harimau', 'sedang', 'jebak', 'perangkap', 'harimau', 'mohon', 'lepas', 'rasa', 'iba', 'kancil', 'bebas', 'bebas', 'harimau', 'justru', 'ingin', 'makan', 'kancil']

### 3.4 Implementasi metode *Long Short-Term Memory* (LSTM)

Metode LSTM dipilih karena keberhasilannya dalam peringkasan abstraktif. LSTM (*Long Short-Term Memory*) merupakan pengembangan dari model RNN yang dirancang untuk menangani data sekuensial dengan mempertahankan informasi dari langkah sebelumnya. RNN sendiri adalah arsitektur jaringan saraf tiruan yang memproses data secara berulang untuk mengelola *input* yang bersifat sekuensial (Afandi et al., 2022). LSTM memiliki keunggulan dibandingkan RNN konvensional, yaitu kemampuannya dalam mengatur informasi dengan mekanisme penyimpanan dan penghapusan sebelum digunakan kembali sebagai *input*. Selain itu, LSTM juga mampu mempertahankan *error* selama proses *backpropagation*, sehingga dapat mencegah peningkatan kesalahan (Ivanedra & Mustikasari, 2019). RNN memiliki sebuah *hidden state* yang berfungsi sebagai memori dari jaringan syaraf tiruan (JST) dan dihitung berdasarkan *hidden state* sebelumnya. LSTM, sebagai salah satu jenis RNN, terdiri dari empat lapisan JST yang saling berinteraksi dalam setiap *hidden state*-nya. LSTM terdiri dari empat jenis *gate* yang saling berinteraksi, serta sebuah *cell state* yang berperan dalam meneruskan informasi dari satu *hidden state* ke *hidden state* berikutnya dengan sedikit atau bahkan tanpa perubahan (Irawati Setiawan et al., 2020).

Tahapan pemrosesan pada LSTM melewati beberapa jaringan yang terdiri dari *cell state*, *input gate*, *output gate*, dan *forget gate*. *Cell State* merupakan jalur bagian atas yang berfungsi sebagai tempat penyimpanan informasi yang dibawa dari satu waktu ke waktu berikutnya. *Input gate* berfungsi mengatur informasi baru yang akan dimasukkan dan memperbarui *cell state*. *Forget gate* berperan dalam

mengatur informasi mana yang akan dipertahankan dalam *cell state*. Sementara itu, *output gate* berfungsi untuk mengatur informasi dalam *cell* yang akan digunakan sebagai *output* dari unit LSTM (Subowo et al., 2022). Langkah awal dalam proses kerja jaringan LSTM adalah menghapus informasi dari *cell state* menggunakan *forget gate*  $f$ , sebagaimana ditunjukkan dalam persamaan 3.2 Pertama, vektor  $h_{t-1}$  disalin dan dilewatkan ke *forget gate*, dengan  $b$  sebagai nilai bias.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3.2)$$

**Keterangan :**

- $f_t$  : nilai *forget gate*
- $\sigma$  : fungsi *sigmoid*
- $W_{fx}$  : bobot *forget gate* yang didapat selama pelatihan
- $h_{t-1}$  : *hidden state* dari waktu sebelumnya
- $x_t$  : input pada waktu  $t$
- $b_f$  : *forget gate bias*

Selanjutnya, jaringan LSTM memperbarui *cell state* dengan informasi baru melalui *input gate*  $i$ , sebagaimana ditunjukkan dalam persamaan 3.3. *Gate* ini berfungsi menentukan informasi yang akan diperbarui dalam *cell state*. Secara bersamaan, vektor yang sama juga melewati lapisan *tanh* untuk menyesuaikan rentang nilainya menjadi bilangan real antara -1 hingga 1, sehingga menghasilkan *candidate cell state*  $\tilde{C}_t$ , sebagaimana ditunjukkan dalam persamaan 3.4.

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (3.3)$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (3.4)$$

**Keterangan :**

- $i_t$  : nilai *input gate*
- $\sigma$  : fungsi *sigmoid*
- $\tilde{C}_t$  : nilai *memory candidate*
- $W_c$  : bobot memori kandidat

Vektor yang melewati *input gate* dan lapisan *tanh* kemudian mengalami operasi perkalian *pointwise*, sehingga hanya *output* yang relevan yang diperbarui dalam *cell state*. Dengan demikian, setelah melalui ketiga *gate* LSTM dan lapisan *tanh*, vektor-vektor tersebut menghasilkan nilai *cell state* yang sesuai seperti dalam persamaan 3.5.

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (3.5)$$

**Keterangan :**

$C_t$  : nilai *memory candidate*  
 $f_t$  : nilai *forget gate*  
 $C_{t-1}$  : *cell state* dari waktu sebelumnya  
 $i_t$  : nilai *input gate*  
 $\tilde{C}_t$  : nilai *memory candidate*

Langkah terakhir adalah menentukan *output gate* sebagaimana ditunjukkan dalam persamaan 3.6. Vektor  $h_{t-1}$  disalin dan diteruskan ke *output gate*. Secara bersamaan,  $C_t$  disalin dan dilewatkan ke lapisan *tanh*. Setelah melewati lapisan *tanh*,  $C_t$  dikombinasikan dengan  $o_t$  melalui operasi perkalian *pointwise*, menghasilkan nilai *output* dari *cell* sesuai dengan persamaan 3.7.

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (3.6)$$

$$h_t = o_t \times \tanh(C_t) \quad (3.7)$$

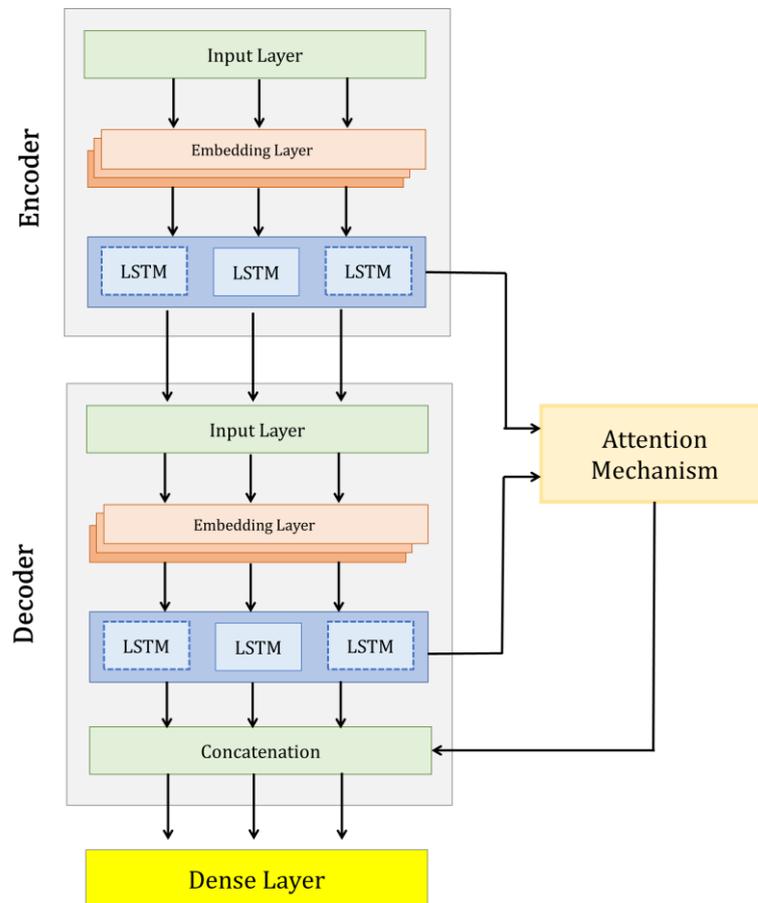
**Keterangan :**

$o_t$  : nilai *output gate*  
 $i_t$  : nilai *input gate*  
 $\sigma$  : fungsi *sigmoid*  
 $h_t$  : nilai *hidden state*  
 $C_t$  : *cell state*  
 $x_t$  : input pada waktu  $t$   
 $b$  : bias

Dalam arsitektur LSTM, terdapat tiga jenis *gate* utama, yaitu *input gate*  $i(t)$ , *forget gate*  $f(t)$ , dan *output gate*  $o(t)$ . Ketiga jenis *gate* ini berfungsi secara sinergis

untuk mengatur aliran informasi dari data baru yang masuk  $x(t)$  ke dalam unit sel, yang secara langsung memengaruhi hasil keluaran  $h(t)$  dari unit LSTM tersebut. *Forget gate* berperan dalam menentukan informasi mana dari keadaan sel sebelumnya yang perlu dihapus atau dipertahankan. Mekanismenya bekerja dengan menghasilkan nilai antara 0 dan 1; semakin mendekati 0, maka informasi tersebut akan dilupakan, dan semakin mendekati 1, maka informasi akan dipertahankan, sehingga LSTM dapat menyaring informasi yang tidak relevan. Sementara itu, *input gate* bertugas mengatur seberapa banyak informasi baru dari *input* saat ini yang akan disimpan dalam keadaan sel. Proses ini melibatkan evaluasi dari *input* saat ini dan keadaan tersembunyi sebelumnya untuk menentukan nilai-nilai kandidat yang akan ditambahkan ke dalam memori sel. Terakhir, *output gate* mengontrol seberapa besar bagian dari keadaan sel saat ini yang akan digunakan sebagai *output*. *Gate* ini memfilter informasi dari memori internal dan menentukan bagian mana yang akan diteruskan ke tahap berikutnya dalam proses pemrosesan data (Kamal Wisyaldin et al., 2020).

Selanjutnya untuk arsitektur yang dipakai pada penelitian ini menggunakan 2 lapisan (layer) yakni *encoder*, dan *decoder*. Berikut adalah gambaran arsitektur yang digunakan pada penelitian ini:



Gambar 3.4 Arsitektur LSTM

Gambar 3.5 menunjukkan arsitektur model jaringan saraf tiruan berbasis *Long Short-Term Memory* (LSTM) yang dilengkapi dengan mekanisme *Attention*. Terdapat dua blok utama yang menggambarkan proses *encoder* dan *decoder* yang masing-masing memiliki struktur serupa: dimulai dari *Input Layer* yang menerima *input* berupa urutan kata atau token. *Input* tersebut kemudian diproses melalui *Embedding Layer*, yang mengubah setiap token menjadi representasi *vektor* berdimensi tetap untuk menangkap informasi semantik.

Setelah *embedding*, *vektor-vektor* tersebut dimasukkan ke dalam beberapa lapisan LSTM. Lapisan LSTM digunakan untuk mempelajari dependensi urutan

dalam data teks, yaitu konteks urutan kata dalam kalimat. LSTM ini disusun secara sejajar dan memproses *input* secara berurutan. *Output* dari LSTM pertama (di bagian atas) digunakan sebagai representasi awal dari teks yang telah diproses.

Selanjutnya, dalam proses *decoder* (bagian bawah), *input* kembali melewati lapisan embedding dan LSTM dengan struktur yang identik. Namun, perbedaannya adalah di bagian ini terdapat *Concatenation Layer*, yang menggabungkan *output* dari seluruh LSTM sebelum dikirim ke *Dense Layer* untuk menghasilkan prediksi akhir.

Yang paling menonjol dari arsitektur ini adalah adanya *Attention Mechanism* yang berada di antara *encoder* dan *decoder*. *Attention* berfungsi untuk menentukan bagian mana dari input yang paling relevan untuk diperhatikan dalam proses decoding, sehingga memperkuat konteks informasi yang digunakan oleh model. Dengan kata lain, *attention* membantu model untuk memfokuskan perhatian pada kata-kata penting saat menghasilkan output. Akhirnya, output dari mekanisme *attention* dan hasil *concatenation* akan masuk ke *Dense Layer* yang menghasilkan keluaran akhir model, misalnya dalam bentuk ringkasan, label, atau prediksi lainnya tergantung dari tugasnya.

### **3.4.1 Splitting Data**

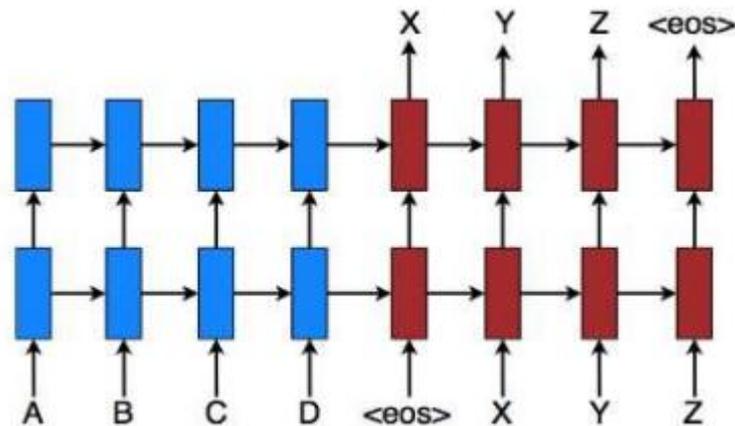
Tahapan selanjutnya adalah membagi dataset menjadi dua bagian utama, yaitu data *training* dan data *testing*. Proporsi untuk data *training* 80% dan data *testing* 20% (Gholamy et al., 2018). Proses ini penting untuk memastikan bahwa model dapat dilatih menggunakan data yang cukup untuk mempelajari pola yang ada, sementara data *testing* digunakan untuk menguji akurasi dan generalisasi

model setelah proses pelatihan. Setelah pembagian data selesai, langkah selanjutnya adalah melakukan proses *training* dan *testing* pada model LSTM. Data yang telah diproses sebelumnya melalui tahapan *preprocessing* teks akan digunakan sebagai *input* bagi model. *Array* hasil *preprocessing* teks tersebut kemudian diberikan kepada input layer model. Pada tahap berikutnya, setiap kata dalam *input* akan diproses melalui *embedding layer*, yang berfungsi untuk mentransformasikan kata-kata tersebut menjadi representasi vektor numerik. Vektor-vektor hasil transformasi ini disimpan dalam vector conversion, yang kemudian diproses lebih lanjut oleh LSTM layer. LSTM layer ini bertujuan untuk menangkap ketergantungan temporal dalam data teks dan mempelajari hubungan antara kata-kata dalam urutan tersebut. Melalui tahap ini, model akan dapat memahami konteks kata dan membangun representasi yang lebih kompleks untuk meningkatkan akurasi dalam memprediksi output yang diinginkan.

### 3.4.2 *Sequence to Sequence*

Model *sequence-to-sequence* (Seq2Seq) pertama kali diperkenalkan oleh Sutskever (Phua et al., 2022). Model ini menjadi solusi atas tantangan dalam memproses data berurutan dengan panjang dan susunan yang bervariasi. Dalam proses peringkasan otomatis, terdapat mekanisme pemetaan dari urutan kata pada *input* ke urutan kata pada output. Namun, dalam praktiknya, struktur urutan kata pada input sering kali tidak identik dengan urutan di *output*. Selain itu, panjang *input* dan *output* bisa berbeda, begitu pula dengan susunan katanya. Untuk menangani perbedaan ini, dikembangkanlah sebuah model yang disebut *Sequence-to-Sequence* (Seq2Seq). Model ini memiliki arsitektur utama berupa dua

komponen, yaitu *Encoder* dan *Decoder*, yang berperan dalam proses penerjemahan urutan *input* menjadi urutan *output* yang sesuai.



Gambar 3.5 Model Seq2Seq

Gambar 3.6 menggambarkan proses kerja *Encoder* dan *Decoder*, di mana keduanya berfungsi untuk mengolah bahasa sumber sebagai *input* dan menghasilkan bahasa target sebagai *output*. *Encoder* (ditunjukkan oleh blok-blok berwarna biru) bertugas menerima dan memproses urutan *input*, yaitu token A, B, C, dan D, secara berurutan. Setiap token dimasukkan ke dalam unit LSTM, yang menyimpan informasi jangka panjang melalui *cell state* dan *hidden state*. Informasi dari seluruh urutan input akan dirangkum dalam bentuk *context vector* yaitu *hidden state* terakhir dari *encoder* yang akan menjadi *input* awal bagi *decoder*. *Decoder* (ditunjukkan oleh blok-blok merah) memulai prosesnya dengan menerima *context vector* dari *encoder* serta simbol khusus <eos> sebagai input pertama. Selanjutnya, *decoder* akan menghasilkan satu token *output* pada setiap langkah, seperti X, Y, dan Z, di mana hasil prediksi pada satu waktu akan digunakan sebagai *input* pada

waktu berikutnya. Proses ini berlanjut hingga token khusus <eos> dihasilkan, yang menandakan akhir dari urutan *output*.

### 3.4.3 Lapisan *Encoder*

*Encoder* adalah bagian dari model yang tugasnya mengubah urutan data masukan menjadi representasi yang lebih padat, sering disebut *feature vector*. Representasi ringkas ini, yang juga dikenal sebagai *encoder hidden state* atau *latent representation*, secara efektif menangkap esensi informasi penting dari *input*, mempersiapkannya untuk diproses oleh jaringan *decoder* (Fauziyah et al., 2022). *Encoder* bertugas membaca setiap kata dalam kalimat masukan dari bahasa sumber. Kemudian, informasi dari kalimat tersebut diringkas dan diubah menjadi sebuah vektor konteks (*context vector*) yang tersimpan dalam *hidden state* LSTM (*Long Short-Term Memory*) (Diarsyah & Setiawan, 2024). *Encoder* bekerja dengan memproses teks yang masuk, lalu merangkum informasi pentingnya ke dalam *hidden state* dan *cell state vector*. Dari ringkasan ini, *encoder* kemudian menghasilkan *context vector*, yang berfungsi sebagai masukan bagi *decoder* (Alfhi Saputra, 2021).

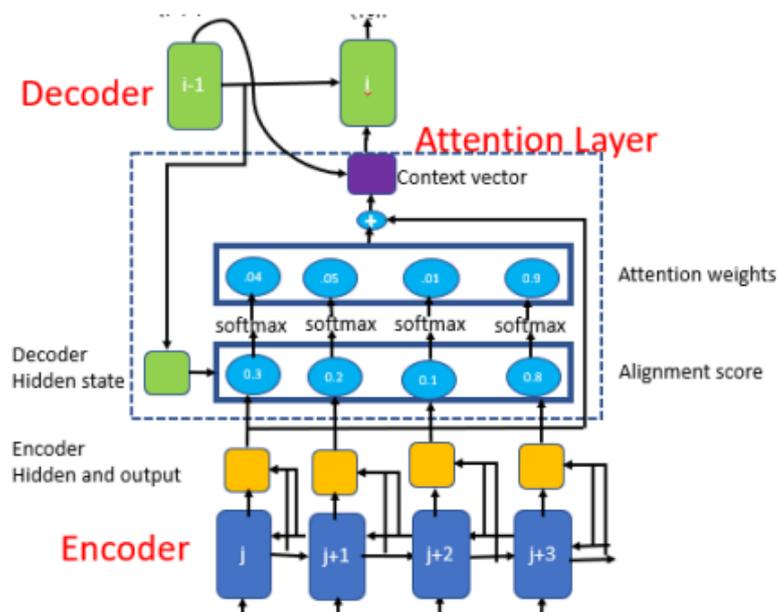
### 3.4.4 Lapisan *Decoder*

*Decoder* berfungsi untuk menerjemahkan informasi yang telah diproses oleh *Encoder*. Pada tahap ini, *context vector* yang dihasilkan sebelumnya akan digabungkan dengan *output Decoder* dari langkah sebelumnya. Gabungan ini kemudian dimasukkan ke dalam sel-sel *RNN Decoder* untuk menghasilkan *hidden state* yang baru. Proses ini terus berulang hingga *Decoder* menghasilkan token

penanda akhir *sekuens*, yaitu token “<EOS>” (Diarsyah & Setiawan, 2024). Arsitektur ini cukup fleksibel, bisa menggunakan berbagai jenis *neural network* seperti CNN, RNN, atau LSTM, tergantung pada permasalahan yang ingin diselesaikan (Fauziyah et al., 2022).

