

**Pembangunan Model Pembelajaran Mesin untuk Identifikasi  
Struktur Kalimat Bahasa Indonesia**

**Tugas Akhir**

**Diajukan untuk Memenuhi Sebagian Persyaratan Mencapai  
Derajat Sarjana Teknik Informatika**



Disusun Oleh:

**Radian Satria Panigas**

13 07 07581

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ATMA JAYA YOGYAKARTA  
2018

# HALAMAN PENGESAHAN

Tugas Akhir Berjudul

**Pembangunan Model Pembelajaran Mesin untuk Identifikasi Struktur  
Kalimat Bahasa Indonesia**

**Disusun Oleh :**

Radian Satria Panigas (13 07 07581)

Dinyatakan Telah Memenuhi Syarat

Pada Tanggal : Oktober 2018

Oleh :

**Pembimbing I**

(Y. Sigit Purnomo W.P., S.T., M.Kom.)

**Pembimbing II**

(Dr. Ir. Alb. Joko Santoso, M.T.)

**Tim Penguji :**

**Penguji I**

(Y. Sigit Purnomo W.P., S.T., M.Kom.)

**Penguji II**

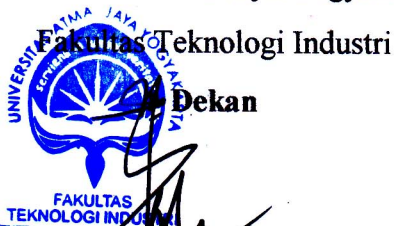
(Patricia Arsanari, S.Si, M.T.)

**Penguji III**

(Yulius Harjoseputro, S.T., M.T.)

