

**Klasifikasi Berita Bahasa Indonesia dengan Recurrent Neural  
Network (RNN)**

**Laporan Tugas Akhir**

Diajukan Untuk Memenuhi  
Persyaratan Guna Meraih Gelar Sarjana  
Informatika Universitas Muhammadiyah Malang



Muh. Ridha Agam  
202010370311035

**Bidang Minat**  
Sains Data

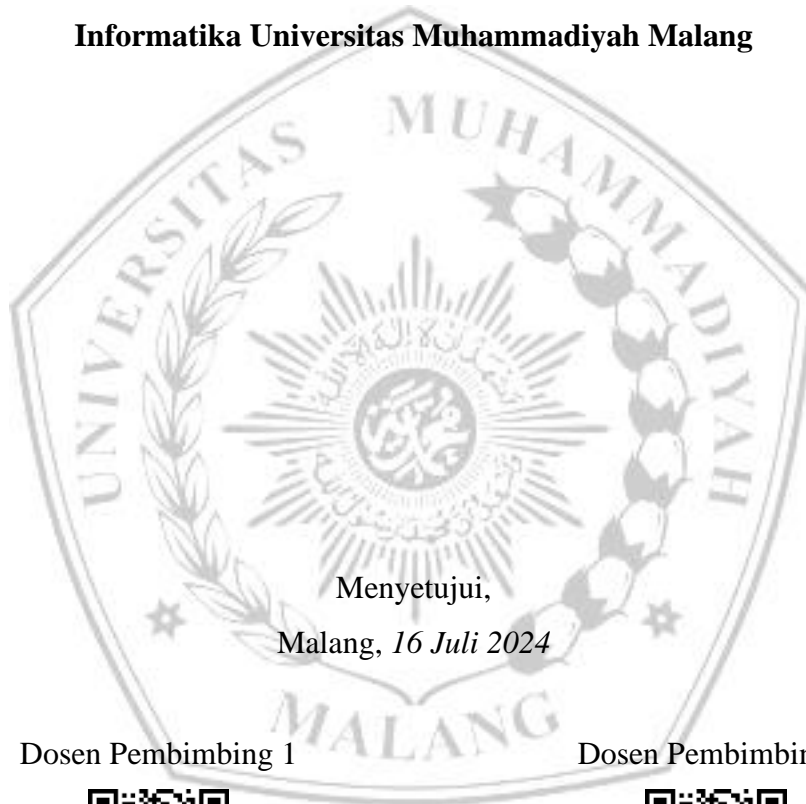
**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MALANG  
2024**

# LEMBAR PERSETUJUAN

## Klasifikasi Berita Bahasa Indonesia dengan Recurrent Neural Network (RNN)

### TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang



Menyetujui,  
Malang, 16 Juli 2024

Dosen Pembimbing 1



Christian Sri Kusuma Aditya S.Kom.,

M.Kom

NIP. 180327021991PNS.

Dosen Pembimbing 2



Ir. Galih Wasis Wicaksono S.kom.

M.Cs.

NIP. 10814100541PNS.

## LEMBAR PENGESAHAN

### Klasifikasi Berita Bahasa Indonesia dengan Recurrent Neural Network (RNN)

### TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

**MUH. RIDHA AGAM**

**202010370311035**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian  
pada tanggal 16 Juli 2024

Menyetujui,

Dosen Penguji 1



Dosen Penguji 2



Vinna Rahmayanti S.Si., M.Si

NIP. 180306071990PNS.

Ir. Yufis Azhar S.Kom., M.Kom.

NIP. 10814100544PNS.

Mengetahui,  
Ketua Jurusan Informatika



Ir. Galih Wasis Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.



## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA : MUH. RIDHA AGAM**

**NIM : 202010370311035**

**FAK./JUR. : Informatika**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “Klasifikasi Berita Bahasa Indonesia dengan Recurrent Neural Network (RNN)” beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,  
Dosen Pembimbing

Malang, 16 Juli 2024  
Yang Membuat Pernyataan



Christian Sri Kusuma Aditya S.Kom.,  
M.Kom

MUH. RIDHA AGAM



## ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model klasifikasi dokumen berita berbahasa Indonesia menggunakan *Recurrent Neural Network* (RNN). Dengan kemajuan teknologi informasi dalam satu dekade terakhir, media online telah menjadi sumber informasi utama bagi masyarakat, yang mengakibatkan perlunya mekanisme yang efisien untuk mengelola dan mengklasifikasikan konten berita yang terus bertambah setiap hari. Teknologi pembelajaran mesin, khususnya RNN, menjadi sangat penting dalam mengklasifikasikan dokumen berita berbahasa Indonesia karena kemampuannya dalam mempelajari ketergantungan waktu pada data sekuensial. Namun, RNN seringkali mendapatkan akurasi yang rendah dan cenderung mengalami *overfitting*. Hal ini disebabkan oleh kondisi data sampel yang tidak terkontrol dan jumlah dataset yang terbatas. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan akurasi RNN melalui augmentasi data untuk menurunkan kemungkinan *overfitting*. Penelitian ini melibatkan beberapa tahapan utama, yaitu pengumpulan data melalui teknik *scraping* dari portal berita Kompas.com, *preprocessing* teks yang mencakup pembersihan, penyaringan, *stemming*, dan tokenisasi, augmentasi data menggunakan *Back Translation*, serta pelatihan model RNN. Augmentasi data dilakukan untuk mengatasi ketidakseimbangan kelas dalam dataset dan meningkatkan akurasi model. Augmentasi dengan *Back Translation* menghasilkan variasi data yang lebih kaya dan membantu model dalam mengenali pola-pola yang lebih beragam, sehingga mengurangi *overfitting* dan meningkatkan kualitas data secara keseluruhan. Hasil penelitian menunjukkan bahwa model RNN yang dilatih dengan augmentasi data menghasilkan peningkatan kinerja yang signifikan dibandingkan dengan metode konvensional seperti *Naïve Bayes Classifier*. Model RNN tanpa augmentasi mencapai akurasi 92%, sementara model RNN dengan augmentasi data mencapai akurasi 96%. Peningkatan ini juga terlihat pada metrik evaluasi lainnya seperti *precision*, *recall*, dan *F1-Score*. Penelitian ini tidak hanya memperkuat literatur dalam bidang klasifikasi teks tetapi juga membuka peluang untuk aplikasi teknik *deep learning* dalam berbagai aspek penelitian dan pengembangan sistem informasi berbasis teks. Teknik augmentasi yang digunakan menunjukkan kontribusi signifikan dalam membantu model RNN untuk mengidentifikasi dan mengklasifikasikan kategori berita dengan lebih tepat dan efisien. Selain itu, penelitian ini diharapkan dapat memberikan kontribusi signifikan bagi pengembangan sistem otomatisasi klasifikasi berita yang lebih canggih dan adaptif terhadap kebutuhan informasi masyarakat Indonesia.