### 3.4.5 Attention Mechanism

*Attention Mechanism* adalah strategi dalam model *Sequence-to-Sequence* (Seq2Seq) yang bertujuan untuk mempelajari representasi vektor tunggal dari kalimat atau untuk mengarahkan fokus pada bagian-bagian penting saat memprediksi *output*. Mekanisme ini terdiri dari tiga lapisan utama: *alignment layer*, *attention weights*, dan *context vector*. *Context vector* ini dihasilkan dengan menggabungkan *hidden state* dari *Encoder* dan *hidden state* dari *Decoder*. Melalui *attention weights*, mekanisme ini memungkinkan model untuk fokus pada vektor input spesifik dari urutan masukan (Fauziyah et al., 2022).



Gambar 3.6 Alur *Attention Mechanism*

### 3.4.6 Dense Layer

*Dense layer*, atau disebut juga *fully connected layer*, umumnya ditempatkan di bagian akhir model. Fungsinya adalah melakukan peringkasan dengan memanfaatkan fitur-fitur yang telah didapatkan. Ciri khas *dense layer* adalah setiap *neuron* di dalamnya terhubung sepenuhnya dengan setiap *neuron* pada lapisan sebelumnya (Putra & Saputra, 2023). Fungsi aktivasi yang dipakai dalam penelitian ini adalah *softmax* dengan fungsi optimasinya *Adam*. Fungsi *Softmax* mengubah keluaran model menjadi nilai probabilitas untuk setiap kelas yang ada.

## 3.5 Evaluasi

Evaluasi sistem ringkasan otomatis dalam penelitian ini dilakukan dengan menggunakan ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*). ROUGE adalah metode evaluasi otomatis yang umum digunakan untuk mengukur kualitas ringkasan yang dihasilkan oleh sistem, dengan cara membandingkannya dengan ringkasan acuan yang disusun oleh manusia (Hidjah & Susilowati, 2025). Semakin tinggi skor ROUGE yang diperoleh, maka semakin baik pula kualitas ringkasan tersebut, karena menunjukkan tingkat kesesuaian yang tinggi dengan ringkasan referensi buatan manusia (Aisyah et al., 2019).

Metode ROUGE yang digunakan dalam penelitian ini mencakup ROUGE-N, ROUGE-N digunakan untuk mengukur kemiripan antara ringkasan yang dihasilkan oleh sistem dengan ringkasan acuan dari manusia, dengan menghitung jumlah *n-gram* (urutan kata sebanyak *n*) yang sama di antara keduanya. Nilai ROUGE berada dalam rentang 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan tingkat kecocokan yang lebih besar antara kedua ringkasan tersebut (Factor, 2024).

Sementara itu, ROUGE-L menilai kesamaan berdasarkan urutan kata terpanjang yang muncul secara berurutan pada hasil ringkasan dan referensi, menggunakan pendekatan *Longest Common Subsequence* (LCS) untuk menentukan tingkat kesesuaian struktur dan isi antar keduanya (Algoritma & Porter, 2025).

$$ROUGE(N) = \frac{\text{Jumlah } n\text{-gram yang sama}}{\text{Jumlah } n\text{-gram pada ringkasan manusia}} \quad (3.8)$$

$$ROUGE(L) = \frac{LCS(\text{ringkasan manusia}, \text{ringkasan mesin})}{\text{panjang ringkasan manusia}} \quad (3.9)$$

### 3.6 Skenario Pengujian

Dalam Dalam penelitian ini, penulis merancang empat skenario berbeda untuk menguji pengaruh teknik *preprocessing* terhadap hasil sistem. Tujuan dari penerapan skenario ini adalah untuk mengevaluasi peran masing-masing tahapan preprocessing, yaitu *stemming* dan *stopwords removal*, dalam meningkatkan performa sistem yang dikembangkan.

Pada skenario pertama, penulis menerapkan kedua proses *preprocessing* secara bersamaan, yaitu *stemming* dan *stopwords removal*. Proses *stemming* digunakan untuk mengubah kata ke bentuk dasarnya (contoh: "berlari" menjadi "lari"), sedangkan penghapusan *stopwords* dilakukan untuk membuang kata-kata umum yang tidak memiliki makna penting dalam analisis (seperti "dan", "yang", "di", dan sebagainya).

Pada skenario kedua, hanya proses *stemming* yang diterapkan tanpa penghapusan *stopwords*. Dengan demikian, kata-kata dasar dipertahankan, namun

semua kata *stopwords removal* tetap ada dalam data. Skenario ini bertujuan untuk melihat sejauh mana *stemming* saja dapat membantu sistem memahami makna teks.

Sebaliknya, pada skenario ketiga, penulis hanya menerapkan *stopwords removal* tanpa melakukan *stemming*. Ini berarti kata-kata umum yang tidak penting telah dibuang, namun bentuk kata tetap dipertahankan seperti aslinya.

Terakhir, pada skenario keempat, tidak dilakukan proses *stemming* maupun *stopwords removal*. Data teks diproses apa adanya tanpa reduksi atau penyaringan kata, yang digunakan sebagai skenario kontrol untuk dibandingkan dengan skenario-skenario lainnya.

Seluruh proses *stemming* dan *stopwords removal* pada penelitian ini dilakukan menggunakan library *Sastrawi* pada bahasa pemrograman Python, yang merupakan library populer dalam pemrosesan bahasa alami untuk teks berbahasa Indonesia.

Tujuan dari penerapan tiga skenario ini adalah untuk menganalisis sejauh mana pengaruh *stemming* dan penghapusan *stopwords* terhadap sistem peringkasan teks yang dikembangkan. Jika hasil perbandingan menunjukkan adanya dampak yang signifikan atau justru tidak berpengaruh, temuan ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya dalam menentukan apakah penggunaan *stemming* dan *stopwords removal* perlu diterapkan atau tidak (Alfhi Saputra, 2021).

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Persiapan Data

Pada tahap awal penelitian ini, proses pengambilan data dilakukan dengan memilih video yang diambil dari platform YouTube. Tidak semua video digunakan dalam penelitian ini, melainkan hanya video yang secara khusus memuat cerita dongeng Nusantara. Video-video yang telah dipilih kemudian ditranskripsi menjadi teks agar dapat diproses lebih lanjut. Setelah diperoleh data hasil transkripsi, teks tersebut diringkas oleh ahli bahasa untuk menghasilkan ringkasan manual yang nantinya akan digunakan sebagai dataset pembandingan. Dengan demikian, dataset yang digunakan dalam penelitian ini terdiri dari dua bagian, yaitu teks asli hasil transkripsi dari video YouTube dan hasil ringkasan manual yang disusun oleh ahli bahasa. Berikut Tabel 4.1 contoh data yang akan menjadi input sistem.

Tabel 4.1 Tabel yang belum diringkas dan sudah diringkas

<b>Teks asli Dongeng Nusantara</b>	<b>Ringkasan Manusia</b>
Dahulu kala, disebuah lembah subur di Sumatra Utara, hiduplah seorang pemuda bernama Toba. sehari hari dia bekerja diladang sebagai seorang petani. Dia sebenarnya termasuk orang yang giat, namun sayang hasil panen tahun ini benar-benar sangat sedikit. Tak seperti musim sebelumnya. Sehingga dia harus mencari penghasilan lain. Salah satunya adalah memancing ikan disungai dekat rumahnya untuk sekedar dimakan sendiri ataupun di jual ke pasar. Suatu hari ketika pulang dari ladang, Toba berencana memancing ikan untuk dijadikan lauk makan malam. Cukup lama menunggu umpannya disambar ikan, membuatnya hampir putus asa. tiba-tiba pancingnya ditarik oleh sesuatu. Toba kegirangan karena penantiannya tidak sia-sia Pancingnya ditarik sekuat tenaga, dan nampak ikan yang amat cantik bersisik keemasan serta berukuran besar. Tanpa Buang waktu lagi, Toba segera memasukan ikan itu kedalam wadah dan segera membawanya pulang untuk lekas memasaknya. Sesampainya dirumah dia sedikit kecewa karena kayu bakarnya telah habis, Sehingga terpaksa harus kembali ke ladang untuk mencari kayu bakar. Ikan itupun dia letakkan didalam tempayan. Selama mencari kayu bakar dalam benak Toba adalah menu makan	Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri. Mereka dikaruniai anak bernama Samosir yang tumbuh menjadi pemalas. Suatu hari, Toba marah karena Samosir memakan bekal sebelum

<p>malam yang begitu istimewa. Dia bisa memakan ikan besar yang ditangkapnya tadi. Sudah cukup mencari kayu bakar, dia segera menyalakan api. Namun ketika akan mengambil ikan, dia keheranan karena ikannya telah lenyap. Yang ada hanya beberapa keping emas. Entah dari mana datangnya. Toba kebingungan dengan keanehan itu dan bermaksud masuk ke kamar untuk menenangkan diri. Namun saat membuka pintu kamar Toba sangat terkejut. Nampak seorang wanita berambut panjang berdiri sambil menyisir rambutnya. dengan gemeteran Toba bertanya kepada wanita itu. “ Siapakah engkau ? ” Wanita itu memalingkan wajahnya ke arah Toba sambil tersenyum manis. Melihat hal itu Toba terpesona dan tertegun sejenak. Ternyata orang yang didepannya memiliki paras yang luar biasa cantik. Belum pernah dia melihat wanita secantik ini sebelumnya. “ Siapakah gerangan bidadari ini ? Apa mungkin dia pencuri ? Tapi tidak mungkin ada pencuri secantik ini ! “ Gadis itu pun mendekatinya serta menyadarkan Toba dari lamunannya...</p>	<p>memberikannya dan tanpa sadar melanggar janjinya dengan menyebut Samosir "anak ikan". Sang istri sedih dan menyuruh Samosir lari ke bukit. Bencana besar pun datang, sang istri kembali menjadi ikan dan Toba hanyut terbawa air bah ...</p>
--	---

Dataset yang telah dikumpulkan kemudian melalui tahap *preprocessing* untuk mempersiapkan data sebelum diolah oleh model. Tidak semua tahapan *preprocessing* digunakan dalam seluruh proses. *Preprocessing* dasar yang selalu diterapkan pada semua data adalah penghapusan tanda baca, *case folding*, dan *tokenizing*. Sementara itu, tahapan *stopword removal* dan *stemming* tidak diterapkan secara keseluruhan, melainkan digunakan sebagai bagian dari skema pengujian. Dalam skema pengujian, dilakukan empat variasi: penggunaan *stopword removal* dan *stemming*, hanya *stemming*, hanya *stopword removal*, serta tanpa menggunakan keduanya. Dengan pembagian skema ini, diharapkan dapat diketahui pengaruh dari masing-masing teknik *preprocessing* terhadap kinerja model dalam melakukan peringkasan otomatis teks dongeng Nusantara.

## 4.2 Tahap Uji Coba

Sebelum memperoleh hasil akurasi dari pengujian sistem peringkasan teks, terdapat beberapa tahapan yang harus dilakukan untuk mencapai hasil yang optimal. Tahapan-tahapan tersebut meliputi:

1. Melakukan *preprocessing* pada dataset Dongeng Nusantara, yang terdiri dari dua kolom utama, yaitu kolom “A” yang memuat teks asli dan kolom “B” yang berisi ringkasan manual yang disusun oleh ahli.
2. Membagi dataset menjadi data latih dan data validasi dengan rasio 80% data *training* dan 20% data *testing*, serta menerapkan empat skenario pengujian *preprocessing*, yaitu: (a) LSTM dengan *preprocessing stemming* dan *stopword removal*, (b) LSTM dengan hanya *preprocessing stemming*, (c) LSTM dengan hanya *preprocessing stopwords removal*, dan (d) LSTM tanpa *preprocessing stemming* maupun *stopword removal*.
3. Mengukur performa metode LSTM dengan menggunakan metrik evaluasi ROUGE-1, ROUGE-2, dan ROUGE-L. Evaluasi dilakukan dengan membandingkan hasil ringkasan yang dihasilkan oleh sistem dengan ringkasan manual yang telah tersedia dalam dataset.
4. Membandingkan hasil pengujian dari keempat skenario *preprocessing* untuk mengidentifikasi skenario yang memberikan performa terbaik, serta menentukan tahapan *preprocessing* yang paling efektif dalam meningkatkan kualitas ringkasan teks.

### 4.3 Konfigurasi Parameter LSTM

Parameter-parameter berikut merupakan komponen utama yang telah ditetapkan dalam pembentukan model dan diaplikasikan secara tetap dalam seluruh skenario, guna memastikan konsistensi dalam proses pelatihan dan pengujian:

a. *hidden\_units* = 150

*Hidden unit* adalah komponen dalam lapisan tersembunyi jaringan saraf yang berperan dalam memproses dan merepresentasikan pola dari data input. Jumlah *hidden unit* yang digunakan akan memengaruhi kinerja dan hasil model. Jika jumlah *hidden unit* terlalu kecil, proses pelatihan akan berlangsung lebih cepat, namun kapasitas model untuk mempelajari pola kompleks dalam data menjadi terbatas. Hal ini dapat berdampak negatif terhadap performa model, khususnya saat diuji pada data validasi karena informasi yang dipelajari menjadi kurang optimal (Sianturi et al., 2023).

b. *Layer* = 2

*Layer* adalah komponen penting dalam arsitektur LSTM yang menentukan jumlah tingkat pemrosesan data dalam jaringan. Pengaturan jumlah *layer* secara tepat diperlukan untuk mengontrol kompleksitas model, sehingga dapat membantu mencegah terjadinya *underfitting* maupun *overfitting* (Abdul et al., 2025).

c. *Timesteps* = 100

*Timestep* adalah jumlah langkah waktu yang digunakan untuk merepresentasikan urutan data dalam model pembelajaran sekuensial. Semakin tinggi nilai *timestep*, semakin panjang rentang informasi sebelumnya yang dapat dipertimbangkan oleh model dalam membuat prediksi. Namun, penggunaan *timestep* yang terlalu besar dapat memperlambat proses pelatihan karena meningkatnya kompleksitas perhitungan (Marietta & ,Rukun Santoso, 2025).

d. *Embedding\_dim* = 100

*Embedding* adalah metode representasi kata dalam bentuk vektor numerik di ruang berdimensi tertentu. Salah satu keunggulan teknik ini adalah kemampuannya dalam menangkap hubungan semantik antar kata, di mana kata-kata yang memiliki makna serupa akan direpresentasikan oleh vektor-vektor yang berdekatan. Sebagai contoh, kata “pergi” dan “berangkat” akan memiliki representasi vektor yang hampir sama karena keduanya memiliki makna yang mirip (M. Kurniawan et al., 2021).

e. *Learning\_rate* = 0,001

*Learning rate* adalah parameter yang mengatur seberapa besar perubahan bobot yang dilakukan oleh model pada setiap iterasi selama proses pelatihan. Nilai *learning rate* yang terlalu tinggi dapat menyebabkan model melewati solusi optimal karena pembaruan yang terlalu agresif, sedangkan nilai yang terlalu rendah dapat memperlambat proses konvergensi karena pembaruan dilakukan secara terlalu lambat (Erzitha et al., 2025). *Learning rate* yang terlalu rendah dapat memperlambat proses konvergensi karena perubahan bobot pada setiap iterasi menjadi sangat kecil. Sebaliknya, jika nilai *learning rate* terlalu tinggi, pembaruan bobot menjadi terlalu besar, sehingga model kesulitan untuk mencapai titik konvergensi secara stabil (Marietta & Rukun Santoso, 2025).

f. *Optimizer* = adam

*Optimizer* adalah algoritma yang digunakan untuk mengatur dan memperbarui bobot dalam model selama proses pelatihan. Adam merupakan

algoritma optimasi berbasis *stochastic gradient descent* yang bekerja dengan menyimpan rata-rata dari momentum gradien serta kuadrat dari momentum tersebut untuk setiap parameter dalam model. *Optimizer* ini menggabungkan keunggulan dari metode RMSprop dan *Stochastic Gradient Descent* (SGD), sehingga mampu memberikan proses pembaruan bobot yang lebih adaptif dan efisien. (Arkadia et al., 2022).

- g. *Dropout*: `dropout_rate = 0,5`, `dropout_type = 'variational'`

*Dropout* adalah salah satu teknik regularisasi dalam pelatihan jaringan saraf yang digunakan untuk mencegah terjadinya *overfitting*. Teknik ini bekerja dengan cara menonaktifkan secara acak sejumlah neuron pada lapisan jaringan selama proses pelatihan. Dengan mengurangi ketergantungan antar neuron, dropout membantu model untuk belajar representasi yang lebih umum, sehingga dapat meningkatkan kemampuan generalisasi terhadap data baru yang belum pernah dilihat sebelumnya (Erzitha et al., 2025).

- h. `batch_size = 1`

*Batch size* adalah jumlah sampel data yang diproses secara bersamaan dalam satu iterasi selama satu *epoch* pelatihan. Ukuran *batch* ini berperan penting dalam memengaruhi kecepatan pelatihan, proses konvergensi, serta tingkat akurasi model. Jika *batch size* terlalu kecil, proses pelatihan bisa menjadi lambat dan bobot model cenderung berubah secara tidak stabil. Sebaliknya, *batch size* yang terlalu besar dapat menyebabkan *overfitting* karena model terlalu menyesuaikan diri dengan data pelatihan, sehingga kesulitannya meningkat dalam menggeneralisasi data baru. *Batch* yang besar juga

membutuhkan kapasitas memori yang lebih besar untuk menyimpan aktivasi dan gradien selama pelatihan (Marietta & Rukun Santoso, 2025).

i. *Epoch* = 100

*Epoch* adalah satu siklus penuh di mana seluruh data pelatihan digunakan untuk melatih model satu kali. Selama setiap *epoch*, model melakukan pembaruan bobot berdasarkan perhitungan gradien dari kesalahan prediksi terhadap data pelatihan. Jumlah *epoch* yang terlalu sedikit dapat menyebabkan proses pelatihan tidak maksimal, sehingga model belum mampu belajar pola secara optimal dari data (Marietta & Rukun Santoso, 2025).