Yogyakarta, Oktober 2018

Universitas Atma Jaya Yogyakarta



Dr. A. Teguh Siswanto

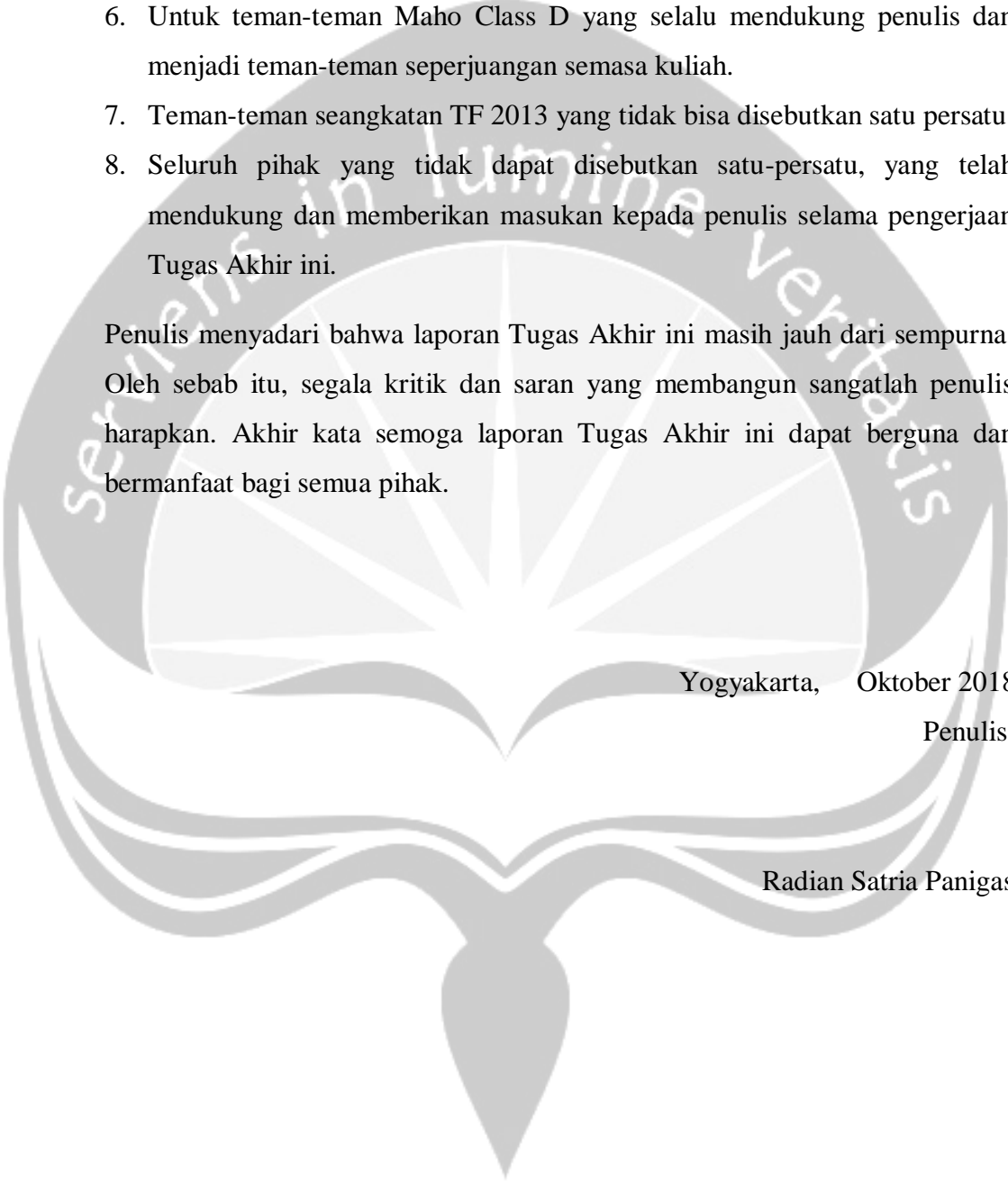
## KATA PENGANTAR

Penulis mengucapkan puji dan syukur kepada Tuhan Yang Maha Esa atas segala kasih dan karunia-Nya sehingga penulis dapat menyelesaikan pembuatan laporan tugas akhir ini dengan baik. Tujuan dari pembuatan laporan tugas akhir ini adalah sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.

Penulisan tugas akhir ini, penulis sebagai mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta mendapat banyak pengalaman dan ilmu-ilmu baru yang sebelumnya belum pernah didapat di kelas perkuliahan.

Dalam pengerjaan Tugas Akhir, penulis banyak mendapat bantuan dari berbagai pihak. Untuk itu pada kesempatan ini, penulis ingin menyampaikan ungkapan terima kasih kepada :

1. Keluarga tercinta yaitu bapak, ibu, kakak, adik yang selalu memberikan dukungan dalam bentuk apapun. Begitu juga dengan nasihat dan semangat yang selalu diberikan kepada penulis.
2. Bapak Y. Sigit Purnomo W.P., S.T., M.Kom., selaku dosen pembimbing I yang telah menerima penulis sebagai mahasiswa bimbingan Tugas Akhir. Begitu banyak masukan dan petunjuk yang telah diberikan sehingga Tugas Akhir ini dapat selesai dengan baik.
3. Bapak Dr. Ir. Alb. Joko Santoso, M.T. selaku dosen pembimbing II yang telah memberi banyak masukan dan saran agar penelitian berjalan dengan baik.
4. Seluruh dosen dan staf Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta yang telah membantu penulis selama kuliah.

- 
5. Untuk Jegar Sahaduta Ginting dan Pier Simak sebagai kawan seperjuangan dalam pengerjaan Tugas Akhir ini yang selalu memberi dukungan dan masukan kepada penulis.
  6. Untuk teman-teman Maho Class D yang selalu mendukung penulis dan menjadi teman-teman seperjuangan semasa kuliah.
  7. Teman-teman seangkatan TF 2013 yang tidak bisa disebutkan satu persatu.
  8. Seluruh pihak yang tidak dapat disebutkan satu-persatu, yang telah mendukung dan memberikan masukan kepada penulis selama pengerjaan Tugas Akhir ini.

Penulis menyadari bahwa laporan Tugas Akhir ini masih jauh dari sempurna. Oleh sebab itu, segala kritik dan saran yang membangun sangatlah penulis harapkan. Akhir kata semoga laporan Tugas Akhir ini dapat berguna dan bermanfaat bagi semua pihak.

Yogyakarta, Oktober 2018

Penulis,

Radian Satria Panigas

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>ii</b>
<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>DAFTAR ISI.....</b>	<b>v</b>
<b>DAFTAR GAMBAR.....</b>	<b>vii</b>
<b>DAFTAR TABEL .....</b>	<b>viii</b>
<b>DAFTAR PERSAMAAN.....</b>	<b>ix</b>
<b>INTISARI.....</b>	<b>x</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	3
1.5 Metodologi Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>5</b>
<b>BAB III LANDASAN TEORI .....</b>	<b>7</b>
3.1 Python.....	7
3.2 Natural Language Processing .....	7
3.3 Struktur Kalimat.....	8
3.4 Machine Learning .....	9
3.5 Deep Learning.....	10
3.6 Recurrent Neural Network.....	10
3.7 Long Short-Term Memory Network.....	11
3.8 Bidirectional LSTM .....	12
3.9 Aktivasi ReLU .....	12
3.10 Dropout.....	13
3.11 ADAM Optimizer .....	14
<b>BAB IV PERSIAPAN DATA DAN MODEL .....</b>	<b>15</b>
4.1 Persiapan Data .....	15
4.2 Alat Penelitian.....	16
4.3 Langkah-Langkah Persiapan Data dan Model.....	18
4.3.1 Pengumpulan Data.....	19
4.3.2 Perancangan Model .....	20
4.3.2.1 Pre-processing Data .....	23
4.3.2.2 Load Training Data.....	25
4.3.2.3 Load Embedding.....	26