**Kata Kunci** : Recurrent Neural Network, Back Translation, Augmentasi Data, Klasifikasi Teks, Klasifikasi Berita.

## ABSTRACT

*This study aims to develop a classification model for Indonesian news documents using Recurrent Neural Network (RNN). With the advancement of information technology over the past decade, online media has become the primary source of information for society, necessitating efficient mechanisms to manage and classify the ever-growing content. Machine learning technology, especially RNN, is crucial for classifying Indonesian news documents due to its ability to learn temporal dependencies in sequential data. However, RNNs often achieve low accuracy and tend to overfit. This issue arises from uncontrolled sample data conditions and limited dataset sizes. Therefore, this study aims to improve RNN accuracy through data augmentation to reduce the likelihood of overfitting. This research involves several main stages, including data collection through scraping techniques from the Kompas.com news portal, text preprocessing that includes cleaning, filtering, stemming, and tokenization, data augmentation using Back Translation, and RNN model training. Data augmentation is performed to address class imbalance in the dataset and enhance model accuracy. Augmentation with Back Translation produces richer data variations and helps the model recognize more diverse patterns, thus reducing overfitting and improving overall data quality. The results show that the RNN model trained with data augmentation significantly improves performance compared to conventional methods like the Naïve Bayes Classifier. The RNN model without augmentation achieved an accuracy of 92%, while the RNN model with data augmentation achieved an accuracy of 96%. This improvement is also reflected in other evaluation metrics such as precision, recall, and F1-Score. This research not only strengthens the literature in the field of text classification but also opens up opportunities for applying deep learning techniques in various aspects of text-based research and information system development. The augmentation technique used shows significant contributions in helping the RNN model to more accurately and efficiently identify and classify news categories. Furthermore, this study is expected to make a significant contribution to the development of more advanced and adaptive automated news classification systems.*

**Keywords:** *Recurrent Neural Network, Back Translation, Data Augmentation, Text Classification, News Classification.*

## LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang telah memberikan rahmat dan ridho-Nya sehingga penulis dapat menyelesaikan Tugas Akhir.
2. Kedua orang tua saya yang selalu mendoakan untuk kebaikan anaknya, dan selalu memberikan dukungan dalam bentuk apapun.
3. Bapak Christian Sri Kusuma Aditya, S.Kom., M.Kom dan Bapak Galih Wasis Wicaksono, S.Kom., M.Cs. selaku dosen pembimbing pada tugas akhir saya.
4. Seluruh teman kontrakkan yang telah membantu dan menemani saya selama proses pembuatan skripsi.
5. Seluruh rekan asisten laboratorium informatika yang membantu dalam meningkatkan kemampuan akademis dan saling berkembang bersama di lingkungan laboratorium informatika.
6. Seluruh rekan kontrakan dan kos selama masa perkuliahan yang telah menemani dari awal perkuliahan.
7. Seluruh teman-teman penulis yang tidak dapat disebutkan satu persatu, yang telah memberikan motivasi, dukungan, serta semangat terhadap penulis.

## KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT. Atas limpahan rahmat dan hidayah-Nya sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul:

### **“Klasifikasi Berita Bahasa Indonesia dengan Recurrent Neural Network (RNN)”**

Tulisan ini mencakup beberapa topik utama, yaitu pendahuluan, tinjauan pustaka, metodologi penelitian, hasil, dan pembahasan. Kesimpulan diberikan berdasarkan hasil yang diperoleh selama proses penelitian. Peneliti sepenuhnya menyadari bahwa dalam penulisan Tugas Akhir ini masih terdapat banyak kekurangan dan keterbatasan. Oleh karena itu, peneliti mengharapkan saran yang konstruktif agar tulisan ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan.

Malang, 30 Juni 2024



Muh. Ridha Agam



## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>LEMBAR PERSETUJUAN .....</b>	<b>ii</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>iii</b>
<b>ABSTRAK .....</b>	<b>iv</b>
<b><i>ABSTRACT</i> .....</b>	<b>v</b>
<b>LEMBAR PERSEMBAHAN .....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR GAMBAR .....</b>	<b>xi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian .....	3
1.4. Batasan Masalah .....	3
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>5</b>
2.1. Penelitian Terdahulu .....	5
2.2. Teks Berita.....	6
2.3. <i>Deep Learning</i> .....	7
2.4. <i>Recurrent Neural Network (RNN)</i> .....	7
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>9</b>
3.1. Pengumpulan Data.....	9
3.2. Tahap Preprocessing .....	10
3.3. RNN.....	13
3.4. Evaluasi.....	14
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>15</b>
4.1. Evaluasi Perbandingan Data RNN.....	15
4.2. Perbandingan Hasil Pelatihan Sebelumnya dengan Neural Network.....	15
4.3. Hasil Implementasi Neural Network .....	17
4.3.1. Peningkatan Variasi Data.....	17