Berdasarkan parameter-parameter di atas penelitian ini menggunakan konfigurasi yang diambil dari hasil penelitian sebelumnya antara lain: *hidden unit* = 150, *layers* = 2, *Embedding\_dim* = 100, *optimizer* = 'adam', *dropout\_type* = 'variational', *dropout\_rate* = 0.5, *batch\_size* = 1 (Reimers & Gurevych, 2016), *learning\_rate* = 0,001 (Zhu et al., 2020), dan *timestep* = 100 (Flows, 2020) *Epoch* = 100 (Moghar, 2020).

#### 4.4 Skema Pengujian

Skema pengujian dalam penelitian ini dirancang untuk mengevaluasi pengaruh teknik *preprocessing* terhadap performa model LSTM dalam menghasilkan ringkasan teks. Empat skenario pengujian diterapkan dengan kombinasi berbeda antara proses *stemming* dan *stopword removal*. Setiap skenario menggunakan dataset yang sama, dengan komposisi 80% cerita sebagai data latih dan 20% cerita sebagai data uji. Perbedaan pada setiap skenario terletak pada jenis

*preprocessing* yang diterapkan sebelum teks dimasukkan ke dalam model. Adapun rincian skenario pengujian dapat dilihat pada Tabel 4.2 berikut:

Tabel 4.2 Tabel Skenario pengujian

No	Skenario
1.	Menggunakan <i>stemming</i> dan <i>stopword removal</i>
2.	<i>Stemming</i>
3.	<i>Stopword removal</i>
4.	Tanpa <i>Stemming</i> dan <i>Stopword removal</i>

#### 4.5 Hasil Pengujian

Pada subbab ini disajikan hasil pengujian dari keempat skenario pra-pemrosesan teks yang telah diterapkan pada model *Long Short-Term Memory* (LSTM). Setiap skenario menghasilkan ringkasan otomatis berdasarkan data uji yang telah diproses dengan konfigurasi yang berbeda, yaitu kombinasi antara penggunaan *stemming* dan *stopword removal*. Hasil pengujian mencakup contoh data setelah *preprocessing*, hasil ringkasan yang dihasilkan oleh model, nilai evaluasi menggunakan metrik ROUGE, serta grafik *Training & Validation Loss* serta *Training & Validation Accuracy*, Loss merupakan ukuran penalti yang diberikan kepada model ketika melakukan prediksi yang tidak sesuai dengan nilai sebenarnya, sedangkan *accuracy* menunjukkan seberapa sering model memberikan hasil prediksi yang benar. Nilai loss merepresentasikan tingkat kesalahan prediksi yang dilakukan oleh model terhadap suatu data. Apabila model mampu memprediksi dengan sempurna, maka nilai *loss* akan mendekati nol. Namun, jika prediksi yang dihasilkan model jauh dari nilai yang seharusnya, maka nilai *loss* akan semakin tinggi. Kondisi sebaliknya berlaku untuk *accuracy*, di mana semakin

banyak prediksi yang benar, maka nilai *accuracy* akan meningkat (Antoko et al., 2021). Penyajian hasil dilakukan secara terpisah untuk setiap skenario guna memudahkan analisis dan perbandingan performa model pada masing-masing konfigurasi.

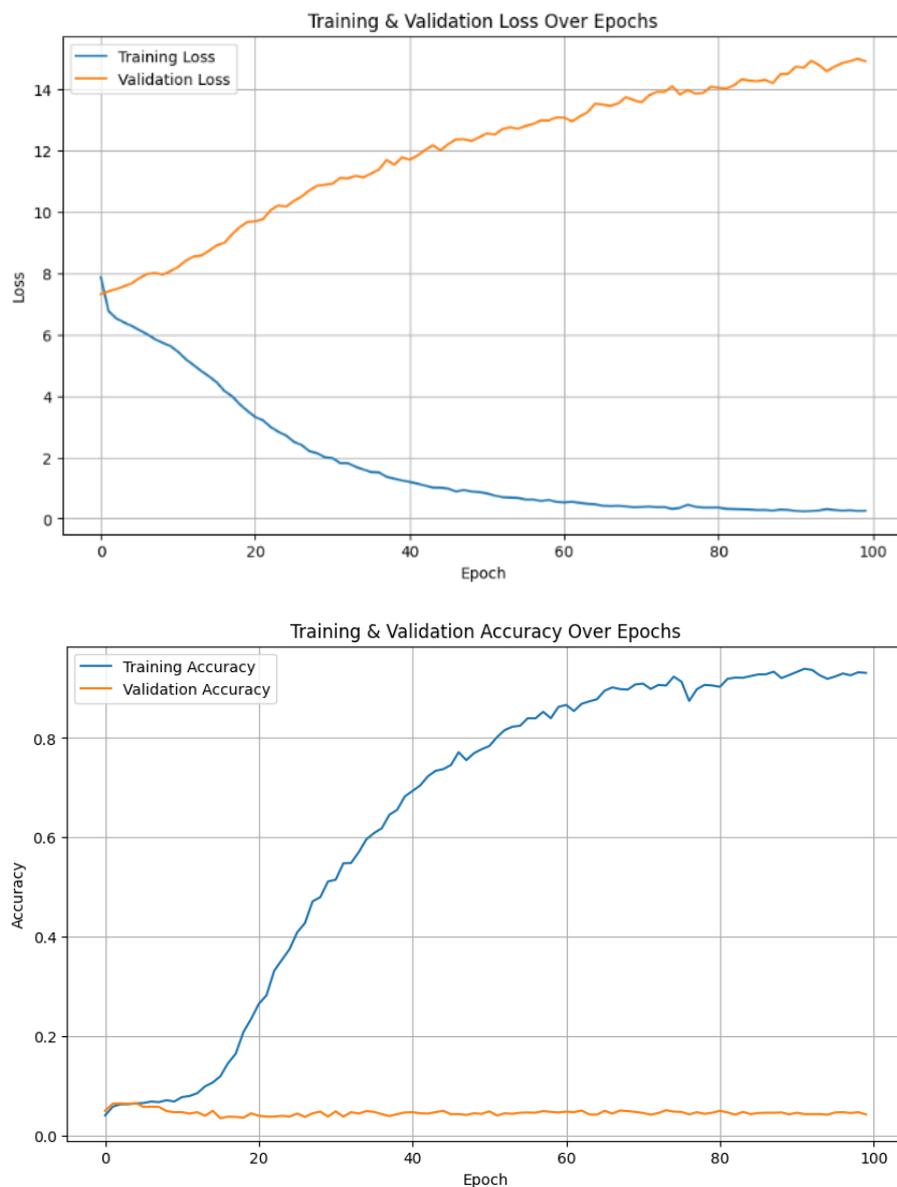
#### 4.5.1 Hasil Skenario 1: Menggunakan *Stemming* dan *Stopword Removal*

Pada skenario pertama, data uji diproses menggunakan teknik *stemming* dan penghapusan *stopword* sebelum dimasukkan ke dalam model LSTM. Tujuan dari langkah ini adalah untuk menyederhanakan struktur kalimat dengan menghilangkan kata-kata yang dianggap tidak memiliki kontribusi semantik tinggi serta mengubah kata ke bentuk dasarnya. Tabel berikut menyajikan beberapa contoh hasil pra-pemrosesan pada skenario ini:

Tabel 4.3 Contoh Hasil *Preprocessing Stemming* dan *stopword removal*

No	Teks Asli	Hasil Preprocessing ( <i>Stemming</i> + <i>Stopword Removal</i> )
1.	Dahulu kala, disebuah lembah subur di Sumatra Utara, hiduplah seorang pemuda bernama Toba. sehari hari dia bekerja diladang sebagai seorang petani. Dia sebenarnya termasuk orang yang giat, namun sayang hasil panen tahun ini benar-benar sangat sedikit. Tak seperti musim sebelumnya. Sehingga dia harus mencari penghasilan lain. Salah satunya adalah memancing ikan disungai dekat rumahnya untuk sekedar dimakan sendiri ataupun di jual ke pasar...	kala buah lembah subur sumatra utara hidup orang pemuda nama toba hari hari kerja ladang bagai orang tani benar masuk orang giat sayang hasil panen tahun benarbenar sangat sedikit tak musim cari hasil salah satu pancing ikan sungai dekat rumah dar makan sendiri di jual pasar suatu hari pulang ladang toba rencana pancing ikan jadi lauk makan malam cukup lama tunggu umpan sambar ikan buat hampir putus asa tibatiba pancing tarik sesuatu ...
2.	kisah dari telur ajaib Alkisah hidup sebuah keluarga kecil sang ayah bernama Keo Dia adalah seorang duda yang mempunyai Putri Tunggal bernama Marika suatu hari Paman Geo jatuh hati kepada seorang janda bernama macan dia juga memiliki seorang putri bernama cantik mereka kemudian menikah dan hidup bersama belakangan Paman Keo harus pergi meninggalkan keluarganya untuk urusan pekerjaan sebelum pergi ia memanggil Marika dan mengatakan sesuatu padanya putriku sementara ...	kisah telur ajaib alkisah hidup buah keluarga kecil sang ayah nama keo adalah orang duda punya putri tunggal nama marika suatu hari paman geo jatuh hati orang janda nama macan milik orang putri nama cantik kemudian meni hidup sama belakang paman keo pergi tinggal keluarga urus kerja pergi panggil marika kata putri tinggal sama ibu macan aku cari uang keluarga cinta ibu kau cinta iya ayah paman...

Untuk menilai performa model LSTM pada skenario pertama, di mana data telah melalui proses *stemming* dan penghapusan *stopword*, dilakukan evaluasi terhadap nilai *loss* dan *accuracy*. Tujuan dari evaluasi ini adalah untuk melihat kestabilan proses pelatihan dan kemampuan model dalam melakukan generalisasi terhadap data uji. Berikut grafik 4.1 hasil pengujian skenario 1:



Gambar 4.1 Grafik Hasil Skenario 1

Evaluasi performa model LSTM pada skenario pertama dilakukan dengan melihat grafik *Training & Validation Loss* serta *Training & Validation Accuracy* terhadap jumlah epoch. Skenario ini menggunakan data yang telah diproses melalui tahapan *stemming* dan penghapusan *stopword*, yang bertujuan untuk menyederhanakan bentuk kata dan menghilangkan kata-kata yang tidak memberikan kontribusi makna signifikan.

Grafik pertama memperlihatkan nilai *loss* selama proses pelatihan dan validasi. Terlihat bahwa *training loss* mengalami penurunan yang stabil, menandakan bahwa model mampu belajar dari data pelatihan dengan baik. Namun, *validation loss* menunjukkan tren meningkat seiring bertambahnya *epoch*, yang mengindikasikan terjadinya *overfitting*. Hal ini berarti model terlalu menyesuaikan diri dengan data latih dan kehilangan kemampuan generalisasi terhadap data uji.

Grafik kedua menunjukkan akurasi model terhadap data pelatihan dan validasi. Akurasi pada data pelatihan meningkat secara signifikan hingga mencapai lebih dari 90%, sementara akurasi pada data validasi justru tetap sangat rendah, berada di kisaran 5%. Pola ini kembali menegaskan bahwa model mengalami *overfitting*, di mana performa terhadap data pelatihan sangat baik namun tidak diikuti dengan peningkatan akurasi terhadap data validasi.

Setelah proses pelatihan dan pengujian model LSTM pada skenario pertama, dilakukan evaluasi terhadap kualitas hasil ringkasan otomatis menggunakan metrik ROUGE. Metrik ROUGE digunakan untuk membandingkan ringkasan yang dihasilkan oleh model dengan ringkasan acuan yang dibuat oleh ahli bahasa. Pada evaluasi ini digunakan tiga jenis metrik utama, yaitu ROUGE-1, ROUGE-2, dan

ROUGE-L. Hasil rata-rata dari evaluasi tersebut ditampilkan dalam tabel 4.3 berikut:

Tabel 4.4 Hasil Evaluasi Matrik Skenario 1

Metrik Evaluasi	Nilai Rata - rata
ROGUE-1	0,5069
ROGUE-2	0,2019
ROGUE-L	0,3505

Tabel di atas menunjukkan bahwa model mencapai skor ROUGE-1 tertinggi, yang berarti banyak kata dalam ringkasan model cocok dengan ringkasan manual. Namun, skor ROUGE-2 yang lebih rendah mengindikasikan bahwa kesesuaian dalam pasangan kata (*bigram*) masih terbatas. Sementara itu, nilai ROUGE-L memberikan gambaran bahwa model cukup baik dalam menghasilkan urutan kata yang mirip dengan ringkasan acuan.

Contoh hasil ringkasan otomatis yang dihasilkan oleh model LSTM pada salah satu cerita dalam skenario pertama. Tabel 4.4 berikut menampilkan perbandingan antara ringkasan manual oleh ahli bahasa, dan ringkasan yang dihasilkan oleh mesin:

Tabel 4.5 Tabel Perbandingan Ringkasan manual dan mesin

No	Jenis	Teks
1.	Ringkasan Manual	di sebuah lembah subur di sumatra utara hiduplah toba seorang petani yang hasil panennya menurun drastis untuk mencukupi kebutuhan ia memancing di sungai dan berhasil menangkap ikan bersisik emas namun ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk mereka menikah dengan syarat toba tak boleh mengungkit asal usul sang istri mereka dikaruniai anak bernama samosir yang tumbuh menjadi pemalas suatu hari toba marah karena samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut samosir anak ikan sang istri sedih dan menyuruh samosir lari ke bukit bencana besar pun datang sang istri kembali menjadi ikan dan toba hanyut terbawa air bah tempat itu akhirnya menjadi danau toba dan bukit tempat samosir berlindung kini dikenal sebagai pulau samosirdi sebuah lembah subur di sumatra utara hiduplah toba seorang petani yang hasil panennya menurun drastis untuk mencukupi kebutuhan ia memancing di sungai dan berhasil menangkap ikan bersisik emas namun ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk mereka menikah dengan syarat toba tak boleh mengungkit asal usul sang istri mereka dikaruniai anak bernama

No	Jenis	Teks
		samosir yang tumbuh menjadi pemalas suatu hari toba marah karena samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut samosir anak ikan sang istri sedih dan menyuruh samosir lari ke bukit bencana besar pun datang sang istri kembali menjadi ikan dan toba hanyut terbawa air bah tempat itu akhirnya menjadi danau toba dan bukit tempat samosir berlindung kini dikenal sebagai pulau samosir
2.	Ringkasan Mesin	<p>kala buah lembah subur sumatra utara hidup orang pemuda nama toba hari hari kerja ladang bagai orang tani benar masuk orang giat sayang hasil panen tahun benarbenar sangat sedikit tak musim cari hasil salah satu pancing ikan sungai dekat rumah dar makan sendiri di jual pasar suatu hari pulang ladang toba rencana pancing ikan jadi lauk makan malam cukup lama tunggu umpan sambar ikan buat hampir putus asa tibatiba pancing tarik sesuatu toba girang nanti siasia pancing tarik kuat tenaga nampak ikan amat cantik sisik emas ukur besar buang waktu toba segera pasu ikan wadah segera bawa pulang lekas masak sampa rumah sedikit kecewa kayu bakar habis sehingga paksa harus kembali ladang cari kayu bakar ikan letak tempayan lama cari kayu bakar benak toba menu makan malam begitu istimewa makan ikan besar ditangkapnya tadi cukup cari kayu bakar segera nyala api ambil ikan heran ikan lenyap beberapa keping emas entah mana datang toba bingung aneh maksud masuk kamar tenang diri buka pintu kamar toba sangat kejut nampak orang wanita rambut panjang diri sisir rambut gemetar toba tanya wanita siapa engkau wanita maling wajah arah toba senyum manis lihat toba pesona tegun jenak nyata orang depan milik paras luar biasa cantik pernah lihat wanita cantik siapa gerangan bidadari apa mungkin curi mungkin ada curi cantik gadis dekat sadar toba lamun jelma ikan kamu tangkap tadi hah jelma ikan tunggu dulu keping emas adalah sisasisa sisik ku bagaimana bisa wanita cantik bisa jadi ekor ikan dulu aku adalah orang putri kutuk dewata jadi ikan berkat engkau kutu kini hilang izin aku tinggal sini bagai ganti aku buat makan kamu lapar toba bingung semua turut wanita ajak dapur cepat makan saji mereka toba sangat takjub lihat wanita sampaisampai makan berandaiandai hati wanita jadi istri indah harihari ku lalu rupa toba tahu wanita aku mau jadi istri aku punya satu syarat engkau langgar dengar kata wanita toba kejut sekaligus senang kata apa syarat meni jangan pernah ungkit memberitahu siapa asal mula engkau langgar akan ada buruk akan jadi baik adalah mudah demi istri cantik kamu pandai diri apa akan ku laku akhir meni toba makin giat kerja penuh butuh hidup keluarga kecil iring jalan waktu sang istri hamil toba sabar tunggu lahir anak singkat cerita sang istri lahir orang putra kemudian beri nama samosir samosir tumbuh jadi anak malas nakal akibat selalu manja orang tua utama sang ibu seharihari bermalasmalasan main keluyur toba terkadang rasa prihatin laku sang anak hingga suatu sang ibu minta samosir antar bekal makan siang ayah mula samosir tolak sang ibu terus bujuk hingga akhir samosir mau berangkat walaupun muka masam tengah jalan samosir putus main lebih yakin sang ayah mau tunggu beberapa saat sementara ladang toba sabar tunggu makan tak kunjung datang sangat lapar lelah main samosir rasa lapar pikir panjang makan bekal sisa sedikit sang ayah lama tunggu jauh tampak samosir datang bawa bekal lihat toba rasa senang toba senyum sambut datang sang anak berfikir sang anak ubah mau bantu dua orang tua nak ayo makan sama ehmmmm ayah aku kenyang ayah terimakasih ayah ayo kita makan sama ayah aku kenyang ayah makan kau jalan cukup jauh rumah ladang kau lelah lapar ajak samosir makan bekal bawah pohon nan rindang ketika toba buka bekal alangkah kejut haaahh apa samosir kau makan jalan iya aku makan sedikit waktu jalan tadi aku sisa sedikit ayah tidak masih cukup kan dengar jawab samosir rasa tidak ada rasa salah toba naik pitam laku sungguh luar batas kau malas selalu susah orang tua kemari kau dasar anak ikan marah toba tak bendung tanpa sadar sudah langgar janji di ucap sambil menang samosir lari pulang tengah jalan masih renung diri bukan</p>

