



**PENERAPAN *RECURRENT NEURAL NETWORK* UNTUK SISTEM  
PERINGKASAN DAN PENERJEMAHAN TEKS BERITA BAHASA  
INGGRIS KE BAHASA INDONESIA**

**SKRIPSI**

**ADIKA RAJENDRA HARIS**

**1910511071**

**S1 INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2022**

## LEMBAR PENGESAHAN

### LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Adika Rajendra Haris  
NIM : 1910511071  
Program Studi : S1 Informatika  
Judul : Penerapan Recurrent Neural Network untuk Sistem  
Peringkasan dan Penerjemahan Teks Berita Bahasa Inggris  
ke Bahasa Indonesia

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta.

(Dr. Widva Cholil, M.I.T.)  
Penguji I

(Nur Hafifah Matondang, S.Kom, MM.)  
Penguji II

(Rio Wirawan, S.Kom., MMSI)  
Dosen Pembimbing



(Dr. Widva Cholil, M.I.T.)  
Ketua Program Studi

Ditetapkan di  
Tanggal Ujian

: Jakarta  
: 3 Juli, 2023



## PERNYATAAN ORISINALITAS

### PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun ditujuk telah saya nyatakan dengan benar.

Nama : Adika Rajendra Haris

NIM : 1910511071

Program Studi : S1 – Informatika

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 05 Juni 2023

Yang Menyatakan,



(Adika Rajendra Haris)

## **ABSTRAK**

Perkembangan internet telah memungkinkan orang untuk membaca berbagai macam berita dari berbagai belahan dunia. Tentu karena semakin bertambah jumlah berita, semakin banyak berita yang terlalu panjang untuk dibaca dan ditulis dengan bahasa selain Bahasa Indonesia sehingga tidak dapat memahami arti dari berita tersebut. Penelitian ini akan meneliti penerapan arsitektur Recurrent Neural Network (RNN) untuk melakukan peringkasan dan penerjemahan teks berita. Penelitian ini fokus pada menerjemahkan Bahasa Inggris ke Bahasa Indonesia. Hasil yang diperoleh merupakan sebuah API yang dapat melakukan peringkasan dan penerjemahan teks berita dari Bahasa Indonesia ke Bahasa Inggris dengan nilai akurasi yang cukup tinggi dan performa cepat.

Kata Kunci: berita, peringkasan, penerjemahan, rnn.

## **ABSTRACT**

*The development of the internet has allowed people to read different kinds of news from various parts of the globe. Of course, because the number of news keeps increasing, more news articles are becoming too long in length to read and written in languages different from Indonesian making the readers unable to grasp the meaning of the news. This research will apply Recurrent Neural Network (RNN) to summarize and translate English news to the Indonesian language. The expected result is two RNN model in which one will summarize the text in English and one will translate said text to Indonesian.*

*Keyword: news, summarization, translation, rnn.*

## **KATA PENGANTAR**

Saat ini, perkembangan teknologi dalam bidang pengolahan bahasa alami semakin pesat. Salah satu bidang yang menarik perhatian adalah sistem peringkasan dan penerjemahan teks, terutama dalam konteks terjemahan teks berita dari bahasa Inggris ke bahasa Indonesia. Terjemahan yang akurat dan efisien dari teks berita asing menjadi bahasa lokal sangat penting untuk menghubungkan komunikasi lintas budaya dan meningkatkan aksesibilitas informasi.

Dalam rangka mencapai tujuan tersebut, penelitian ini bertujuan untuk menerapkan pendekatan menggunakan Recurrent Neural Network (RNN) dalam pengembangan sistem peringkasan dan penerjemahan teks berita dari bahasa Inggris ke bahasa Indonesia. RNN adalah sebuah model *neural network* yang memiliki kemampuan untuk mengolah dan memahami konteks teks, sehingga diharapkan dapat meningkatkan kualitas terjemahan.

Melalui penelitian ini, diharapkan dapat ditemukan solusi yang efektif dalam meningkatkan kualitas terjemahan teks berita, sehingga mendorong pemahaman yang lebih baik tentang berita internasional dan memfasilitasi akses informasi yang lebih luas di antara masyarakat yang berbicara bahasa Indonesia.

Penulis menyadari bahwa penelitian ini memiliki batasan-batasan tertentu, seperti keterbatasan perangkat yang ada dan kemungkinan adanya kesalahan terjemahan yang dihasilkan oleh sistem. Namun, penulis optimis bahwa hasil penelitian ini akan memberikan kontribusi positif dalam mengembangkan bidang pengolahan bahasa alami, terutama dalam konteks terjemahan teks berita.

Akhir kata, penulis ingin mengucapkan terima kasih kepada beberapa pihak yang telah membantu pengerjaan skripsi melalui arahan dan dukungan yang diberikan:

1. Keluarga dan Kedua almarhum orang tua yang selalu memberi dukungan dan rasa semangat dalam proses pengerjaan skripsi
2. Bapak Ing. Artambo Benjamin Pangaribuan, M.Eng. dan Bapak Hamonangan Kinantan Prabu, M.T sebagai Dosen Pembimbing Pembimbing Akademik.
3. Bapak Rio Wirawan, S.Kom, MMSI sebagai Dosen Pembimbing Skripsi
4. Para sahabat Pejuang dan UNO sebagai kawan seperjuangan skripsi dan sumber dukungan mental
5. Bang Sam dari Brew Inc sebagai tempat nongkrong dan bercerita.
6. Eiichiro Oda yang ceritanya telah menginspirasi penulis untuk terus menyelesaikan penelitian ini

