

**Augmentasi Data untuk Deteksi Berita Palsu menggunakan  
Metode *Convolutional Neural Network* pada Berita Bahasa  
Indonesia**

**Tesis**  
**untuk memenuhi sebagian persyaratan**  
**mencapai derajat Sarjana S-2 Program Studi**  
**Magister Sistem Informasi**



**Atik Zilziana Muflihati Noor**

**30000320420054**

**SEKOLAH PASCASARJANA**  
**UNIVERSITAS DIPONEGORO**  
**SEMARANG**

**2023**

HALAMAN PENGESAHAN

TESIS

**Augmentasi Data untuk Deteksi Berita Palsu menggunakan Metode  
Convolutional Neural Network pada Berita Bahasa Indonesia**

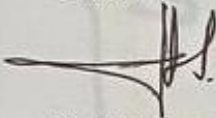
Oleh:

**Atik Zilziana Muflihati Noor**  
30000320420054

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 16 Agustus 2023 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 16 Agustus 2023  
Mengetahui,

**Penguji I**



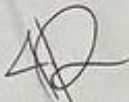
Dr. Adian Fatchur Rochim, S.T., M.T.  
NIP. 197302261998021001

**Penguji II**



Dr. Budi Warsito, S.Si., M.Si.  
NIP : 197508241999031003

**Pembimbing I**



Dr. Rahmat Gernowo, M.Si.  
NIP: 196511231994031003

**Pembimbing II**



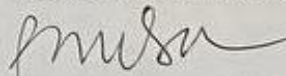
Dr. Oky Dwi Nurhayati, S.T., M.T.  
NIP 197910022009122001

**Mengetahui:**  
**Dekan Sekolah Pascasarjana**  
**Universitas Diponegoro**



Dr. R.B. Sanjoto, S.H., M.Hum  
NIP 196701011991031005

**Ketua Program Studi**  
**Magister Sistem Informasi**



Drs. Bayu Surarso, M.Sc., Ph.D.  
NIP 196311051988031001

## PERNYATAAN PERSETUJUAN

### PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini :

Nama : Atik Zilziana Muflihati Noor  
NIM : 30000320420054  
Program Studi : Magister Sistem Informasi  
Program : Sekolah Pascasarjana  
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul :

**Augmentasi Data untuk Deteksi Berita Palsu menggunakan  
Metode *Convolutional Neural Network* pada Berita Bahasa  
Indonesia**

beserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang

Pada tanggal : 16 Agustus 2023

SEKOLAH PASCASARJANA  
Yang menyatakan



Atik Zilziana Muflihati Noor

NIM. 30000320420054

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 16 Agustus 2023



Atik Zilziana Muflihati Noor

SEKOLAH PASCASARJANA

## KATA PENGANTAR

Segala puji syukur bagi Tuhan Yang Maha Esa yang telah memberikan Rahmat dan Karunia-Nya sehingga penulis mampu menyelesaikan penelitian tesis yang berjudul “Augmentasi Data untuk Deteksi Berita Palsu menggunakan Metode *Convolutional Neural Network* pada Berita Bahasa Indonesia”. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister S-2 Sistem Informasi pada Sekolah Pascasarjana Universitas Diponegoro. Dalam pelaksanaan serta penyusunan tesis ini, tidak terlepas dukungan, bimbingan dan bantuan dari berbagai pihak sehingga akhirnya skripsi ini dapat diselesaikan. Oleh karena itu, melalui kesempatan ini penulis ingin menyampaikan ucapan terimakasih kepada:

1. Bapak Prof. Dr. Yos Johan Utama, S.H, M.Hum selaku Rektor Universitas Diponegoro,
2. Bapak Dr. R.B. Sularto, S.H., M.Hum, selaku Dekan Sekolah Pascasarjana Universitas Diponegoro,
3. Bapak Drs.Bayu Surarso, M.Sc., Ph.D, selaku Ketua Program Studi Magister Sistem Informasi Universitas Diponegoro,
4. Bapak Dr. Rahmat Gernowo, M.Si., selaku Dosen Pembimbing I,
5. Ibu Dr. Oky Dwi Nurhayati, S.T., M.T., selaku Dosen Pembimbing II,

Penulis menyadari bahwa laporan ini masih banyak kekurangan yang disebabkan oleh keterbatasan kemampuan dan pengetahuan penulis. Sehingga penulis sangat mengharapkan kritik dan saran. Semoga laporan ini dapat memberikan manfaat bagi semua.