No	Jenis	Teks
		anak ikan tetapi anak ibunda lama kandung sampa rumah samosir adu ibu terisakisak dia cerita sang ayah sebut bagai anak ikan malas jadi beban orang tua ibu aku bukan anak ikan bu aku bukan anak ikan dengar cerita ibu samosir kejut rasa sedih dia tidak sangka suami langgar janji ungkit asal usul katakata ayah tadi jangan kamu masuk hati kamu tetap anak ibu ibu pesan samosir jadi anak mandiri sabar selalu buat baik kepada orang sekarang cepat kau lari atas bukit belum bencana besar datang jaga diri baikbaik nak samosir gegas turut ibu ketika langit ubah jadi gelap guntur kilat sambar berkalikali jadi hujan badai begitu dahsyat badai buat air sungai meluapluap hingga buat banjir bandang lalu istri toba lompat air bah kembali wujud ikan toba sedang kerja ladang kejut air yang begitu besar tibatiba datang dia tidak sempat lari diri toba pun hanyut dan tenggelam dalam air bah kau langgar janji maaf aku air telah genang tempat dan hanya sisa puncak bukit tengah tempat itu samosir sembunyi kini bukit itu sebut pulau samosir dan air bah yang telah jadi danau kenal bagai danau toba

#### 4.5.2 Hasil Skenario 2: Menggunakan *Stemming*

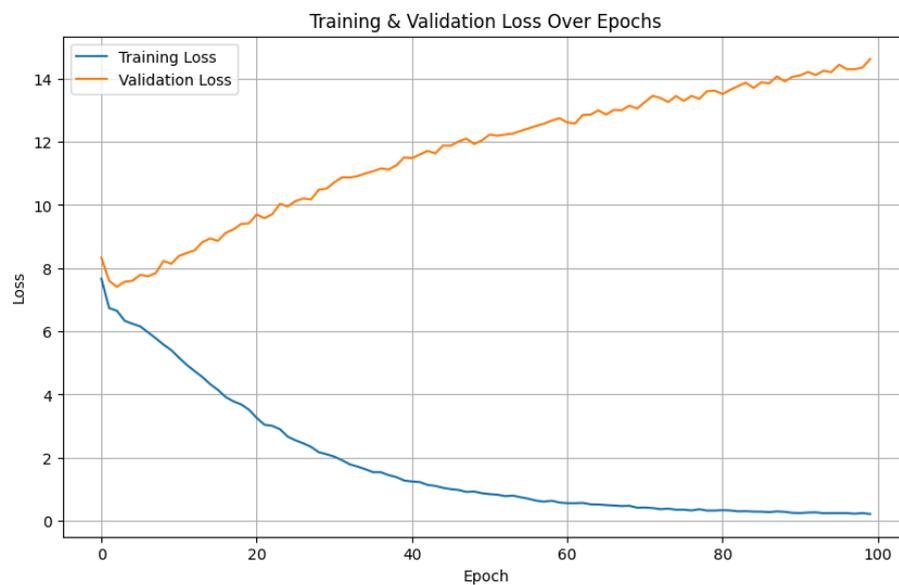
Pada skenario kedua, proses pra-pemrosesan teks dilakukan hanya dengan menerapkan *stemming*, tanpa menghilangkan *stopword*. Tujuan dari skenario ini adalah untuk mengevaluasi sejauh mana pengaruh *stemming* saja terhadap kinerja model LSTM dalam merangkum teks. Dengan tetap mempertahankan *stopword*, model masih menerima kata-kata umum yang sering muncul dalam teks, seperti "dan", "yang", atau "adalah". Hal ini memungkinkan pengamatan terhadap apakah keberadaan kata-kata tersebut memberikan kontribusi terhadap struktur kalimat yang lebih alami atau justru menjadi *noise* yang mengganggu proses pembelajaran model. Proses pelatihan dan pengujian dilakukan dengan dataset yang sama seperti skenario sebelumnya, dan hasil evaluasi akan dibandingkan menggunakan metrik ROUGE serta analisis visual grafik *loss* dan *accuracy*. Tabel 4.5 berikut menyajikan beberapa contoh hasil pra-pemrosesan pada skenario ini:

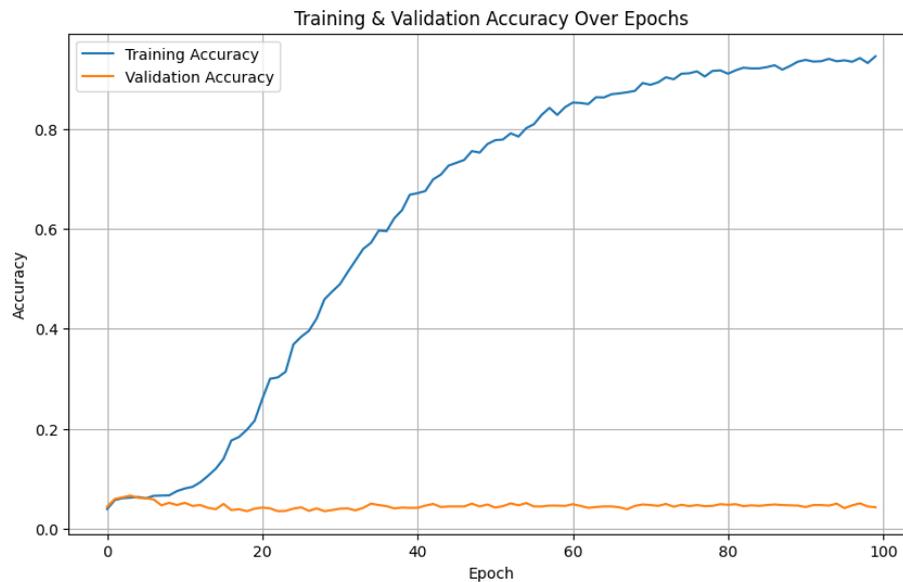
Tabel 4.6 Contoh Hasil *Preprocessing Stemming*

No	Teks Asli	Hasil <i>Preprocessing Stemming</i>
1.	Dahulu kala, disebuah lembah subur di Sumatra Utara, hiduplah seorang pemuda bernama Toba. sehari hari dia bekerja diladang sebagai seorang petani. Dia sebenarnya termasuk orang yang giat,	kala buah lembah subur sumatra utara hidup orang pemuda nama toba hari hari kerja ladang bagai orang tani benar masuk orang giat

No	Teks Asli	Hasil <i>Preprocessing Stemming</i>
	namun sayang hasil panen tahun ini benar-benar sangat sedikit...	sayang hasil panen tahun benarbenar sangat sedikit...
2.	kisah dari telur ajaib Alkisah hidup sebuah keluarga kecil sang ayah bernama Keo Dia adalah seorang duda yang mempunyai Putri Tunggal bernama Marika suatu hari Paman Geo jatuh hati kepada seorang janda Bernama...	kisah telur ajaib alkisah hidup buah keluarga kecil sang ayah nama keo adalah orang duda punya putri tunggal nama marika suatu hari paman geo jatuh hati orang janda nama macan milik orang putri...

Untuk menganalisis performa model pada skenario 2, dilakukan pelatihan dan evaluasi terhadap model LSTM menggunakan data yang telah diproses berdasarkan pendekatan skenario kedua. Evaluasi dilakukan dengan mengamati perubahan nilai akurasi dan loss selama proses pelatihan dan validasi dalam 100 *epoch*. Hasil evaluasi disajikan dalam dua grafik 4.2 berikut yang menggambarkan performa model dari segi akurasi dan loss.





Gambar 4.2 Grafik Hasil Skenario 2

Pada grafik pertama menunjukkan perkembangan nilai *training accuracy* dan *validation accuracy* terhadap jumlah *epoch*. Terlihat bahwa *training accuracy* mengalami peningkatan yang signifikan hingga mencapai lebih dari 0.9, menandakan bahwa model belajar dengan sangat baik terhadap data latih. Namun, *validation accuracy* tetap sangat rendah dan stagnan di kisaran 0.05 hingga akhir *epoch*. Hal ini menunjukkan adanya *overfitting*, di mana model sangat baik dalam menghafal data pelatihan, namun gagal untuk menggeneralisasi pada data validasi. Sementara itu, Gambar kedua menampilkan grafik *training loss* dan *validation loss* selama proses pelatihan. *Training loss* secara konsisten menurun, yang sejalan dengan peningkatan *training accuracy*. Namun, *validation loss* justru mengalami peningkatan secara bertahap, menunjukkan bahwa model tidak mampu meminimalkan kesalahan terhadap data validasi. Pola ini memperkuat indikasi bahwa model mengalami *overfitting*.

Dibandingkan dengan skenario pertama, hasil evaluasi ROUGE pada skenario kedua menunjukkan penurunan performa. Tabel 4.6 berikut menunjukkan nilai rata-rata ROUGE-1, ROUGE-2, dan ROUGE-L pada skenario kedua:

Tabel 4.7 Hasil Metrik Evaluasi Skenario 2

Metrik Evaluasi	Hasil Rata-rata
ROUGE-1	0,4071
ROUGE-2	0,1932
ROUGE-L	0,2911

Penurunan ini menunjukkan bahwa pada skenario kedua, model memiliki kemampuan yang lebih rendah dalam menangkap kata-kata penting, pasangan kata, serta urutan kalimat yang sesuai dengan ringkasan referensi. Hal ini dapat disebabkan oleh berbagai faktor, seperti parameter model yang digunakan, kualitas data pelatihan, atau kompleksitas struktur kalimat dalam data. Dengan demikian, skenario kedua belum mampu mengungguli skenario pertama dalam hal kualitas ringkasan yang dihasilkan menurut metrik ROUGE.

Untuk memberikan gambaran lebih jelas terkait kualitas ringkasan yang dihasilkan oleh model, berikut ditampilkan tabel 4.7 contoh perbandingan antara ringkasan referensi (buatan manusia) dengan ringkasan yang dihasilkan secara otomatis oleh mesin pada salah satu data uji.

Tabel 4.8 Perbandingan Ringkasan Manual dan mesin (LSTM)

No	Jenis	Teks
1.	Ringkasan Manual	Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri. Mereka dikaruniai anak bernama Samosir yang tumbuh menjadi pemalas. Suatu hari, Toba marah karena Samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut Samosir "anak ikan". Sang istri sedih dan menyuruh Samosir lari ke bukit. Bencana besar pun datang, sang istri kembali menjadi ikan dan Toba hanyut terbawa air bah. Tempat itu akhirnya menjadi Danau Toba dan bukit

No	Jenis	Teks
		tempat Samosir berlindung kini dikenal sebagai Pulau Samosir. Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri. Mereka dikaruniai anak bernama Samosir yang tumbuh menjadi pemalas. Suatu hari, Toba marah karena Samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut Samosir "anak ikan". Sang istri sedih dan menyuruh Samosir lari ke bukit. Bencana besar pun datang, sang istri kembali menjadi ikan dan Toba hanyut terbawa air bah. Tempat itu akhirnya menjadi Danau Toba dan bukit tempat Samosir berlindung kini dikenal sebagai Pulau Samosir.
2.	Ringkasan Mesin	di buah lembah subur di sumatra utara hidup toba orang tani yang hasil panen turun drastis untuk cukup butuh ia pancing di sungai dan hasil tangkap ikan sisik emas namun ikan itu ubah jadi wanita cantik yang nyata jelma putri yang kutuk mereka meni dengan syarat toba tak boleh ungit asal usul sang istri mereka karunia anak nama samosir yang tumbuh jadi malas suatu hari toba marah karena samosir makan bekal belum beri dan tanpa sadar langgar janji dengan sebut samosir anak ikan sang istri sedih dan suruh samosir lari ke bukit bencana besar pun datang sang istri kembali jadi ikan dan toba hanyut bawa air bah tempat itu akhir jadi danau toba dan bukit tempat samosir lindung kini kenal bagai pulau samosirdi buah lembah subur di sumatra utara hidup toba orang tani yang hasil panen turun drastis untuk cukup butuh ia pancing di sungai dan hasil tangkap ikan sisik emas namun ikan itu ubah jadi wanita cantik yang nyata jelma putri yang kutuk mereka meni dengan syarat toba tak boleh ungit asal usul sang istri mereka karunia anak nama samosir yang tumbuh jadi malas suatu hari toba marah karena samosir makan bekal belum beri dan tanpa sadar langgar janji dengan sebut samosir anak ikan sang istri sedih dan suruh samosir lari ke bukit bencana besar pun datang sang istri kembali jadi ikan dan toba hanyut bawa air bah tempat itu akhir jadi danau toba dan bukit tempat samosir lindung kini kenal bagai pulau samosir

### 4.5.3 Hasil Skenario 3: Menggunakan *Stopword Removal*

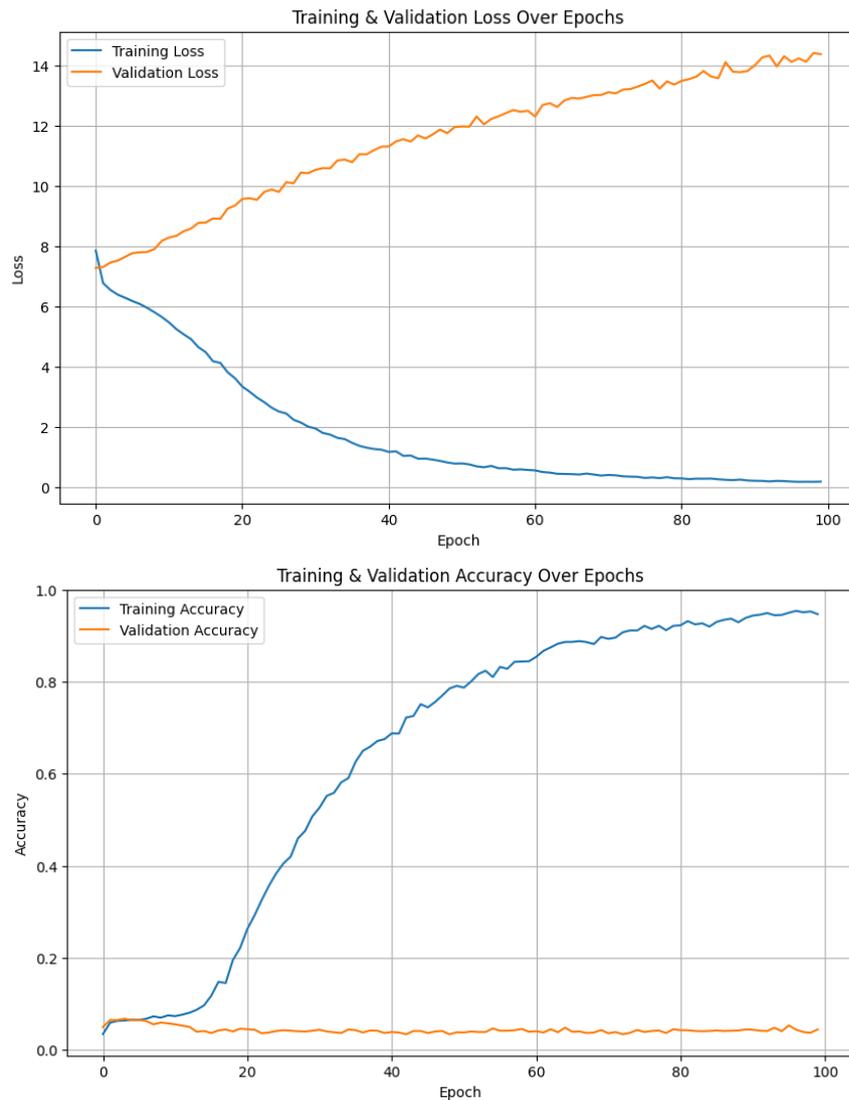
Pada skenario ketiga, model diuji dengan data yang telah diproses hanya menggunakan *stopword removal* tanpa melalui proses *stemming*. Tujuan dari skenario ini adalah untuk mengetahui bagaimana penghapusan kata-kata umum yang kurang bermakna (seperti “yang”, “dan”, “adalah”) dapat memengaruhi performa model, tanpa mengubah bentuk kata ke dasarnya. Dengan menghilangkan *stopword*, diharapkan model lebih fokus pada kata-kata bermakna tinggi dalam

teks, meskipun masih terdapat variasi morfologis yang tidak disederhanakan. Proses pelatihan dan evaluasi dilakukan seperti pada skenario sebelumnya, menggunakan dataset yang sama serta parameter pelatihan yang konsisten, sehingga hasilnya dapat dibandingkan secara adil dengan skenario 1 dan 2. Evaluasi dilakukan melalui grafik *training/validation loss, accuracy*, serta skor ROUGE. Dan berikut tabel 4.8 contoh hasil teks setelah *preprocessing Stopword Removal*.

Tabel 4.9 Hasil *Preprocessing Stopword Removal*

No	Teks Asli	Hasil <i>Preprocessing Stopword Removal</i>
1.	Dahulu kala, disebuah lembah subur di Sumatra Utara, hiduplah seorang pemuda bernama Toba. sehari hari dia bekerja diladang sebagai seorang petani. Dia sebenarnya termasuk orang yang giat, namun sayang hasil panen tahun ini benar-benar sangat sedikit...	dahulu kala buah lembah subur di sumatra utara hidup orang pemuda nama toba hari hari dia kerja ladang bagai orang tani dia benar masuk orang yang giat namun sayang hasil panen tahun ini benarbenar sangat sedikit...
2.	kisah dari telur ajaib Alkisah hidup sebuah keluarga kecil sang ayah bernama Keo Dia adalah seorang duda yang mempunyai Putri Tunggal bernama Marika suatu hari Paman Geo jatuh hati kepada seorang janda bernama macan dia juga memiliki seorang putri...	kisah dari telur ajaib alkisah hidup buah keluarga kecil sang ayah nama keo dia adalah orang duda yang punya putri tunggal nama marika suatu hari paman geo jatuh hati kepada orang janda nama macan dia juga milik orang putri...

Untuk mengevaluasi kinerja model pada skenario ketiga, dilakukan pemantauan terhadap perubahan akurasi dan loss selama proses pelatihan dan validasi. Grafik 4.3 menunjukkan perkembangan kedua metrik tersebut dari awal hingga akhir epoch. Melalui grafik ini, dapat dilihat seberapa baik model belajar dari data pelatihan sekaligus bagaimana kemampuannya dalam menggeneralisasi pada data validasi, sehingga memberikan gambaran menyeluruh mengenai performa model dalam skenario ini.



Gambar 4.3 Grafik Hasil Skenario 3

Berdasarkan dua grafik yang ditampilkan Gambar 4.3, kita dapat melihat performa model LSTM dalam melakukan peringkasan teks dengan skenario *preprocessing* yang hanya menggunakan stopword removal. Pada grafik pertama, yang menunjukkan *Training & Validation Accuracy Over Epochs*, terlihat bahwa akurasi pelatihan (*training accuracy*) meningkat secara signifikan seiring bertambahnya jumlah *epoch*, mencapai lebih dari 95% pada akhir pelatihan. Namun, akurasi validasi (*validation accuracy*) tetap sangat rendah dan nyaris tidak

menunjukkan peningkatan, stagnan di sekitar 5%. Hal ini mengindikasikan bahwa model mengalami *overfitting*, di mana model terlalu menyesuaikan diri dengan data pelatihan dan gagal menggeneralisasi pada data validasi atau data *testing*.