## DAFTAR ISI

PENERAPAN <i>RECURRENT NEURAL NETWORK</i> UNTUK SISTEM PERINGKASAN DAN PENERJEMAHAN TEKS BERITA BAHASA INGGRIS KE BAHASA INDONESIA .....	i
LEMBAR PENGESAHAN .....	ii
PERNYATAAN ORISINALITAS .....	iii
ABSTRAK .....	iv
<i>ABSTRACT</i> .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	viii
DAFTAR GAMBAR .....	x
DAFTAR TABEL .....	xi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat Penelitian .....	2
1.5 Batasan Masalah .....	3
1.6 Luaran yang diharapkan .....	3
1.7 Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA .....	5
2.1 Machine Learning .....	5
2.2 Deep Learning .....	5
2.3 Recurrent Neural Network .....	6
2.4 Long Short Term Memory .....	8
2.5 Natural Language Processing .....	9
2.6 Text Summarization .....	9
2.7 Machine Translation .....	10
2.8 Tensorflow .....	10
2.9 Text Preprocessing .....	11
2.10 Confusion Matrix .....	12
2.11 Recall-Oriented Understudy for Gisting Evaluation .....	14
2.12 <i>Encoder-Decoder</i> Model .....	15
2.13 Google Colaboratory .....	15



2.14	Flask .....	16
2.15	Studi Literatur.....	16
BAB III METODOLOGI PENELITIAN.....		19
3.1	Kerangka Pikir.....	19
3.2	Identifikasi Masalah .....	19
3.3	Studi Literatur.....	19
3.4	Akuisisi Data .....	20
3.5	Preprocessing Data .....	21
3.6	Pembuatan Model LSTM Peringkasan dan Penerjemahan .....	21
3.7	Pengujian dan Evaluasi Kedua Model LSTM.....	22
3.8	Pembuatan API dengan Kedua Model .....	22
3.9	Pengujian API.....	23
3.10	Perangkat Penelitian .....	23
BAB IV HASIL DAN PEMBAHASAN .....		24
4.1.	Arsitektur Model dan API .....	24
4.2.	Pengumpulan Data .....	25
4.3.	<i>Preprocessing</i> Data .....	28
4.4.	Pembuatan Model.....	35
4.5.	Pelatihan Model.....	38
4.6.	Evaluasi Model.....	39
4.7.	Pembuatan Model <i>Inference</i> .....	42
4.8.	Pengujian Model <i>Inference</i> .....	45
4.9.	Penilaian Model <i>Inference</i> .....	48
4.10.	Pembuatan API Peringkasan-Penerjemahan.....	51
BAB V PENUTUP.....		54
5.1	Kesimpulan.....	54
5.2	Saran .....	55
DAFTAR PUSTAKA .....		56
RIWAYAT HIDUP.....		58
LAMPIRAN.....		59

## DAFTAR GAMBAR

Gambar 2. 1	Arsitektur Deep Neural Network .....	6
Gambar 2. 2	Arsitektur LSTM .....	8
Gambar 2. 3	Confusion Matrix .....	12
Gambar 3. 1	Kerangka Pikir.....	19
Gambar 3. 2	Alur Pengujian.....	22
Gambar 4. 1	Arsitektur Model .....	24
Gambar 4. 2	Alur API .....	25
Gambar 4. 3	Pengambilan Data Gigaword .....	26
Gambar 4. 4	Pengambilan Data CCMatrix .....	27
Gambar 4. 5	Fungsi Preprocess.....	29
Gambar 4. 6	Penerapan Fungsi ke Gigaword.....	29
Gambar 4. 7	Penerapan Fungsi ke CCMatrix .....	30
Gambar 4. 8	Drop Null CCMatrix .....	30
Gambar 4. 9	Jumlah Kata Gigaword.....	31
Gambar 4. 11	Hapus Data Outlier CCMatrix.....	32
Gambar 4. 10	Jumlah Kata Awal CCMatrix .....	32
Gambar 4. 12	Jumlah Kata Setelah CCMatrix.....	33
Gambar 4. 13	Kode Tokenizer .....	34
Gambar 4. 14	Encoder Model .....	35
Gambar 4. 15	Decoder Model.....	36
Gambar 4. 16	Pembangunan Model.....	37
Gambar 4. 17	Arsitektur Model .....	37
Gambar 4. 18	Pelatihan Model.....	38
Gambar 4. 19	Graf Accuracy Peringkasan.....	39
Gambar 4. 20	Graf Loss Peringkasan .....	40
Gambar 4. 21	Graf Accuracy Penerjemahan.....	41
Gambar 4. 22	Graf Loss Penerjemahan .....	42
Gambar 4. 23	Model Inference .....	43
Gambar 4. 24	Inference Encoder .....	44
Gambar 4. 25	Inference Decoder .....	45
Gambar 4. 26	Fungsi Inference Teks .....	46

Gambar 4. 27 Sampel ROUGE Peringkasan .....	49
Gambar 4. 28 Rata-rata ROUGE Peringkasan .....	49
Gambar 4. 29 Sampel ROUGE Penerjemahan .....	50
Gambar 4. 30 Rata-rata ROUGE Penerjemahan.....	50
Gambar 4. 31 API .....	51
Gambar 4. 32 Bentuk Input API .....	51
Gambar 4. 33 Hasil API.....	52

### **DAFTAR TABEL**

Tabel 4. 1 Sampel Data Gigaword .....	26
Tabel 4. 2 Sampel Data CCMatrix .....	27
Tabel 4. 3 Data Split.....	33
Tabel 4. 4 Jumlah Kata Tokenizer .....	34
Tabel 4. 5 Contoh Tokenization.....	35
Tabel 4. 6 Contoh Hasil Ringkasan.....	47
Tabel 4. 7 Contoh Hasil Penerjemahan.....	48
Tabel 4. 8 Input dan Output API.....	52