Semarang, 13 Agustus 2023

SEKOLAH PASCASARJANA

Penulis

## DAFTAR ISI

Halaman

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>i</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI .....</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN NON ROYALTI BERMATERAI .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>DAFTAR ISI.....</b>	<b>v</b>
<b>DAFTAR GAMBAR.....</b>	<b>vii</b>
<b>DAFTAR TABEL .....</b>	<b>viii</b>
<b>ABSTRAK .....</b>	<b>x</b>
<b>ABSTRACT .....</b>	<b>xi</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
<b>1.1 Latar Belakang.....</b>	<b>1</b>
<b>1.2 Tujuan Penelitian .....</b>	<b>2</b>
<b>1.3 Manfaat Penelitian .....</b>	<b>3</b>
<b>BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....</b>	<b>4</b>
<b>2.1 Tinjauan Pustaka .....</b>	<b>4</b>
<b>2.2 Landasan Teori.....</b>	<b>6</b>
2.2.1 Berita Palsu / Berita <i>Hoax</i> .....	6
2.2.2 Artificial Intelligian .....	9
2.2.3 <i>Machine Learning</i> .....	11
2.2.4 <i>Deep Learning</i> .....	12
2.2.5 Metode CNN.....	14
2.2.6 Operasi Konvolusi.....	15
2.2.7 Augmentasi Data.....	21
<b>BAB III METODE PENELITIAN .....</b>	<b>24</b>

<b>3.1</b>	<b>Cara Pendekatan Penelitian.....</b>	<b>24</b>
<b>3.2</b>	<b>Bahan dan peralatan eksplorasi digunakan .....</b>	<b>24</b>
<b>3.3</b>	<b>Prosedur Penelitian .....</b>	<b>25</b>
3.3.1	<i>Data Collecting</i> .....	26
3.3.2	<i>Preprocessing</i> .....	27
3.3.3	<i>Splitting Data</i> .....	29
3.3.4	Augmentasi .....	29
3.3.5	Metode CNN.....	30
3.3.6	<i>Performance Test</i> .....	36
3.3.7	<i>Output</i> .....	38
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>39</b>
<b>4.1</b>	<b>Hasil Penelitian.....</b>	<b>39</b>
4.1.1	<i>Preprocessing</i> .....	39
4.1.2	<i>Split Data</i> .....	49
<b>4.2</b>	<b>Hasil Augmentasi.....</b>	<b>50</b>
<b>4.3</b>	<b>Deteksi Menggunakan CNN.....</b>	<b>57</b>
<b>4.4</b>	<b>Evaluasi .....</b>	<b>77</b>
<b>4.5</b>	<b>Pembahasan .....</b>	<b>77</b>
4.5.1	<i>Preprocessing</i> .....	77
4.5.2	<i>Split Data</i> .....	78
4.5.3	Augmentasi Data.....	78
4.5.4	Deteksi Menggunakan CNN.....	79
4.5.5	Evaluasi .....	82
<b>BAB V PENUTUP.....</b>		<b>84</b>
<b>5.1</b>	<b>Kesimpulan .....</b>	<b>84</b>
<b>5.2</b>	<b>Saran.....</b>	<b>84</b>
<b>DAFTAR PUSTAKA.....</b>		<b>86</b>



## DAFTAR GAMBAR

	Halaman
Gambar 2.1 Perbedaan <i>machine learning</i> dengan deep learning (Nikoskinen, 2015).....	13
Gambar 2.2 Arsitektur CNN (Nugroho dkk., 2020).....	15
Gambar 2.3 Ilustrasi operasi konvolusi ( (Ilahiyah & Nilogiri, 2018)) .....	17
Gambar 2.4 Plot Distribusi Fungsi ReLU (Brownlee, 2019).....	18
Gambar 2.5 Ilustrasi operasi <i>pooling</i> .....	19
Gambar 2.6 Ilustrasi <i>Flattening</i> (Kurnia & Wibowo, 2021) .....	19
Gambar 2.7 Ilustrasi <i>Fully Connected Layer</i> (Kurnia & Wibowo, 2021) .....	21
Gambar 3.1 Rancangan Sistem menggunakan EDA .....	26
Gambar 3.2 <i>Feature Extraction</i> .....	31
Gambar 3.3 <i>Flowchart Convolutional Neural Network</i> (CNN).....	34
Gambar 4.1 Hasil <i>Preprocessing</i> .....	78
Gambar 4.2 Hasil Augmentasi.....	79
Gambar 4.3 Hasil Akurasi Algoritma CNN.....	80
Gambar 4.4 Hasil <i>Loss Function</i> Algoritma CNN.....	81
Gambar 4.5 Hasil <i>Confusion Matrix</i> dalam Tabel.....	82
Gambar 4.6 Hasil <i>Confusion Matrix</i> dalam Presentase .....	83