Hal ini juga diperkuat oleh grafik kedua, yaitu *Training & Validation Loss Over Epochs*. Grafik tersebut menunjukkan bahwa nilai *training loss* terus mengalami penurunan secara konsisten seiring bertambahnya jumlah *epoch*, yang menandakan bahwa model semakin mampu menyesuaikan diri terhadap data pelatihan. Artinya, model berhasil mempelajari pola-pola dalam data pelatihan dengan baik, sehingga kesalahan prediksi terhadap data tersebut menjadi semakin kecil.

Namun, di sisi lain, terlihat bahwa *validation loss* justru meningkat secara signifikan seiring waktu. Peningkatan ini mengindikasikan bahwa model mulai kehilangan kemampuannya dalam melakukan generalisasi terhadap data baru yang tidak pernah dilihat sebelumnya. Fenomena ini dikenal sebagai *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap data pelatihan termasuk terhadap pola data yang tidak relevan sehingga model tidak mampu menggeneralisasi dengan baik saat dihadapkan pada data *testing* atau data nyata di luar sampel pelatihan. *Overfitting* tetap terjadi jika model gagal generalisasi pada data baru, tetapi skor ROUGE bisa tetap tinggi jika model berhasil meniru kata-kata penting atau struktur umum ringkasan.

Hasil evaluasi menggunakan metrik ROUGE pada skenario ketiga menunjukkan peningkatan yang cukup signifikan dibandingkan skenario

sebelumnya. Tabel 4.9 berikut merangkum nilai rata-rata ROUGE-1, ROUGE-2, dan ROUGE-L yang diperoleh:

Tabel 4.10 Hasil Evaluasi Metrik Skenario 3

Evaluasi Matrik	Hasil Rata-rata
ROGUE-1	0,6060
ROGUE-2	0,3291
ROGUE-L	0,4613

Nilai ROUGE-1 yang mencapai lebih dari 0,6 mengindikasikan bahwa model mampu menangkap banyak kata penting dalam ringkasan referensi. Selain itu, peningkatan pada ROUGE-2 dan ROUGE-L menunjukkan kemampuan model yang lebih baik dalam menangkap pasangan kata dan struktur kalimat yang relevan. Hal ini mengisyaratkan bahwa penghapusan *stopword* tanpa proses *stemming* memberikan dampak positif terhadap kualitas ringkasan yang dihasilkan oleh model LSTM pada skenario ini.

Untuk memberikan gambaran yang lebih konkret mengenai hasil ringkasan yang dihasilkan pada skenario ketiga, berikut disajikan perbandingan antara ringkasan referensi buatan manusia dan ringkasan otomatis yang dihasilkan oleh model. Melalui perbandingan ini, dapat diamati sejauh mana model mampu mengekstraksi informasi penting serta menyusun kalimat secara koheren dan relevan dengan isi cerita. Ringkasan mesin pada skenario ini terlihat lebih ringkas dan informatif dibandingkan skenario sebelumnya, mencerminkan pengaruh positif dari penghapusan *stopword* dalam proses prapemrosesan. Tabel 4.10 berikut menyajikan hasil ringkasan dari kedua sumber tersebut.

Tabel 4.11 Hasil perbandingan Ringkasan Manual dan Mesin

No	Jenis	Teks
1.	Ringkasan Manual	Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan

No	Jenis	Teks
		<p>bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri. Mereka dikaruniai anak bernama Samosir yang tumbuh menjadi pemalas. Suatu hari, Toba marah karena Samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut Samosir "anak ikan". Sang istri sedih dan menyuruh Samosir lari ke bukit. Bencana besar pun datang, sang istri kembali menjadi ikan dan Toba hanyut terbawa air bah. Tempat itu akhirnya menjadi Danau Toba dan bukit tempat Samosir berlindung kini dikenal sebagai Pulau Samosir. Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri. Mereka dikaruniai anak bernama Samosir yang tumbuh menjadi pemalas. Suatu hari, Toba marah karena Samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut Samosir "anak ikan". Sang istri sedih dan menyuruh Samosir lari ke bukit. Bencana besar pun datang, sang istri kembali menjadi ikan dan Toba hanyut terbawa air bah. Tempat itu akhirnya menjadi Danau Toba dan bukit tempat Samosir berlindung kini dikenal sebagai Pulau Samosir.</p>
2.	Ringkasan Mesin (LSTM)	<p>di sebuah lembah subur di sumatra utara hiduplah toba seorang petani yang hasil panennya menurun drastis untuk mencukupi kebutuhan ia memancing di sungai dan berhasil menangkap ikan bersisik emas namun ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk mereka menikah dengan syarat toba tak boleh mengungkit asal usul sang istri mereka dikaruniai anak bernama samosir yang tumbuh menjadi pemalas suatu hari toba marah karena samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut samosir anak ikan sang istri sedih dan menyuruh samosir lari ke bukit bencana besar pun datang sang istri kembali menjadi ikan dan toba hanyut terbawa air bah tempat itu akhirnya menjadi danau toba dan bukit tempat samosir berlindung kini dikenal sebagai pulau samosirdi sebuah lembah subur di sumatra utara hiduplah toba seorang petani yang hasil panennya menurun drastis untuk mencukupi kebutuhan ia memancing di sungai dan berhasil menangkap ikan bersisik emas namun ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk mereka menikah dengan syarat toba tak boleh mengungkit asal usul sang istri mereka dikaruniai anak bernama samosir yang tumbuh menjadi pemalas suatu hari toba marah karena samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut samosir anak ikan sang istri sedih dan menyuruh samosir lari ke bukit bencana besar pun datang sang istri kembali menjadi ikan dan toba hanyut terbawa air bah tempat itu akhirnya menjadi danau toba dan bukit tempat samosir berlindung kini dikenal sebagai pulau samosir</p>

#### 4.5.4 Hasil Skenario 4: Tanpa *Stemming* dan *Stopword Removal*

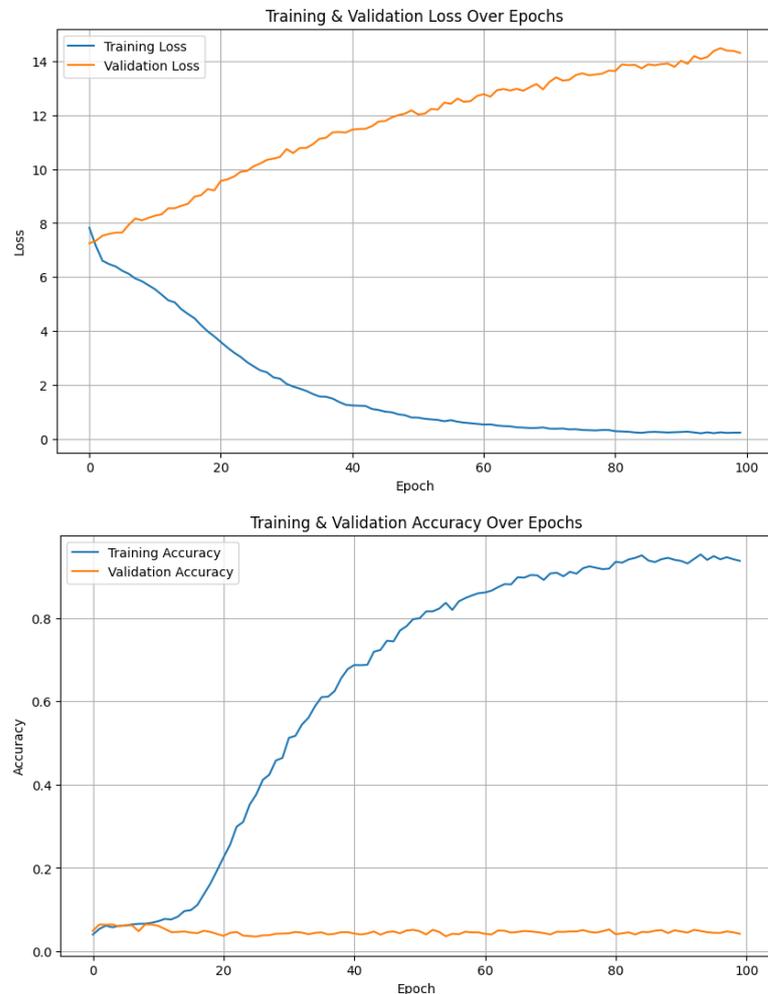
Pada skenario keempat, model diuji menggunakan data yang tidak melalui proses *stemming* maupun penghapusan *stopword*. Artinya, teks dimasukkan ke dalam model dalam bentuk yang lebih alami, tanpa normalisasi bentuk kata maupun penyaringan kata-kata umum. Tujuan dari skenario ini adalah untuk mengevaluasi sejauh mana model LSTM dapat memahami struktur bahasa dan mengekstraksi informasi penting tanpa intervensi prapemrosesan yang biasa digunakan untuk menyederhanakan data. Dengan tetap mempertahankan bentuk kata asli dan seluruh komponen dalam kalimat, model dihadapkan pada data yang lebih kompleks dan bervariasi. Proses pelatihan dan evaluasi dilakukan dengan parameter yang sama seperti pada skenario sebelumnya agar hasilnya dapat dibandingkan secara adil. Hasil evaluasi akan ditampilkan dalam bentuk grafik performa serta perhitungan metrik ROUGE. Tabel 4.11 berikut menyajikan contoh data teks asli yang digunakan pada skenario keempat tanpa melalui tahapan *preprocessing* apa pun.

Tabel 4.12 Contoh Teks Asli Tanpa *Preprocessing*

No	Teks Asli
1.	Dahulu kala, disebuah lembah subur di Sumatra Utara, hiduplah seorang pemuda bernama Toba. sehari hari dia bekerja diladang sebagai seorang petani. Dia sebenarnya termasuk orang yang giat, namun sayang hasil panen tahun ini benar-benar sangat sedikit. Tak seperti musim sebelumnya. Sehingga dia harus mencari penghasilan lain...
2.	kisah dari telur ajaib Alkisah hidup sebuah keluarga kecil sang ayah bernama Keo Dia adalah seorang duda yang mempunyai Putri Tunggal bernama Marika suatu hari Paman Geo jatuh hati kepada seorang janda bernama macan dia juga memiliki seorang putri bernama cantik mereka kemudian menikah dan hidup bersama belakangan Paman Keo...

Evaluasi performa model pada skenario keempat ditunjukkan melalui grafik akurasi dan loss terhadap data pelatihan dan validasi. Pada skenario ini, model dilatih menggunakan data mentah tanpa melalui proses *stemming* maupun

penghapusan *stopword*, sehingga seluruh bentuk kata dan kata-kata umum tetap dipertahankan. Berikut Grafik 4.4 hasil dari evaluasi performa pada skenario 4:



Gambar 4.4 Grafik Hasil Skenario 4

Berdasarkan grafik yang ditampilkan Gambar di mana proses *preprocessing* tidak menggunakan *stemming* maupun *stopword removal*, terlihat bahwa performa model LSTM dalam melakukan peringkasan teks mengalami permasalahan serius dalam hal generalisasi.

Pada grafik pertama yang memperlihatkan *Training & Validation Accuracy Over Epochs*, akurasi pelatihan (*training accuracy*) terus meningkat secara

signifikan hingga mendekati 95%, menandakan bahwa model mampu belajar dengan sangat baik terhadap data pelatihan. Namun, akurasi validasi (*validation accuracy*) tetap sangat rendah, tidak menunjukkan perbaikan berarti dan stagnan di kisaran 5% sepanjang proses pelatihan. Hal ini menunjukkan bahwa model mengalami *overfitting* yang parah, di mana model terlalu menyesuaikan diri dengan data pelatihan, namun gagal memahami atau mengenali pola pada data validasi.

Hal ini diperkuat oleh grafik kedua, *Training & Validation Loss Over Epochs*. Di sini, *training loss* menurun tajam dan konsisten, sementara *validation loss* terus meningkat secara signifikan seiring bertambahnya *epoch*. Peningkatan *validation loss* ini merupakan indikator kuat bahwa model tidak dapat menggeneralisasi dan kemungkinan besar hanya menghafal data pelatihan (*memorization*).

Kondisi ini menunjukkan bahwa *preprocessing* yang tidak menyaring *stopword* dan tidak melakukan *stemming* menyebabkan data masukan menjadi terlalu kompleks dan berisik, sehingga menyulitkan model dalam mempelajari pola yang umum dan relevan. Oleh karena itu, pendekatan *preprocessing* yang lebih baik seperti menghapus *stopword*, melakukan *stemming*, atau normalisasi lainnya sangat dianjurkan untuk meningkatkan performa model secara keseluruhan dan mencegah *overfitting*.

Kemudian untuk hasil pengujian kualitas ringkasan pada skenario keempat menunjukkan bahwa performa model mengalami penurunan ketika teks digunakan dalam bentuk mentah, tanpa melalui proses *stemming* dan penghapusan *stopword*. Penilaian menggunakan metrik ROUGE memperlihatkan bahwa kemampuan model dalam menghasilkan ringkasan yang relevan menjadi lebih terbatas. Nilai

rata-rata ROUGE-1, ROUGE-1, dan ROUGE-L dapat dilihat pada Tabel 4.12 berikut:

Tabel 4.13 Hasil Evaluasi Metrik Skenario 4

<b>Evaluasi Metrik</b>	<b>Hasil Rata-rata</b>
ROUGE-1	0,3745
ROUGE-2	0,1832
ROUGE-L	0,2812

Dari hasil evaluasi ROUGE pada skenario keempat, dapat disimpulkan bahwa penggunaan data teks mentah tanpa proses *stemming* dan penghapusan *stopword* berdampak negatif terhadap kualitas ringkasan yang dihasilkan oleh model LSTM. Nilai ROUGE-1, ROUGE-2, dan ROUGE-L yang relatif rendah menunjukkan bahwa model kesulitan dalam menangkap informasi penting dan menghasilkan ringkasan yang akurat dan relevan. Hal ini menegaskan pentingnya tahapan pra-pemrosesan dalam meningkatkan performa model, khususnya dalam hal penyederhanaan dan pembersihan data teks sebelum dilakukan pelatihan dan pengujian. Oleh karena itu, penerapan teknik *stemming* dan penghapusan *stopword* sangat disarankan untuk memperoleh hasil ringkasan yang lebih baik.

Untuk menilai kualitas ringkasan yang dihasilkan oleh model otomatis, dilakukan perbandingan langsung dengan ringkasan manual yang dibuat oleh ahli bahasa. Perbandingan ini bertujuan untuk melihat sejauh mana ringkasan mesin mampu mereplikasi inti informasi dan kejelasan yang terdapat pada ringkasan manual. Dengan demikian, evaluasi ini membantu mengidentifikasi kekuatan dan kelemahan model dalam menghasilkan ringkasan yang informatif dan mudah dipahami. Hasil perbandingan tersebut disajikan dalam Tabel 4.13.

Tabel 4.14 Tabel Perbandingan Ringkasan Manual dan Mesin

No	Jenis	Teks
1.	Ringkasan manual	<p>Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri. Mereka dikaruniai anak bernama Samosir yang tumbuh menjadi pemalas. Suatu hari, Toba marah karena Samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut Samosir "anak ikan". Sang istri sedih dan menyuruh Samosir lari ke bukit. Bencana besar pun datang, sang istri kembali menjadi ikan dan Toba hanyut terbawa air bah. Tempat itu akhirnya menjadi Danau Toba dan bukit tempat Samosir berlindung kini dikenal sebagai Pulau Samosir.</p> <p>Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri. Mereka dikaruniai anak bernama Samosir yang tumbuh menjadi pemalas. Suatu hari, Toba marah karena Samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut Samosir "anak ikan". Sang istri sedih dan menyuruh Samosir lari ke bukit. Bencana besar pun datang, sang istri kembali menjadi ikan dan Toba hanyut terbawa air bah. Tempat itu akhirnya menjadi Danau Toba dan bukit tempat Samosir berlindung kini dikenal sebagai Pulau Samosir.</p>
2.	Ringkasan mesin (LSTM)	<p>di sebuah lembah subur di sumatra utara hiduplah toba seorang petani yang hasil panennya menurun drastis untuk mencukupi kebutuhan ia memancing di sungai dan berhasil menangkap ikan bersisik emas namun ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk mereka menikah dengan syarat toba tak boleh mengungkit asal usul sang istri mereka dikaruniai anak bernama samosir yang tumbuh menjadi pemalas suatu hari toba marah karena samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut samosir anak ikan sang istri sedih dan menyuruh samosir lari ke bukit bencana besar pun datang sang istri kembali menjadi ikan dan toba hanyut terbawa air bah tempat itu akhirnya menjadi danau toba dan bukit tempat samosir berlindung kini dikenal sebagai pulau samosirdi sebuah lembah subur di sumatra utara hiduplah toba seorang petani yang hasil panennya menurun drastis untuk mencukupi kebutuhan ia memancing di sungai dan berhasil menangkap ikan bersisik emas namun ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk mereka menikah dengan syarat toba tak boleh mengungkit asal usul sang istri mereka dikaruniai anak bernama samosir yang tumbuh menjadi pemalas suatu hari toba marah karena samosir memakan bekal sebelum memberikannya dan tanpa sadar melanggar janjinya dengan menyebut samosir anak ikan sang istri sedih dan menyuruh samosir lari ke bukit bencana besar pun datang sang istri kembali menjadi ikan dan toba hanyut terbawa air bah tempat itu akhirnya</p>

No	Jenis	Teks
		menjadi danau toba dan bukit tempat samosir berlandung kini dikenal sebagai pulau samosir

#### 4.6 Perbandingan Hasil

Untuk memperoleh gambaran menyeluruh mengenai performa masing-masing skenario, dilakukan perbandingan nilai rata-rata dari metrik evaluasi ROUGE-1, ROUGE-2, dan ROUGE-L. Perbandingan ini bertujuan untuk mengetahui kinerja LSTM dan skenario mana yang menghasilkan kualitas ringkasan terbaik serta seberapa besar pengaruh proses *preprocessing* terhadap hasil akhir. Nilai-nilai yang dibandingkan merupakan hasil rata-rata dari 18 data *testing* pada tiap skenario. Hasil lengkap dari perbandingan tersebut ditampilkan pada Tabel 4.14 berikut:

Tabel 4.15 Tabel hasil matrik dari semua skenario

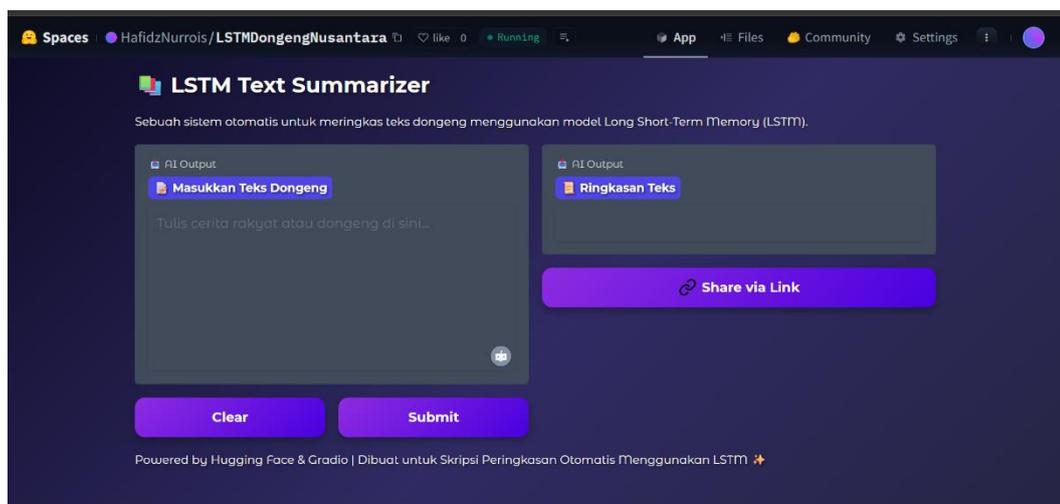
No	Skenario	ROGUE-1	ROGUE-2	ROGUE-L
1.	<i>Stemming &amp; Stopword removal</i>	0,5069	0,2019	0,3505
2.	<i>Stemming</i>	0,4071	0,1932	0,2911
3.	<i>Stopword removal</i>	0,6060	0,3291	0,4613
4.	Tanpa <i>Stemming &amp; Stopword removal</i>	0,3745	0,1832	0,2812

Berdasarkan data pada Tabel 4.14, dapat disimpulkan bahwa kinerja LSTM sudah baik karena mampu menghasilkan ringkasan mesin dengan skenario *preprocessing* terbaik yaitu penghapusan *stopword* saja yang menghasilkan nilai tertinggi pada ketiga metrik evaluasi. Hal ini menunjukkan bahwa menghilangkan kata-kata yang tidak memiliki makna penting dapat meningkatkan fokus model terhadap informasi inti dalam teks, tanpa mengganggu bentuk kata yang mungkin penting secara kontekstual. Sebaliknya, skenario tanpa prapemrosesan sama sekali menunjukkan performa terendah, yang mengindikasikan bahwa data mentah

cenderung menyulitkan model dalam menangkap informasi penting. Maka dari itu, proses *preprocessing* terbukti memiliki pengaruh signifikan terhadap performa model, dan kombinasi teknik yang tepat perlu dipertimbangkan untuk menghasilkan ringkasan teks yang optimal.