4.3.2. Peningkatan Metrik Evaluasi .....	18
<b>BAB V KESIMPULAN.....</b>	<b>19</b>
5.1. Kesimpulan .....	19
5.2. Saran .....	19
<b>DAFTAR PUSTAKA.....</b>	<b>21</b>
<b>LAMPIRAN.....</b>	<b>23</b>



## DAFTAR TABEL

Tabel 1.	Proses Pembersihan dan Penyaringan .....	10
Tabel 2.	Proses Stemming dan Tokenisasi .....	13
Tabel 3.	Kategori Dataset sebelum Augmentasi .....	12
Tabel 4.	Kategori Dataset Setelah Augmentasi .....	12
Tabel 5.	Hyperparameter Value .....	15
Tabel 6.	Perbandingan Performa Hasil Klasifikasi dengan Penelitian Sebelumnya .....	16





## DAFTAR GAMBAR

Gambar 1.	Model RNN .....	8
Gambar 2.	Alir Penelitian .....	9
Gambar 3.	Alur Augmentasi dengan <i>Back Translation</i> .....	11
Gambar 4.	Hasil Augmentasi Data dengan Back Translation .....	12
Gambar 5.	RNN unit .....	14
Gambar 6.	Confusion Matrix Dataset Tanpa Augmentasi.....	16
Gambar 7.	Confusion Matrix Data Augmentasi dengan <i>Back Translation</i> ...	17



## DAFTAR PUSTAKA

- [1] I. Y. R. Pratiwi, “Hoax news identification using machine learning model from online media in Bahasa Indonesia,” *MATRIX: Jurnal Manajemen Teknologi dan Informatika*, vol. 12, no. 2, pp. 58–67, Jul. 2022, doi: 10.31940/matrix.v12i2.58-67.
- [2] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, “Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations,” *Organ Res Methods*, vol. 25, no. 1, pp. 114–146, Jan. 2022, doi: 10.1177/1094428120971683.
- [3] A. Arif Siswandi, Y. Permana, and A. Emarilis, “Stemming Analysis Indonesian Language News Text with Porter Algorithm,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1845/1/012019.
- [4] M. I. Utari and H. Medyawati, “Classification Of News Types By Implementing Enhanced Confix Stripping Stemmer,” *International Journal of Engineering Technologies and Management Research*, vol. 6, no. 5, pp. 135–141, Mar. 2020, doi: 10.29121/ijetmr.v6.i5.2019.380.
- [5] K. Ivanedra and M. Mustikasari, “Implementasi Metode Recurrent Neural Network Pada Text Summarization Dengan Teknik Abstraktif The Implementation Of Text Summarization With Abstractive Techniques Using Recurrent Neural Network Method,” vol. 6, no. 4, pp. 377–382, 2019, doi: 10.25126/jtiik.201961067.
- [6] B. S. Prakoso, D. Rosiyadi, H. S. Utama, and D. Aridarma, “Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting,” *masa berlaku mulai*, vol. 1, no. 3, pp. 227–232, 2017.
- [7] A. Setiawan, L. W. Santoso, and R. Adipranata, “Klasifikasi Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier.”
- [8] I. Athiyyah Rahma and L. Hulliyyatus Suadaa, “Penerapan Text Augmentation Untuk Mengatasi Data Yang Tidak Seimbang Pada Klasifikasi Teks Berbahasa Indonesia Studi Kasus: Deteksi Judul Clickbait Dan Komentar Hate Speech Pada Berita Online”, doi: 10.25126/jtiik.2023107325.
- [9] R. Yunanto, A. P. Purfini, and A. Prabuwisesa, “Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning”, doi: 10.34010/jamika.v11i2.493.
- [10] A. D. Rendragraha, M. Arif Bijaksana, and A. Romadhony, “Pendekatan Metode Transformers untuk Deteksi Bahasa Kasar dalam Komentar Berita Online Indonesia.”