SEKOLAH PASCASARJANA



## DAFTAR TABEL

	Halaman
Tabel 2.1 Contoh Data Berita .....	7
Tabel 3.1 <i>Case folding</i> .....	27
Tabel 3.2 <i>Normalization</i> .....	28
Tabel 3.3 <i>Tokenizeing</i> .....	28
Tabel 3.4 <i>Stopword Removal</i> .....	28
Tabel 3.5 <i>Stemming</i> .....	29
Tabel 3.6 Contoh kalimat hasil EDA .....	29
Tabel 3.7 Parameter yang disarankan (Wei, Zou, 2019) .....	30
Tabel 3.8 <i>Confusion Matrix</i> .....	37
Tabel 4.1 Hasil <i>Case Folding</i> .....	39
Tabel 4.2 Hasil <i>Normalization</i> .....	41
Tabel 4.3 Hasil <i>Tokenizeing</i> .....	43
Tabel 4.4 Hasil <i>Filtering</i> .....	45
Tabel 4.5 Hasil <i>Stemming</i> .....	48
Tabel 4.6 Hasil pembagian data <i>Train</i> .....	49
Tabel 4.7 Hasil pembagian data <i>Test</i> .....	50
Tabel 4.8 Hasil <i>Synonym Replacement</i> .....	51
Tabel 4.9 Hasil <i>Random Insertion</i> .....	52
Tabel 4.10 Hasil <i>Random Swapping</i> .....	54
Tabel 4.11 Hasil <i>Random Deletion</i> .....	55
Tabel 4.12 Hasil Total Kata tiap Dokumen .....	58
Tabel 4.13 Hasil <i>Max Length</i> dan <i>Num Words</i> .....	59
Tabel 4.14 Parameter yang digunakan <i>Tokenize</i> .....	59
Tabel 4.15 Hasil <i>Tokenize</i> .....	59
Tabel 4.16 Hasil dari <i>Text To Sequences Data Train</i> .....	63
Tabel 4.17 Hasil dari <i>Text To Sequences Data Test</i> .....	63
Tabel 4.18 Hasil <i>Pad Sequences Data Train</i> .....	64

Tabel 4.19 Hasil <i>Pad Sequences Data Test</i> .....	66
Tabel 4.20 Hasil <i>Embedding Layer</i> .....	68
Tabel 4.21 Hasil <i>Embedding Layer</i> berdasarkan <i>Input</i> .....	72
Tabel 4.22 Hasil Conv1D .....	73
Tabel 4.23 Hasil Transformasi <i>Matrix</i> sebelum dilakukan aktivasi pada <i>Dense Layer</i> .....	74
Tabel 4.24 Perkalian Bobot Dense dengan Inputan Conv1D .....	75
Tabel 4.25 Total Dense dan Hasil Aktivasi .....	75
Tabel 4.26 Hasil Prediksi.....	76
Tabel 4.27 Tabel <i>Confusion Matrix</i> 2 Kelas.....	77
Tabel 4.28 Perhitungan <i>Confusion Matrix</i> .....	83



SEKOLAH PASCASARJANA

**AUGMENTASI DATA UNTUK DETEKSI BERITA PALSU MENGGUNAKAN  
METODE *CONVOLUTIONAL NEURAL NETWORK* PADA BERITA  
BAHASA INDONESIA**

**ABSTRAK**

Berita palsu masih menjadi masalah yang harus mendapat perhatian khusus. Berita palsu disebarkan dengan sengaja untuk menyesatkan pembaca agar mempercayai berita yang tidak valid. Beberapa pengguna menggunakan cara yang tidak etis dengan menyebarkan berita palsu dalam berbagai media seperti teks. Sehingga, informasi terbaru yang muncul di jaringan sosial *online* diragukan dan dalam banyak kasus, menyesatkan pengguna lain. Tujuan utama dari penelitian ini adalah merancang model yang memiliki kemampuan terbaik untuk melakukan tugas *teks classification* pada konteks berita bahasa Indonesia. Model ini diharapkan dapat digunakan untuk berkontribusi dalam menanggulangi masalah penyebaran berita palsu, khususnya di Indonesia. Metode Augmentasi Data dan *Convolutional Neural Network* (CNN) diusulkan untuk digunakan dalam melakukan penelitian. Augmentasi yang digunakan pada penelitian ini adalah *Easy Data Augmentasi* (EDA), EDA sendiri merupakan pendekatan tradisional yang langsung untuk meningkatkan volume data. Metode CNN dapat menjadi salah satu pilihan tepat untuk melakukan *teks classification* dengan kinerja klasifikasi yang baik dalam melakukan tugas klasifikasi teks. Model pada penelitian ini mampu menghasilkan akurasi sebesar 82.26%, dengan akurasi Recall 97.51% dan F1-Score sebesar 90.07%.

**Kata Kunci:** Augmentasi Data, Berita Palsu, CNN, Deteksi, Teknik EDA.

# ***DATA AUGMENTATION FOR FAKE NEWS DETECTION USING CONVOLUTIONAL NEURAL NETWORK METHODS ON***

## ***ABSTRACT***

*Fake news is still a problem that should receive special attention. Fake news is spread intentionally to mislead readers into believing invalid news. Some users use unethical means by spreading fake news in various media such as text. Thus, the latest information appearing on online social networks is doubtful and in many cases, misleading other users. The main objective of this research is to design a model that has the best ability to perform text classification tasks in the context of Indonesian language news. This model is expected to be used to contribute to tackling the problem of spreading fake news, especially in Indonesia. Data Augmentation and Convolutional Neural Network (CNN) methods are proposed to be used in conducting the research. The augmentation used in this research is Easy Data Augmentation (EDA), EDA is a traditional approach that directly increases the volume of data. CNN method can be one of the right choices to perform text classification with good classification performance in performing text classification tasks. The model in this study was able to produce an accuracy of 82.26%, with Recall accuracy of 97.51% and F1-Score of 90.07%.*

***Keywords:*** *Data Augmentation, Fake News, CNN, Detection, EDA Techniques.*

**SEKOLAH PASCASARJANA**