#### 4.7 Hasil Implementasi Sistem

Setelah melalui tahap perancangan dan pengembangan, sistem peringkasan otomatis berbasis *Long Short-Term Memory* (LSTM) berhasil diimplementasikan dalam bentuk *website*. Sistem ini dirancang untuk menerima masukan berupa teks dongeng dan menghasilkan ringkasan secara otomatis. Sebelum proses penerapan sistem, dilakukan persiapan model yang mencakup pembuatan *embedding matrix*, proses tokenisasi teks, serta pelatihan model untuk memperoleh bobot LSTM yang optimal. Seluruh komponen hasil pelatihan tersebut kemudian diunggah ke platform *Hugging Face* dan diintegrasikan ke dalam *Spaces* menggunakan framework Gradio. Dengan langkah ini, model dapat diakses dan dijalankan secara langsung melalui tampilan website yang sederhana dan *user-friendly*. Berikut adalah tampilan implementasinya:



Gambar 4.5 Hasil Implementasi Sistem

Website ini menyediakan antarmuka sederhana yang memungkinkan pengguna untuk langsung mencoba sistem peringkasan teks dongeng secara otomatis. Pada sisi kiri antarmuka, tersedia kolom *input* yang digunakan untuk memasukkan teks dongeng atau cerita rakyat yang ingin diringkaskan. Setelah teks dimasukkan dan tombol *Submit* ditekan, sistem akan memproses *input* menggunakan model LSTM yang telah dilatih sebelumnya. Hasil ringkasan kemudian akan ditampilkan di kolom kanan secara otomatis. Selain itu, terdapat tombol *Clear* untuk menghapus *input*. Desain *website* dibuat responsif dan mudah digunakan, sehingga mempermudah interaksi pengguna dalam menguji performa sistem peringkasan ini secara langsung.

#### 4.8 Pembahasan

Penelitian ini bertujuan untuk menghasilkan ringkasan otomatis dari cerita-cerita dongeng Nusantara menggunakan metode *Long Short-Term Memory*. Fokus utama penelitian adalah untuk menghasilkan ringkasan otomatis menggunakan LSTM dengan menentukan teknik *preprocessing* terbaik antara *stemming* dan

*stopword removal* serta mengevaluasi akurasi menggunakan ROGUE-1, ROGUE-2, dan ROGUE-L, baik digunakan secara terpisah maupun bersamaan, guna meningkatkan kualitas ringkasan otomatis. Hasil ringkasan mesin tersebut kemudian dievaluasi dengan membandingkannya terhadap ringkasan manual yang disusun oleh ahli bahasa sebagai acuan kebenaran. Data cerita dikumpulkan melalui platform YouTube dan diubah dari bentuk video menjadi teks melalui proses transkripsi. Teks transkrip inilah yang kemudian diringkas secara manual oleh ahli dan dijadikan dasar evaluasi kualitas ringkasan mesin.

Hasil pengujian menunjukkan bahwa pemilihan skenario *stemming* dan *stopword removal* memiliki pengaruh signifikan terhadap performa model LSTM. Setiap kombinasi *preprocessing* memberikan dampak yang berbeda terhadap kemampuan model dalam menghasilkan ringkasan yang relevan dan informatif. Skenario ketiga, yaitu penggunaan *stopword removal* tanpa *stemming*, memberikan hasil terbaik dengan nilai ROUGE-1 sebesar 0,6060, ROUGE-2 sebesar 0,3291, dan ROUGE-L sebesar 0,4613. Ini menunjukkan bahwa menghilangkan kata-kata tidak penting sambil mempertahankan bentuk kata asli memungkinkan model memahami konteks dan struktur kalimat secara lebih menyeluruh.

Sebaliknya, skenario keempat yang menggunakan data mentah tanpa prapemrosesan menghasilkan performa terendah, dengan nilai ROUGE-1 sebesar 0,3745, ROUGE-2 sebesar 0,1832, dan ROUGE-L sebesar 0,2812. Rendahnya hasil ini mengindikasikan bahwa keberadaan kata-kata tidak penting serta variasi bentuk kata yang tidak distandarisasi dapat mengganggu proses pembelajaran model, sehingga menghasilkan ringkasan yang kurang akurat. Temuan ini memperkuat

pentingnya penerapan *preprocessing* yang tepat untuk mendukung kinerja optimal dalam sistem peringkasan otomatis berbasis LSTM.

Visualisasi grafik pada keempat skenario pengujian menunjukkan pola yang cukup konsisten dalam proses pelatihan model. Grafik *Training & Validation Loss* memperlihatkan penurunan nilai loss secara bertahap seiring bertambahnya epoch, yang menandakan bahwa model berhasil belajar dari data dengan stabil. Skenario ketiga tampak memiliki penurunan loss yang paling signifikan dan stabil, yang sejalan dengan hasil evaluasi ROUGE-nya yang tertinggi. Sementara itu, pada skenario keempat, penurunan loss berjalan lebih lambat dan fluktuatif, yang menunjukkan kesulitan model dalam memahami pola data mentah tanpa prapemrosesan.

Grafik *Training & Validation Accuracy* juga mendukung temuan tersebut, di mana skenario ketiga memperlihatkan akurasi validasi yang lebih tinggi dan konsisten dibandingkan skenario lainnya. Ini menunjukkan bahwa model tidak hanya belajar dengan baik selama pelatihan, tetapi juga mampu menggeneralisasi pada data uji dengan lebih baik. Sebaliknya, pada skenario keempat, akurasi validasi cenderung stagnan dan tidak menunjukkan peningkatan yang signifikan, menandakan adanya *overfitting* atau kurangnya pembelajaran yang efektif dari data mentah. Dengan demikian, grafik ini memperkuat kesimpulan bahwa proses *preprocessing* memainkan peran penting dalam mendukung performa model LSTM dalam tugas peringkasan teks.

Berdasarkan tabel perbandingan dengan penelitian sebelumnya, kinerja LSTM pada skenario preprocessing terbaik menunjukkan hasil yang kompetitif, seperti ditunjukkan pada Tabel 4.15 berikut.

Tabel 4.16 Penelitian sebelumnya mengenai peringkasan teks LSTM

No	Penelitian	Metode	ROUGE-1	ROUGE-2	ROUGE-L
1.	(Alfhi Saputra, 2021)	LSTM	0.13846	-	-
2.	(Suleiman & Awajan, 2020)	LSTM	0,43	0,2	0,39
3.	(Sakhare, 2023)	LSTM	0,44	0,2	0,4
4.	(Al-jumaili, 2021)	LSTM	0,43	0,36	0,43
5.	(Awais et al., 2024)	LSTM	0,26	0,9	0,3
<b>6.</b>	<b>Penelitian ini</b>	<b>LSTM</b>	<b>0,6</b>	<b>0,32</b>	<b>0,46</b>

Tabel 4.15 menyajikan perbandingan skor evaluasi ROUGE antara beberapa penelitian terdahulu dan penelitian ini yang sama-sama menggunakan LSTM untuk tugas peringkasan teks otomatis. Berdasarkan hasil yang ditampilkan, penelitian ini menunjukkan kinerja yang paling unggul dengan skor ROUGE-1 sebesar 0,60, ROUGE-2 sebesar 0,32, dan ROUGE-L sebesar 0,46. Jika dibandingkan dengan penelitian sebelumnya, model pada penelitian ini menghasilkan peningkatan skor yang cukup konsisten di hampir semua metrik. Hal ini menunjukkan bahwa kinerja LSTM pada penelitian ini lebih baik dibandingkan penelitian sebelumnya. Berdasarkan hal tersebut, dapat disimpulkan bahwa metode LSTM yang digunakan dalam penelitian ini memiliki performa yang baik dalam melakukan peringkasan teks dongeng Nusantara. Selain itu, temuan ini juga selaras dengan hasil penelitian Uysal & Gunal (2014) yang menyatakan bahwa tahapan *preprocessing* berpengaruh signifikan dalam meningkatkan akurasi model.

Walau hasil evaluasi menunjukkan performa model yang lebih baik, perlu dipahami bahwa perbedaan skor ROUGE antara penelitian ini dan penelitian lain

bisa sangat dipengaruhi oleh jenis dan karakteristik dataset yang digunakan. Setiap data memiliki kompleksitas, variasi bahasa, dan jumlah sampel yang berbeda-beda, yang pada akhirnya dapat berdampak pada seberapa baik model mampu memahami dan meringkas teks.

#### 4.9 Integrasi Islam

Integrasi nilai-nilai Islam dalam penelitian ini tercermin dari perhatian Islam terhadap pentingnya menyampaikan dan mengambil pelajaran dari kisah-kisah umat terdahulu. Hal ini sebagaimana dijelaskan dalam surah Hud ayat 120:

وَكُلًّا نَقُصُّ عَلَيْكَ مِنْ أَنْبَاءِ الرُّسُلِ مَا نُثَبِّتُ بِهِ فُؤَادَكَ وَجَاءَكَ فِي هَذِهِ الْحَقُّ وَمَوْعِظَةٌ وَذِكْرَى لِلْمُؤْمِنِينَ ﴿١٢٠﴾

*"Dan semua kisah dari para rasul Kami ceritakan kepadamu, agar dengan kisah itu Kami teguhkan hatimu. Dan padanya telah datang kepadamu kebenaran serta pengajaran dan peringatan bagi orang-orang yang beriman." (QS. Hud:120).*

Ayat ini menegaskan bahwa penyampaian kembali kisah-kisah terdahulu memiliki nilai spiritual yang kuat, yaitu sebagai sarana untuk menguatkan hati, memberi pelajaran, dan meningkatkan keimanan. Menurut tafsir Wajiz (Ulama, 2025c), ayat ini menerangkan manfaat dari kisah-kisah para rasul bagi umat setelahnya. Semua kisah yang Allah sampaikan kepada Nabi Muhammad bertujuan untuk menguatkan hati beliau dalam menghadapi tantangan dan menjalankan tugas kenabian. Kisah-kisah itu mengandung kebenaran, nasihat yang mengarahkan kepada kebaikan, serta peringatan yang dapat diambil hikmahnya oleh orang-orang beriman. Dalam konteks skripsi ini, yang berjudul *"Peringkasan Teks Otomatis Dongeng Nusantara Menggunakan Metode LSTM"*, pengolahan cerita rakyat melalui teknologi tidak hanya bertujuan melestarikan budaya dan mengembangkan

sistem kecerdasan buatan di bidang pemrosesan bahasa alami, tetapi juga berperan dalam menyampaikan nilai-nilai moral dan keteladanan. Oleh karena itu, pendekatan ini mengandung unsur *hablumminallah*, yaitu membangun kedekatan dengan Allah SWT melalui pemanfaatan ilmu pengetahuan yang mendukung penyebaran hikmah serta memperkuat iman dalam masyarakat.

Keberagaman budaya dalam dongeng Nusantara mencerminkan kekayaan nilai yang ada di Indonesia. Penelitian ini bertujuan untuk melestarikan cerita-cerita tersebut agar lebih mudah dipahami melalui teknologi. Sebagaimana dijelaskan dalam Al-Qur'an Surah Al-Hujurat ayat 13:

يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ أَتْقَىٰكُمْ إِنَّ اللَّهَ عَلِيمٌ  
 خَبِيرٌ ﴿١٣﴾

*“Wahai manusia! Sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan seorang perempuan, kemudian Kami jadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah ialah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui, Maha Mengenal.” (QS. Al-Hujurat:13)*

Menurut Tafsir Wajiz (Ulama, 2025b) Wahai manusia! Sungguh, Kami telah menciptakan kamu dari seorang laki-laki dan seorang perempuan, yakni berasal dari keturunan yang sama yaitu Adam dan Hawa. Semua manusia sama saja derajat kemanusiaannya, tidak ada perbedaan antara satu suku dengan suku lainnya. Kemudian Kami jadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal dan dengan demikian saling membantu satu sama lain, bukan saling mengolok-olok dan saling memusuhi antara satu kelompok dengan lainnya. Allah tidak menyukai orang yang memperlihatkan kesombongan dengan keturunan,

kekayaan atau kepangkatan karena sungguh yang paling mulia di antara kamu di sisi Allah ialah orang yang paling bertakwa. Karena itu berusaha untuk meningkatkan ketakwaan agar menjadi orang yang mulia di sisi Allah. Sungguh, Allah Maha Mengetahui segala sesuatu baik yang lahir maupun yang tersembunyi, Mahateliti sehingga tidak satu pun gerak-gerik dan perbuatan manusia yang luput dari ilmu-Nya. Dari pemahaman tersebut, dapat disimpulkan bahwa keberagaman bukanlah penghalang, melainkan sumber kekayaan yang memperkuat nilai-nilai kemanusiaan. Hal ini sejalan dengan penelitian yang dilakukan, di mana dongeng-dongeng Nusantara yang berasal dari berbagai daerah dan latar budaya dikumpulkan, diproses, dan diringkas menggunakan teknologi kecerdasan buatan. Dengan menghadirkan dongeng-dongeng dari berbagai suku di Indonesia dalam bentuk yang lebih ringkas dan mudah dipahami, penelitian ini turut memperkuat jembatan sosial dan komunikasi antarmanusia, sesuai dengan semangat ayat tersebut. Inilah bentuk penerapan *hablum minannas* melalui pendekatan teknologi dan pelestarian nilai budaya dalam kehidupan masyarakat modern.

Penyampaian pesan moral dalam dongeng perlu disesuaikan dengan perkembangan teknologi agar lebih efisien dan mudah diterima oleh generasi saat ini. Penelitian ini menggunakan metode peringkasan teks otomatis untuk menyajikan kembali nilai-nilai dalam dongeng Nusantara secara ringkas namun tetap bermakna. Pendekatan ini sejalan dengan semangat dalam menyampaikan pesan secara efektif dan tepat sasaran, sebagaimana tertuang dalam Surah An-Nahl ayat 125:

كُلُوا وَاشْرَبُوا مِنْ رِزْقِ اللَّهِ وَلَا تَعْتُوا فِي الْأَرْضِ مُفْسِدِينَ ﴿٦٠﴾

“Makan dan minumlah rezeki (yang diberikan) Allah dan janganlah melakukan kejahatan di bumi dengan berbuat kerusakan.” (QS. Al-Baqarah:60)

Menurut tafsir Wajiz (Ulama, 2025a), ayat ini mengandung perintah agar manusia memanfaatkan rezeki yang Allah berikan dari bumi secara seimbang, dan larangan untuk berkeliaran di muka bumi dengan membuat kerusakan. Ayat ini memberikan arahan bahwa keberadaan manusia di bumi harus disertai dengan tanggung jawab menjaga dan melestarikan lingkungan.

Kaitannya dengan skripsi yang berjudul “*Peringkasan Teks Otomatis Dongeng Nusantara Menggunakan LSTM*”, banyak dongeng lokal mengandung pesan kearifan lingkungan seperti menjaga hutan, sungai, dan tanah sebagai bagian dari kehidupan. Dengan menerapkan teknologi untuk meringkas teks-teks tersebut, penelitian ini membantu menyebarkan kembali nilai-nilai cinta lingkungan kepada generasi digital, sehingga pesan kebaikan terhadap alam tidak hilang ditelan zaman. Ini mencerminkan penerapan *hablum minal ‘alam*, yakni menjalin hubungan yang baik dengan alam sebagai bagian dari ibadah dan tanggung jawab moral manusia di muka bumi.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Penelitian ini bertujuan untuk mengidentifikasi tahap *preprocessing* terbaik yang dapat meningkatkan performa model LSTM dalam menghasilkan ringkasan otomatis pada teks cerita dongeng Nusantara. Dalam penelitian ini, dilakukan empat skenario pengujian, yaitu: (1) penggunaan *stemming* dan *stopword removal*, (2) hanya *stemming*, (3) hanya *stopword removal*, dan (4) tanpa keduanya (teks mentah). Hasil pengujian menunjukkan bahwa skenario pertama, menghasilkan nilai ROUGE-1 sebesar 0.5069, ROUGE-2 sebesar 0.2019, dan ROUGE-L sebesar 0.3505. Pada skenario kedua, diperoleh nilai ROUGE-1 sebesar 0.4071, ROUGE-2 sebesar 0.1932, dan ROUGE-L sebesar 0.2911. Skenario ketiga, memberikan performa terbaik dengan nilai rata-rata ROUGE-1 sebesar 0.6060, ROUGE-2 sebesar 0.3291, dan ROUGE-L sebesar 0.4613. Sementara itu, skenario keempat memberikan hasil terburuk dengan nilai ROUGE-1 sebesar 0.3745, ROUGE-2 sebesar 0.1832, dan ROUGE-L sebesar 0.2812. Keempat skenario yang diuji menunjukkan bahwa kinerja LSTM sudah baik karena mampu menghasilkan ringkasan dengan hasil yang cukup baik dibandingkan penelitian sebelumnya. Dengan kinerja terbaik dicapai pada skenario ketiga, yaitu dengan menggunakan *stopword removal* tanpa *stemming*, yang menghasilkan ringkasan dengan nilai rata-rata ROUGE-1 sebesar 0.6060, ROUGE-2 sebesar 0.3291, dan ROUGE-L sebesar 0.4613.