4.3.2.3 Pembuatan model .....	26
4.3.2.5 Training Model .....	28
4.3.2.6 Penyimpanan Model .....	29
4.3.2.7 Load Model .....	29
4.3.2.8 Prediksi.....	30
4.3.3 Pengkodean Model .....	30
4.3.3.1 Kode Pembuatan Model.....	30
4.3.3.2 Kode Training Model.....	33
4.3.3.3 Kode Prediksi .....	34
4.3.4 Pengujian Model.....	35
<b>BAB V ANALISIS DAN PEMBAHASAN.....</b>	<b>36</b>
5.1 Pengujian .....	36
5.1.1 Hasil Pembelajaran dan Analisis .....	36
5.1.2 Hasil Pengujian dan Analisis .....	38
<b>BAB VI KESIMPULAN DAN SARAN.....</b>	<b>40</b>
6.1 Kesimpulan .....	40
6.2 Saran.....	40
<b>DAFTAR PUSTAKA.....</b>	<b>41</b>
<b>LAMPIRAN .....</b>	<b>43</b>

## DAFTAR GAMBAR

<i>Gambar 3. 1 Recurrent Neural Network (Olah,2015).....</i>	<i>10</i>
<i>Gambar 3. 2 RNN Dijabarkan terhadap Timestep (Olah, 2015).....</i>	<i>11</i>
<i>Gambar 3. 3 Cell LSTM dengan 4 Lapis Fungsi Aktivasi (Olah, 2015) .....</i>	<i>11</i>
<i>Gambar 3. 4 Arsitektur BiLSTM (Olah, 2015).....</i>	<i>12</i>
<i>Gambar 3. 5 Neural Network Sebelum Diaplikasikan Dropout .....</i>	<i>13</i>
<i>Gambar 3. 6 Neural Network Setelah Diaplikasikan Dropout. ....</i>	<i>14</i>
<i>Gambar 4. 1 Contoh IOB Tagging .....</i>	<i>15</i>
<i>Gambar 4. 2 Diagram Persiapan Data dan Model.....</i>	<i>18</i>
<i>Gambar 4. 3 Proses Anotasi dengan WebAnno. ....</i>	<i>19</i>
<i>Gambar 4. 4 Contoh Dataset Sesuai Format Penelitian .....</i>	<i>19</i>
<i>Gambar 4. 5 Alur Kerja Proses Pelatihan.....</i>	<i>22</i>
<i>Gambar 4. 6 Alur Kerja Proses Prediksi.....</i>	<i>23</i>
<i>Gambar 4. 7 Kode Preprocessing Input .....</i>	<i>24</i>
<i>Gambar 4. 8 Graph Model LSTM .....</i>	<i>27</i>
<i>Gambar 4. 9 Graph Model BiLSTM.....</i>	<i>28</i>
<i>Gambar 4. 10 Kode Inisialisasi Embedding .....</i>	<i>31</i>
<i>Gambar 4. 11 Kode Pembuatan Model .....</i>	<i>32</i>
<i>Gambar 4. 12 Kode untuk Memulai Training.....</i>	<i>33</i>
<i>Gambar 4. 13 Kode Prediksi.....</i>	<i>34</i>
<i>Gambar 5. 1 Proses Training Model.....</i>	<i>36</i>

## DAFTAR TABEL

<i>Tabel 4. 1 Contoh IOB Tagging</i> .....	16
<i>Tabel 4. 2 Jumlah kalimat dengan pola tertentu pada dataset.</i> .....	20
<i>Tabel 5. 1 Hasil proses uji data validasi model.</i> .....	37
<i>Tabel 5. 2 Hasil proses testing model.</i> .....	38





## DAFTAR PERSAMAAN

<i>Persamaan 3. 1 Fungsi Aktivasi ReLU.....</i>	<i>12</i>
---	-----------



## INTISARI

**Radian Satria Panigas (13 07 07581)**

Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Atma  
Jaya Yogyakarta  
Jl. Babarsari 43, Yogyakarta 55281

Penelitian dan penerapan *Natural Language Processing* dalam dunia teknologi semakin meningkat dan berkembang. Umumnya pengembangan NLP terjadi secara pesat dalam pemrosesan Bahasa Inggris. Metode – metode yang digunakan dalam proses *training* model NLP umumnya menggunakan algoritma *Machine Learning*. Salah satu penyebab berkembangnya NLP adalah meningkatnya jumlah data linguistik yang tersedia. Namun mayoritas sistem dan data yang tersedia saat ini adalah bagian dari *high-resource language* seperti Bahasa Inggris, Perancis, Jerman, Spanyol dan China.

Dalam pembangunan model Machine Learning ini, model akan digunakan untuk mengidentifikasi struktur kalimat dalam Bahasa Indonesia. Model akan dirancang berdasarkan beberapa referensi model NLP yang dipakai untuk pemrosesan Bahasa Inggris. Pembuatan model akan dirancang dengan Bahasa pemrograman Python. Data yang digunakan dikumpulkan dari berbagai macam sumber seperti Kompas, Historia, dan Beritagar. Data dibagi menjadi 2538 data latih, 725 data validasi dan 364 data tes.

Model terbaik yang dibuat memiliki skor F1 mencapai 81.57% pada data *testing*. Model dapat memprediksi struktur kalimat Bahasa Indonesia. Dari proses pengujian, dapat disimpulkan bahwa model mampu mengidentifikasi struktur kalimat Bahasa Indonesia dengan cukup baik.

**Kata Kunci:** *Natural Language Processing, Machine Learning*, Struktur Kalimat Bahasa Indonesia, Bahasa Indonesia.