- [11] Nitha Kumala Dewi, “Identifikasi Berita Hoax dengan Menerapkan Algoritma Text Mining,” *Journal of Informatics, Electrical and Electronics Engineering*, vol. 2, no. 3, pp. 65–74, Mar. 2023, doi: 10.47065/jieee.v2i3.888.
- [12] O. Widarsih, ; Didi Yulistio, ; Agus Trianto, K. Bengkulu, and T. Bulan, “Analisis Tipe Teks Rekon (Struktur dan Aspek Kebahasaan) pada Teks Berita Media Center,” 2021. [Online]. Available: <https://mediacenter.bengkulukota.go.id>.
- [13] Y. Widhiyasana, T. Semiawan, I. Gibran, A. Mudzakir, and M. R. Noor, “Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia (Convolutional Long Short-Term Memory Implementation for Indonesian News Classification),” 2021.
- [14] N. Alvi Hasanah, Nanik Suciati, and Diana Purwitasari, “Pemantauan Perhatian Publik terhadap Pandemi COVID-19 melalui Klasifikasi Teks dengan Deep Learning,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 193–202, Feb. 2021, doi: 10.29207/resti.v5i1.2927.
- [15] I. N. Husada and H. Toba, “Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, Aug. 2020, doi: 10.28932/jutisi.v6i2.2743.
- [16] D. Christianto and H. Heryanto, “Pengenalan Emosi Dalam Teks Dengan Algoritme Long Short Term Memory.”
- [17] B. Nurdewanto, E. Sonalitha, A. Zubair, and S. Ratih Asriningtias, “Enhanced Confix Stripping Stemmer Dan Fuzzy Clustering Pada Data Pelaku Seni Di Malang Raya,” *Seminar Nasional Sistem Informasi*, 2019.
- [18] Natasya and A. S. Girsang, “Modified EDA and Backtranslation Augmentation in Deep Learning Models for Indonesian Aspect-Based Sentiment Analysis,” *Emerging Science Journal*, vol. 7, no. 1, pp. 256–272, Feb. 2023, doi: 10.28991/ESJ-2023-07-01-018.
- [19] I. Luthfi Rais, “Klasifikasi Data Kuesioner dengan Metode Recurrent Neural Network.”





# FAKULTAS TEKNIK

## INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id

UNIVERSITAS  
MUHAMMADIYAH  
MALANG

### FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Muh. Ridha Agam  
 NIM : 202010370311035  
 Judul TA : Klasifikasi Berita Bahasa Indonesia dengan Recurrent Neural Network (RNN)



#### Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	9%
2.	Bab 2 – Daftar Pustaka	25 %	20%
3.	Bab 3 – Analisis dan Perancangan	25 %	9%
4.	Bab 4 – Implementasi dan Pengujian	15 %	0%
5.	Bab 5 – Kesimpulan dan Saran	5 %	3%
6.	Makalah Tugas Akhir	20%	3%

*\*) Hasil cek plagiarisme diisi oleh pemeriksa (staf TU)*

*\*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)*

Mengetahui,  
 Pemeriksa (Staff TU)

  
 (.....)



Kampus I  
 Jl. Bandung 1 Malang, Jawa Timur  
 P. +62 341 551 253 (Hunting)  
 F. +62 341 460 435

Kampus II  
 Jl. Bendungan Sutami No 188 Malang, Jawa Timur  
 P. +62 341 551 149 (Hunting)  
 F. +62 341 582 060

Kampus III  
 Jl. Raya Tlogomas No 248 Malang, Jawa Timur  
 P. +62 341 464 318 (Hunting)  
 F. +62 341 460 435  
 E. webmaster@umm.ac.id