## 5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan dan penelitian lebih lanjut dalam bidang peringkasan otomatis menggunakan metode LSTM:

- a. Jumlah data yang digunakan dalam penelitian ini masih terbatas. Untuk mendapatkan hasil yang lebih representatif dan meningkatkan kemampuan generalisasi model, disarankan untuk menambah jumlah data cerita, baik dari segi kuantitas maupun keragaman asal daerah, sehingga model dapat memahami berbagai gaya penulisan dan struktur narasi.
- b. Meskipun LSTM telah menunjukkan hasil yang cukup baik, akan lebih optimal jika dilakukan perbandingan dengan arsitektur deep learning lainnya seperti GRU, Transformer, atau BERT untuk mengetahui pendekatan mana yang paling efektif dalam tugas peringkasan cerita.
- c. Penelitian ini hanya menggunakan dua teknik utama dalam *preprocessing*, yaitu *stemming* dan *stopword removal*. Penelitian selanjutnya dapat mencoba teknik lain seperti *lemmatization*, *named entity recognition*, atau *part-of-speech tagging* untuk melihat pengaruhnya terhadap kualitas ringkasan.
- d. Selain evaluasi kuantitatif dengan ROUGE, sebaiknya ditambahkan juga evaluasi kualitatif seperti uji keterbacaan (*readability*) atau penilaian manual oleh ahli bahasa terhadap ringkasan mesin untuk memperoleh gambaran kualitas hasil secara lebih menyeluruh.

## DAFTAR PUSTAKA

- Abdillah, M. F. (2024). *Analisis Perbandingan Textrank Dan Long Short Term Memory Dalam Peringkasan Teks Berita Bahasa Inggris* [Universitas Islam Negeri Maulana Malik Ibrahim Malang]. <http://etheses.uin-malang.ac.id/65984/1/200605110068.pdf>
- Abdul, M., Dani, Q., Pratama, C. A., & Raihan, I. (2025). *Pemodelan Runtun Waktu Harga Nikel dengan Algoritma LSTM dan GRU*. 8, 392–398.
- Afandi, W., Saputro, S. N., Kusumaningrum, A. M., Adriansyah, H., Kafabi, M. H., & Sudianto, S. (2022). Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM. *Jurnal Informatika: Jurnal Pengembangan IT*, 7(2), 85–89. <https://doi.org/10.30591/jpit.v7i2.3401>
- Ahsan, T. H. (2023). *Peringkasan Teks Multi Dokumen Berita Berbahasa Indonesia Menggunakan Fasttext Dan K-Means Clustering* [Universitas Islam Negeri Maulana Malik Ibrahim Malang]. <http://etheses.uin-malang.ac.id/52460/7/19650026.pdf>
- Aisyah, K. N., Anggraini, S., & Arifin, A. Z. (2019). Peringkasan Teks Multi-Dokumen Berdasarkan Metode Sentence Extraction Dan Word Sense Disambiguation. *NJCA (Nusantara Journal of Computers and Its Applications)*, 4(1), 1–8. <https://doi.org/10.36564/njca.v4i1.89>
- Akbar, G. (2021). *Automatic Text Summarization Berita Bahas Indonesia Menggunakan Metode Attentional Encoder Decoder*. <https://dspace.uin.ac.id/handle/123456789/37691%0Ahttps://dspace.uin.ac.id/bitstream/handle/123456789/37691/14523091.pdf?sequence=1>
- AL-Hafidh, D. F., Rozi, I. F., & Putri, I. K. (2022). Peringkasan Teks Otomatis pada Portal Berita Olahraga menggunakan metode Maximum Marginal Relevance. *Jurnal Informatika Polinema*, 8(3), 21–30. <https://doi.org/10.33795/jip.v8i3.519>
- Al-jumaili, A. S. A. (2021). *Abstractive Text Summary of COVID-19 Documents based on LSTM Method and Abstractive Text Summary of COVID-19 Documents based on LSTM Method and Word Embedding. December 2021*. <https://doi.org/10.14704/WEB/V18I2/WEB18370>
- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, 2(1), 1–6.
- Alfhi Saputra, M. (2021). Peringkas Teks Otomatis Bahasa Indonesia secara Abstraktif Menggunakan Metode Long Short-Term Memory. *E-Proceeding of Engineering : Vol.8, No.2 April 2021* |, 8(2), 3474–3488.

- Alfin, M., Abidin, Z., & Basid, P. M. N. S. A. (2024). Peringkasan Multi Dokumen Berbahasa Indonesia Menggunakan Metode Recurrent Neural Network. *Techno.Com*, 23(1), 187–197. <https://doi.org/10.62411/tc.v23i1.9605>
- Algoritma, P., & Porter, N. D. A. N. (2025). *Peringkasan Teks Otomatis Dengan Latent Semantic Analysis Automatic Text Summarization In Indonesian Language Learning*. 4(1), 53–59.
- Alim, M. N. (2023). Pemodelan Time Series Data Saham LQ45 dengan Algoritma LSTM, RNN, dan Arima. *Prosiding Seminar Nasional Matematika*, 6, 694–701. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Anggraeni, D., & Rafiyanti, S. (2022). Pengaruh Dongeng terhadap Pendidikan Karakter Anak Usia Dini. *Jurnal Pendidikan Tambusai*, 6(1), 2485–2490.
- Antoko, T. D., Ridani, M. A., & Minarno, A. E. (2021). *Klasifikasi Buah Zaitun Menggunakan Convolution Neural Network*. 10(28), 119–126. <https://doi.org/10.34010/komputika.v10i2.4475>
- Arkadia, A., Hananto, B., & Prasvita, D. S. (2022). Optimasi Long Short Term Memory Dengan Adam Menggunakan Data Udara Kota DKI Jakarta. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 92–101.
- Awais, M., Muhammad, R. A. O., & Nawab, A. (2024). Abstractive Text Summarization for the Urdu Language : Data and Methods. *IEEE Access*, 12(May), 61198–61210. <https://doi.org/10.1109/ACCESS.2024.3378300>
- Ayu Latifah, S., & Suprayitno, E. (2021). Nilai Pendidikan Karakter dan Pesan Edukatif Dalam Dongeng Nusantara Bertutur. *Bahasa Dan Sastra*, 127–136.
- Ayu Syahfitri, R., Kurniawan, A., & Irsan Humaidy, M. (2022). Bulletin of Data Science Penerapan Algoritma Maximum Marginal Relevance Dalam Peringkasan Teks Secara Otomatis. *Media Online*, 1(2), 49–56. <https://ejurnal.seminar-id.com/index.php/bulletinds>
- Bahari, A., & Dewi, K. E. (2024). Peringkasan Teks Otomatis Abstraktif Menggunakan Transformer Pada Teks Bahasa Indonesia. *Komputa : Jurnal Ilmiah Komputer Dan Informatika*, 13(1), 83–91. <https://doi.org/10.34010/komputa.v13i1.11197>
- Caesarardhi, M. R., Vinarti, R. A., & Kusumawardani, R. P. (2023). Implementasi Aplikasi Peringkasan Teks Otomatis Untuk Atlas Penyakit Menular Pada Manusia Menggunakan Metode Ordered Abstractive Summarization. *Jurnal Linguistik Komputasional (JLK)*, 6(2), 70–86. <https://doi.org/10.26418/jlk.v6i2.182>
- Chamalah, E., Arsanti, M., & Setiana, L. N. (2021). Dongeng Islami Untuk Meningkatkan Budaya Literasi Sastra Anak Bagi Sisiwa Tpq Az-Zuhri. *Jurnal Pengabdian Masyarakat Bumi Raflesia*, 4(1), 512–519. <https://doi.org/10.36085/jpmbr.v4i1.1310>

- Diarsyah, M. G., & Setiawan, D. (2024). Implementasi CNN-LSTM untuk Music Captioning. *Media Informatika*, 23(1), 21–33. <https://doi.org/10.37595/mediainfo.v23i1.213>
- Dimas, E., Riberu, H., Kurniati, A., Oktaviani, U. D., Verawanto, T. J., Dimas, E., Riberu, H., Kurniati, A., Oktaviani, U. D., Joni, T., & Aristo, V. (2023). *Membentuk Karakter Siswa Dengan Dongeng Digital Nusantara Pendahuluan Pada karakter didefinisikan sebagai usaha menanamkan yang peserta didik mampu bersikap dan bersandarkan konsisten dan bagi kemajuan bangsa ( Khusniati , adanya karakter siswa . Guru san. 9(April), 219–233.*
- Dwi Maharani, & Nita priyanti. (2023). Pemanfaatan Youtube Cerita Rakyat Malin Kundang dalam Menumbuhkan Perkembangan Moral dan Kemampuan Bercerita Anak di TKIT Al-Wildan Bekasi. *EDUKASIA: Jurnal Pendidikan Dan Pembelajaran*, 4(1), 603–608. <https://doi.org/10.62775/edukasia.v4i1.310>
- Erzitha, R., Novaliendry, D., Budayawan, K., & Padang, U. N. (2025). *Analysis Of Dropout And Learning Rate On Bilstm- Ann Performance In Hate Speech Detection. 3(2), 221–232.*
- Factor, C. (2024). *Jurnal Rekayasa Sistem Informasi dan Teknologi Volume 1, No 3 – Februari 2024 e-ISSN : 3025-888X. 1(3), 137–149.*
- Fadhila, I. N., Fathurohman, I., & Setiawaty, R. (2023). Nilai Peduli Sesama Pada Kumpulan Dongeng Nusantara Karya Wahyu Astuti. *Prasasti Ilmu: Jurnal Ilmiah Pendidikan*, 3(1), 1–8.
- Fadlilah, M. F., Atmadja, A. R., & Firdaus, M. D. (2024). *Pemanfaatan Transformer untuk Peringkasan Teks : Studi Kasus pada Transkripsi Video Pembelajaran. 6(3), 2111–2119. https://doi.org/10.47065/bits.v6i3.6342*
- Fauzi, A. (2022). Penerapan Algoritma Text Mining dan Lexrank dalam Meringkas Teks Secara Otomatis. *Bulletin of Data Science*, 1(2), 65–72. <https://ejurnal.seminar-id.com/index.php/bulletinds>
- Fauziah, Y., Ilyas, R., & Kasyidi, F. (2022). Mesin Penterjemah Bahasa Indonesia-Bahasa Sunda Menggunakan Recurrent Neural Networks. *Jurnal Teknoinfo*, 16(2), 313. <https://doi.org/10.33365/jti.v16i2.1930>
- Fenardo, R., Mawardi, V. C., & Lauro, M. D. (2021). Indonesian Children’s Stories Summarization using Maximum Marginal Relevance Method. *Researchgate.Net, January 2022. https://www.researchgate.net/profile/Ricky-Fenardo/publication/358187551\_Indonesian\_Children’s\_Stories\_Summarization\_using\_Maximum\_Marginal\_Relevance\_Method/links/61f4152d11a1090a79b99e94/Indonesian-Childrens-Stories-Summarization-using-Maximum-Marginal-R*
- Firmansyah, M. R., Ilyas, R., & Kasyidi, F. (2020). Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network. *Prosiding The 11th Industrial*

*Research Workshop and National Seminar, 11(1), 488–495.*

- Flows, V. (2020). *Using Deep Learning to Forecast Maritime Vessel Flows*. 1–17. <https://doi.org/10.3390/s20061761>
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70 / 30 or 80 / 20 Relation Between Training and Testing Sets : A Pedagogical. *Departmental Technical Reports (CS), 1209*, 1–6. [https://scholarworks.utep.edu/cs\\_techrep](https://scholarworks.utep.edu/cs_techrep)
- Halimah, Surya Agustian, & Siti Ramadhani. (2022). Peringkasan teks otomatis (automated text summarization) pada artikel berbahasa indonesia menggunakan algoritma lexrank. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 371–381. <https://doi.org/10.37859/coscitech.v3i3.4300>
- Hartawan, G., Maylawati, D. S., & Uriawan, W. (2024). Bidirectional and Auto-Regressive Transformer (BART) for Indonesian Abstractive Text Summarization. *Jurnal Informatika Polinema, 10(4)*, 535–542. <https://doi.org/10.33795/jip.v10i4.5242>
- Hidjah, K., & Susilowati, D. (2025). *Implementasi Konsultasi Stunting Balita Menggunakan Large Language Models (LLMs)*. 6(1), 13–20.
- Husniyah, A. M. (2022). Media Aplikasi Dora (Dongeng Nusantara) Pada Pembelajaran Menyimak Dongeng Di Era Digital Di Sekolah Dasar. *Jurnal Cakrawala Pendas, 8(1)*, 316–325. <http://dx.doi.org/10.31949/jcp.v8i1.1975>
- Idhafi, Z., Agustian, S., Yanto, F., & Safaat H, N. (2023). Peringkas teks otomatis pada artikel berbahasa indonesia menggunakan metode maximum marginal relevance. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(3), 609–618. <https://doi.org/10.37859/coscitech.v4i3.6311>
- Irawati Setiawan, E., Ferdianto, A., Santoso, J., Kristian, Y., Gunawan, G., Sumpeno, S., & Hery Purnomo, M. (2020). Analisis Pendapat Masyarakat terhadap Berita Kesehatan Indonesia menggunakan Pemodelan Kalimat berbasis LSTM. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi, 9(1)*, 8–17. <https://doi.org/10.22146/jnteti.v9i1.115>
- IRHAMI, N. A. (2023). *Representasi Matan Jauharah Tauhid Menggunakan Metode Recurrent Neural Network Dengan Model Arsitektur Long Short-Term Memory Fakultas Sains Dan Teknologi Universitas Islam Negeri Ar-Raniry Banda Aceh 2023 M / 1444 H Menggunakan Metode Recurrent Neural Ne.*
- Ivanedra, K., & Mustikasari, M. (2019). Implementasi Metode Recurrent Neural Network Pada Text Summarization Dengan Teknik Abstraktif The Implementation Of Text Summarization With Abstractive Techniques Using Recurrent Neural Network Method. *Jurnal Teknologi Informasi Dan Ilmu Komputer, 6(4)*, 377–382. <https://doi.org/10.25126/jtiik.201961067>
- Kamal Wisyaldin, M., Maya Luciana, G., Pariaman, H., & Pembangkitan Jawa Bali, P. (2020). Pendekatan Long Short-Term Memory untuk Memprediksi

- Kondisi Motor 10 kV pada PLTU Batubara. *Kilat*, 9(2), 311–318. <https://doi.org/10.33322/kilat.v9i2.997>
- Kurniawan, B., & Romzi, M. (2022). Pembuatan dan Pelatihan Administrator Website pada Dinas Kesehatan Kabupaten Ogan Komering Ulu. *Jurnal Pengabdian Masyarakat (Abdira)*, 2(3), 253–258. <https://doi.org/10.31004/abdira.v2i3.202>
- Kurniawan, M., Kusriani, K., Arief, M. R., Studi, P., Teknik, M., & Yogyakarta, U. A. (2021). *Part of Speech Tagging Pada Teks Bahasa Indonesia dengan BiLSTM + CNN + CRF dan ELMo*. 29–37. <https://doi.org/10.30864/eksplora.v11i1.506>
- Marietta, B., & Rukun Santoso, A. A. S. (2025). *Peramalan Nilai Tukar Rupiah Terhadap Dolar Amerika Dengan Metode Long Short-Term Memory Dan Gated Recurrent Unit*. 14, 13–22. <https://doi.org/10.14710/j.gauss.14.1.13-22>
- Mawaridi, B. H. (2024). *Sistem Peringkasan Teks Berita Berbahasa Indonesia Menggunakan Latent Dirichlet Allocation Dan Maximum Marginal Relevance* [Universitas Islam Negeri Maulana Malik Ibrahim Malang]. <http://etheses.uin-malang.ac.id/66473/1/200605110011.pdf>
- Migunani, & Kevin Aditama. (2020). Pemanfaatan Natural Language Processing Dan Pattern Matching Dalam Pembelajaran Melalui Guru Virtual. *Elkom : Jurnal Elektronika Dan Komputer*, 13(1), 121–133. <https://doi.org/10.51903/elkom.v13i1.187>
- Moch Farryz Rizkilloh, & Sri Widiyanesti. (2022). Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(1), 25–31. <https://doi.org/10.29207/resti.v6i1.3630>
- Moghar, A. (2020). ScienceDirect ScienceDirect April Using Stock Market Prediction LSTM Neural Network a LSTM Recurrent Stock Market Prediction Using. *Procedia Computer Science*, 170, 1168–1173. <https://doi.org/10.1016/j.procs.2020.03.049>
- Mubarok, M. M. (2021). *Indonesian Abstractive News Summarization Berbasis Deep Learning Dengan Metode Sequence-*.
- Murtiningsih, R. S. (2022). Dimensi Ideologis Dongeng Anak Nusantara Relevansinya Bagi Pendidikan Karakter Di Indonesia. *Humanika*, 29(1), 36–45. <https://doi.org/10.14710/humanika.v29i1.43668>
- Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (CNN) Pada Ekspresi Manusia. *Algor*, 2(1), 12–21.
- Phua, Y. T., Navaratnam, S., Kang, C. M., & Chew, W. S. (2022). Sequence-to-sequence neural machine translation for English-Malay. *IAES International Journal of Artificial Intelligence*, 11(2), 658–665.

<https://doi.org/10.11591/ijai.v11.i2.pp658-665>

- Pratama, F. D. (2024). *Sentimen Pengguna Sosial Media Youtube Terhadap Pencalonan Presiden Ganjar Pranowo Dan Prabowo Subianto Pada Pemilu 2024*. 2024.
- Pratama, W., Ilyas, R., & Kasyidi, F. (2022). Peringkasan Otomatis Makalah Menggunakan Maximum Marginal Relevance. *Informatics and Digital Expert (INDEX)*, 3(1), 32–37. <https://doi.org/10.36423/index.v3i1.677>
- Pratiwi, H. A., Cahyanti, M., & Lamsani, M. (2021). Implementasi Deep Learning Flower Scanner Menggunakan Metode Convolutional Neural Network. *Sebatik*, 25(1), 124–130. <https://doi.org/10.46984/sebatik.v25i1.1297>
- Putra, D. R. R., & Saputra, R. A. (2023). Implementasi Convolutional Neural Network (Cnn) Untuk Mendeteksi Penggunaan Masker Pada Gambar. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3), 710–714. <https://doi.org/10.23960/jitet.v11i3.3286>
- Qori Qordofa, M., & As'ad, M. (2022). Metode Dakwah KH. Ahmad Baha'uddin Nursalim (Gus Baha) Melalui Channel Santri Gayeng Di Media Youtube. *Syiar | Jurnal Komunikasi Dan Penyiaran Islam*, 2(1), 1–10. <https://doi.org/10.54150/syiar.v2i1.52>
- Rakhman, R. T., Piliang, Y. A., Ahmad, H. A., & Gunawan, I. (2021). Pemetaan Jenis Dongeng Nusantara Dalam Infografis. *ANDHARUPA: Jurnal Desain Komunikasi Visual & Multimedia*, 7(01), 59–78. <https://doi.org/10.33633/andharupa.v7i01.3938>
- Reimers, N., & Gurevych, I. (2016). *Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks*.
- Rochmawati, N., Hidayati, H. B., Yamasari, Y., Tjahyaningtjas, H. P. A., Yustanti, W., & Prihanto, A. (2021). Analisa Learning Rate dan Batch Size pada Klasifikasi Covid Menggunakan Deep Learning dengan Optimizer Adam. *Journal of Information Engineering and Educational Technology*, 5(2), 44–48. <https://doi.org/10.26740/jieet.v5n2.p44-48>
- Rusbandi, M., Rozi, I. F., & Batubulan, K. S. (2021). Otomatisasi Peringkasan Teks Pada Dokumen Hukum Menggunakan Metode Latent Semantic Analysis. *Jurnal Informatika Polinema*, 7(3), 9–16. <https://doi.org/10.33795/jip.v7i3.515>
- Sakhare, D. Y. (2023). *A Sequence-to-Sequence Text Summarization Using Long Short-Term Memory Based Neural Approach*. 16(2), 142–151. <https://doi.org/10.22266/ijies2023.0430.12>
- Santosa, A., Purnamasari, I., & Mayasari Rini. (2022). Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6, 81–93.

- Sari, I. P., Azzahrah, A., Qathrunada, I. F., Lubis, N., & Anggraini, T. (2022). Perancangan Sistem Absensi Pegawai Kantoran Secara Online pada Website Berbasis HTML dan CSS. *Blend Sains Jurnal Teknik*, 1(1), 8–15. <https://doi.org/10.56211/blendsains.v1i1.66>
- Setiawan, A. Y., Darmawiguna, I. G. M., & Pradnyana, G. (2022). *Sentiment Summarization Learning Evaluation Using LSTM (Long Short Term Memory) Algorithm*. <https://doi.org/10.4108/eai.27-11-2021.2315533>
- Sianturi, T. B., Cholissodin, I., & Yudistira, N. (2023). Penerapan Algoritma Long Short-Term Memory (LSTM) berbasis Multi Fungsi Aktivasi Terbobot dalam Prediksi Harga Ethereum. 7(3), 1101–1107.
- Sitanggang, C. E., Firda, D. A., Ramadhini, R., & Panjaitan, J. M. (2024). *Studi Literatur : Penggunaan Media Sosial Sebagai Alat Promosi Usaha*. 14(1), 23–29.
- Somantri, A. (2022). Perancangan Model Penilaian Kualitas Informasi di Fanpage Facebook. *Jurnal Pasundan Informatika*, 1(01), 7. <https://doi.org/10.23969/pasinformatik.v1i01.5119>
- St Tuhpatussania, & Erniwati, S. (2025). Perbandingan Algoritma Nazief-Adriani Dan Porter Untuk Peringkasan Teks Otomatis Dengan Latent Semantic Analysis Pada Modul Pembelajaran Berbahasa Indonesia. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, 4(1), 53–59. <https://doi.org/10.69916/jkbti.v4i1.202>
- Subowo, E., Adi Artanto, F., Putri, I., & Umaedi, W. (2022). Algoritma Bidirectional Long Short Term Memory untuk Analisis Sentimen Berbasis Aspek pada Aplikasi Belanja Online dengan Cicilan. *Jurnal Fasilkom*, 12(2), 132–140. <https://doi.org/10.37859/jf.v12i2.3759>
- Suharto, A. (2023). Fundamental Bahasa Pemrograman Python. *Eureka Media Aksara*, 1–25.
- Suleiman, D., & Awajan, A. (2020). *Deep Learning Based Abstractive Text Summarization : Approaches , Datasets , Evaluation Measures , and Challenges*. 2020. <https://doi.org/10.1155/2020/9365340>
- Tarkus, D., Sompie, S. R. U. A., & Jacobus, A. (2020). Implementasi Metode Recurrent Neural Network pada Pengklasifikasian Kualitas Telur Puyuh. *Jurnal Teknik Informatika*, 15(2), 137–144.
- Ulama, N. (2025a). *Tafsir Qur'an Surat Al-Baqarah ayat 60*. <https://quran.nu.or.id/al-baqarah/60>
- Ulama, N. (2025b). *Tafsir Qur'an Surat Al-Hujurat ayat 13*. <https://quran.nu.or.id/al-hujurat/13>
- Ulama, N. (2025c). *Tafsir Qur'an Surat Hud ayat 120*. <https://quran.nu.or.id/hud/120>

- Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, 50, 104–112. <https://doi.org/10.1016/j.ipm.2013.08.006>
- Yosia Wibowo, L., Annisa, N., Ananda Khairunnisa, P., Handrianus Pranatawijaya, V., & Priskila, R. (2024). Implementasi Long Short-Term Memory Dalam Analisis Sentimen Pengguna Aplikasi Twitter Yang Mengandung Ujaran Kebencian. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3170–3174. <https://doi.org/10.36040/jati.v8i3.9654>
- Yuliska, Y., & Syaliman, K. U. (2020). Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia. *IT Journal Research and Development*, 5(1), 19–31. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).4688](https://doi.org/10.25299/itjrd.2020.vol5(1).4688)
- Zahara, S., Sugianto, & M. Bahril Ilmiddafiq. (2019). Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 357–363. <https://doi.org/10.29207/resti.v3i3.1086>
- Zhu, Y., Chen, Y., & Zhang, B. (2020). *Automatic , Dynamic , and Nearly Optimal Learning Rate Specification by Local Quadratic Approximation*.

# LAMPIRAN

## Lampiran I. Profil Ahli bahasa

Profil ahli sebagai pembuat ringkasan manusia

Nama : Ahmad Anas Aular Rofiq S.Pd

Profesi : Guru Bahasa Indonesia dan Pendidikan Agama Islam di SDN Merjosari 5



## Lampiran II

### Data Teks Berita Asli dan Ringkasan Manual

No	Teks Dongeng	Ringkasan Manual
1.	Dahulu kala, disebuah lembah subur di Sumatra Utara, hiduplah seorang pemuda bernama Toba. sehari hari dia bekerja diladang sebagai seorang petani. Dia sebenarnya termasuk orang yang giat, namun sayang hasil panen tahun ini benar-benar sangat sedikit. Tak seperti musim sebelumnya. Sehingga dia harus mencari penghasilan lain. ...	Di sebuah lembah subur di Sumatra Utara, hiduplah Toba, seorang petani yang hasil panennya menurun drastis. Untuk mencukupi kebutuhan, ia memancing di sungai dan berhasil menangkap ikan bersisik emas. Namun, ikan itu berubah menjadi wanita cantik yang ternyata jelmaan putri yang dikutuk. Mereka menikah dengan syarat Toba tak boleh mengungkit asal usul sang istri....
2.	kisah dari telur ajaib Alkisah hidup sebuah keluarga kecil sang ayah bernama Keo Dia adalah seorang duda yang mempunyai Putri Tunggal bernama Marika suatu hari Paman Geo jatuh hati kepada seorang janda bernama macan dia juga memiliki seorang putri bernama cantik mereka kemudian menikah dan hidup bersama belakangan Paman Keo ...	Marika adalah gadis baik yang tinggal bersama ayahnya, Keo, dan ibu tirinya, Macan, serta saudari tirinya, Cantik. Setelah ayahnya pergi bekerja, Macan dan Cantik menyiksa Marika. Suatu hari, Marika kabur dan bertemu nenek tua misterius yang memberinya tempat tinggal dan tiga telur ajaib. Saat pulang, telur-telur itu mengeluarkan emas dan berlian ...
3.	Alkisah dahulu kala di negeri Tonjeng Beru berdirilah sebuah kerajaan bernama kerajaan Sekar Kuning yang terletak di tepi pantai menghadap Samudra Hindia. Sang Raja bernama Raden Pandji Kusuma atau biasa dikenal dengan sebutan raja Tonjeng Beru. Dia memiliki istri bernama Dewi Seranting. Keduanya memerintah dengan bijaksana hingga rakyatnya pun hidup makmur dan sejahtera...	Di kerajaan Sekar Kuning, Putri Mandalika tumbuh jadi gadis cantik, rendah hati, dan dicintai rakyat. Banyak pangeran dari negeri lain datang melamar, membawa hadiah dan pujian. Namun, Mandalika justru bingung dan sedih karena lamaran itu memicu persaingan dan ancaman perang antar kerajaan. Setelah bermeditasi di tebing Pantai Seger,...
4.	Alkisah dahulu kala di negeri Tonjeng Beru berdirilah sebuah kerajaan bernama kerajaan Sekar Kuning yang terletak di tepi pantai menghadap Samudra Hindia. Sang Raja bernama Raden Pandji Kusuma atau biasa dikenal dengan sebutan raja Tonjeng Beru. Dia memiliki istri bernama Dewi Seranting. Keduanya memerintah dengan bijaksana hingga rakyatnya pun hidup makmur dan sejahtera...	Di kerajaan Sekar Kuning, Putri Mandalika tumbuh jadi gadis cantik, rendah hati, dan dicintai rakyat. Banyak pangeran dari negeri lain datang melamar, membawa hadiah dan pujian. Namun, Mandalika justru bingung dan sedih karena lamaran itu memicu persaingan dan ancaman perang antar kerajaan. Setelah bermeditasi di tebing Pantai Seger,...
5.	cerita dari pahat ajaib pada suatu ketika hidup seorang laki-laki bernama Wirata dia merupakan seorang yatim piatu Dia adalah orang yang kekurangan tetapi tak	Wirata, seorang pemuda miskin yatim piatu, hidup dengan bekerja keras dan berbagi dengan sesama. Suatu hari, ia bertemu dengan seorang kakek tua yang memberinya

No	Teks Dongeng	Ringkasan Manual
	pernah lelah bekerja keras untuk memenuhi kebutuhan hidup sehari-hari. Seti Ap hari setelah menebang kayu secukupnya dia selalu beristirahat dan makan siang di bawah pohon besar...	pahat ajaib. Pahat itu mampu mengubah patung kayu menjadi benda nyata, bahkan emas. Dengan kemampuan ini, Wirata menjadi pematung terkenal dan kaya raya...
...	...	...
30.	Alkisah dahulu kala, ada sebuah kerajaan besar bernama Kerajaan Jenggala. Kerajaan ini dipimpin oleh raja bernama Raden Putra. Raden Putra adalah raja yang arif dan bijaksana. Raden Putra memiliki seorang permaisuri dan juga seorang selir...	Kerajaan Jenggala dipimpin oleh Raja Raden Putra yang bijaksana, memiliki permaisuri yang baik hati dan selir yang iri. Suatu hari, permaisuri hamil, membuat selir cemburu dan berencana menyingkirkannya dengan racun...
31.	Alkisah hidup sebuah keluarga kecil sang ayah bernama Keo. Dia adalah seorang duda yang mempunyai Putri Tunggal bernama Marika. Suatu hari Paman Geo jatuh hati kepada seorang janda bernama Macan. Dia juga memiliki seorang putri bernama cantik mereka kemudian menikah dan hidup bersama...	Suatu hari, Paman Keo, seorang duda dengan putri bernama Marika, menikahi Macan yang memiliki putri cantik. Keo harus pergi bekerja, dan sebelum berangkat, ia berpesan kepada Marika untuk menghormati Macan....
32.	Dikisahkan bahwa pada suatu waktu, Dewi Sembadra sedang duduk sendirian dan bersantai di taman Maduganda. Keadaan sekitar Dewi Sembadra saat itu lengang tanpa pengawasan berarti dari pengawal istana terutama Dewi Srikandi...	Dewi Sembadra sedang duduk sendirian di taman Maduganda ketika Burisrawa, yang telah lama mencintainya, mencoba mendekatinya meskipun Sembadra telah bersuami pada Arjuna. Meskipun Sembadra menolak, Burisrawa memaksa dan mengancam dengan keris...
33.	apa Kalian sedang mencari kedua anak kalian saat aku pergi ke ladang aku melihat anak kalian berubah menjadi kelinci-kelinci ini sudah Jangan ber orang itu Han akan mendatangkan amarah jadikim kisah gucianah Li beris sebandang saat..	Surya, seorang pedagang, menitipkan guci berisi emas kepada Pamanjo karena takut dicuri. Namun, setelah kembali, Surya mendapati guci itu berisi biji semangka dan Pamanjo menyangkal pernah menerima emas. Merasa tertipu, Surya menyembunyikan kedua anak Pamanjo dan mengubah mereka menjadi kelinci..
34.	Harimau Tunggu dengarkan aku dulu berhenti Ah ternyata kau membawaku ke sana hanya untuk melunasi hutangmu Hah makan Li ekor Harimau saja tidak cukup binatang apa itu kisah harimau yang dibodohi seekor kelinci diceritakan ada seekor harimau...	Seekor harimau yang sedang lapar ingin menangkap ikan seperti yang dilakukan elang laut. Dia mencoba menirunya, tetapi gagal dan hampir tenggelam. Seorang pria yang melihat kejadian itu berjanji kepada harimau untuk membawa buruan setiap hari agar kejadian tersebut tidak tersebar...
35.	kisah ikan mas ajaib Pada suatu hari ada sepasang suami istri yang sudah tua mereka hidup dengan kemiskinan dan sudah sejak lama mereka membangun	Suatu hari, seorang lelaki tua menangkap ikan mas ajaib yang bisa berbicara. Ikan itu meminta dibebaskan, menawarkan untuk mengabdikan permintaan lelaki tua jika ia

No	Teks Dongeng	Ringkasan Manual
	gubuk kecil di dekat laut pasangan lelaki mencari nafkah dengan cara menangkap ikan sedangkan pasangan perempuannya selalu memotong Rami dan menjualnya...	melepaskannya. Lelaki tua itu melepaskan ikan tanpa meminta imbalan, namun istrinya marah...
...	...	...
60	Dahulu kala ada sebuah pulau di daerah Sulawesi tengah, yang memiliki pesisir pantai nan indah. Di sekitar pantai itu. Hiduplah satu keluarga nelayan. Sang ayah mencari ikan dilaut untuk memenuhi kebutuhan sehari-hari keluarga kecil itu. "Hei anak- anak. Lihatlah ! Hari ini ayah mendapat tangkapan lumayan banyak !"...	Pada suatu waktu, di sebuah pulau di Sulawesi Tengah, tinggal sebuah keluarga nelayan. Sang ayah selalu melaut mencari ikan, sementara ibu mengurus rumah dan kebun. Suatu hari, setelah pulang melaut, sang ayah membawa ikan banyak, namun si bungsu yang kelaparan akhirnya memakan semua ikan tersebut, meskipun ibunya sudah berpesan untuk menyisakan dua ikan untuk ayah...
61.	Alkisah Dahulu kala, di sebelah utara kota Garut, terdapat sebuah desa, dimana kebanyakan warga di desa itu bekerja sebagai petani. Tanahnya pun subur dan sumber air tumpah ruah. Hasil panen yang melimpah tak membuat warga desa itu makmur...	Di sebuah desa di utara Garut, Nyai Endit, seorang janda kaya, menguasai hidup petani dengan cara menekan mereka menjual hasil panen dengan harga murah. Ketika musim panen tiba, petani harus membeli kembali beras dengan harga tinggi...
62.	Alkisah dahulu kala hiduplah seorang pertapa sakti bertubuh raksasa bernama Syeh Tuan Tapa. Karena kebiasaannya bertapa di Gua Kalam, di Teluk Aceh Selatan. Tuan Tapa sangat taat dalam beribadah. Dia menghabiskan seluruh waktunya untuk selalu berzikir dan terus mengingat nama sang pencipta...	Dahulu, seorang pertapa sakti bernama Syeh Tuan Tapa tinggal di Gua Kalam, Teluk Aceh Selatan. Suatu hari, dua naga dari Cina yang diusir karena tidak memiliki keturunan, menemukan bayi yang terombang ambing di laut dan membawanya kepada Tuan Tapa...
63.	Dahulu kala, hidup lah seorang Begawan Sakti Mandraguna bernama Begawan Sidik Wacana. Sang Begawan memiliki anak laki-laki bernama Joko Lelono. Setelah Joko Lelono beranjak dewasa, Begawan Sidik Wacana merasa bahwa sang anak sudah saatnya memiliki pendamping hidup untuk menemani dan mengurus semua kebutuhannya....	Begawan Sidik Wacana memiliki anak bernama Joko Lelono yang telah dewasa. Setelah percakapan tentang mencari pendamping hidup, Joko Lelono menjadi buta akibat salah paham dengan ayahnya. Begawan Sidik Wacana meminta Joko Lelono menemui Begawan Sidik Waseso untuk penyembuhan...
...	...	...
89	Alkisah dahulu kala di wilayah barat bumi perdikan Sendang Kamulyan terdapat sebuah padepokan, yang bernama Padepokan Sinawang. Wilayah tersebut berada di bawah kekuasaan Majapahit yang saat itu sedang dirundung duka...	Alkisah, Raden Ayu Saraswati, putri Majapahit, diasingkan ke Padepokan Sinawang karena penyakit aneh yang membuat tubuhnya berbau amis. Ia disembuhkan oleh pemuda tampan bernama Srobo, yang ternyata buaya putih penguasa Sungai Bagong...

<b>No</b>	<b>Teks Dongeng</b>	<b>Ringkasan Manual</b>
90.	Setelah beberapa hari berperang, kerajaan Klungkung mengalami kekalahan oleh kerajaan Blambangan yang dipimpin oleh Prabu Menak Prakoso. Putra dan putri raja Klungkung yaitu Raden Rupaksa dan Putri Surati berhasil selamat. Mereka melarikan diri ke dalam hutan sebelum perang berkecamuk...	Kerajaan Klungkung kalah oleh Blambangan, namun Raden Rupaksa dan Putri Surati selamat dan melarikan diri ke hutan. Untuk mengelabui musuh, mereka berpisah. Di hutan, Surati bertemu Raden Banterang, putra Prabu Menak Prakoso. Mereka jatuh cinta dan menikah di Blambangan meski sempat ditentang. Suatu hari, Surati bertemu kembali dengan kakaknya..
91.	Dahulu kala, di sebuah desa bernama Karang Kejambon. Desa ini subur makmur yang berada di Jawa Timur. Hiduplah seorang perempuan bernama Wandan Wanguri dan memiliki seorang anak laki-laki bernama Joko Tulus....	Di desa Karang Kejambon, Joko Tulus, yang dikenal dengan julukan Kebo Kicak, tumbuh sebagai anak yang kuat namun keras kepala. Dia mencari ayahnya, Patih Pangulang Jagad di Majapahit, dan setelah mengalahkan Bajul Ijo di Sungai Brantas sebagai bukti,...