

**IMPLEMENTASI MODEL *NATURAL LANGUAGE  
PROCESSING* (NLP) PADA SISTEM REKOMENDASI  
PEKERJAAN MENGGUNAKAN *CONVOLUTIONAL NEURAL  
NETWORKS* (CNN) DAN *LONG SHORT TERM MEMORY*  
(LSTM)**

**SKRIPSI**

**Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer**



**OLEH :**

**MHD. HAIDIR FIKRI**

**09011282025064**

**JURUSAN SISTEM KOMPUTER  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA**

**2024**

# HALAMAN PENGESAHAN

## IMPLEMENTASI MODEL *NATURAL LANGUAGE PROCESSING* (NLP) PADA SISTEM REKOMENDASI PEKERJAAN MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORKS* (CNN) DAN *LONG SHORT* *TERM MEMORY* (LSTM)

### SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

OLEH :

**MHD. HAIDIR FIKRI**

**09011282025064**

Palembang, 17 Desember 2024

Mengetahui,

**Ketua Jurusan Sistem Komputer**



**Dr. Ir. Sukemi. M.T.**

**NIP. 196612032006041001**

**Pembimbing Tugas Akhir**

**Sutarno, S.T., M.T.**

**NIP. 197811012010121003**

## AUTHENTICATION PAGE

**IMPLEMENTASI MODEL *NATURAL LANGUAGE PROCESSING* (NLP)  
PADA SISTEM REKOMENDASI PEKERJAAN MENGGUNAKAN  
*CONVOLUTIONAL NEURAL NETWORKS* (CNN) DAN *LONG SHORT  
TERM MEMORY* (LSTM)**

**SKRIPSI**

**Submitted To Complete One Of The Requirements For  
Obtaining A Bachelor's Degree in Computer Science**

**By :**

**MHD. HAIDIR FIKRI**

**09011282025064**

Palembang,  Desember 2024

Acknowledge,

**Head Of Computer System  
Departement**



**Dr. Ir. Sukemi. M.T.**

**NIP. 196612032006041001**

**Final Project Advisor**



**Sutarno, S.T., M.T.**

**NIP. 197811012010121003**

**HALAMAN PERSETUJUAN**

**Telah diuji dan lulus pada :**

**Hari : Jumat**

**Tanggal : 15 November 2024**

**Tim Penguji :**

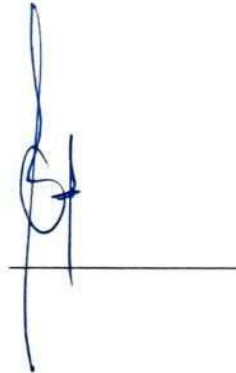
**1. Ketua : Huda Ubaya, S.T., M.T.**



**2. Penguji : Dr. Rossi Passarella, M.Eng.**



**3. Pembimbing : Sutarno, S.T., M.T.**



**Mengetahui,**   
**Ketua Jurusan Sistem Komputer**



**Dr. Ir. Sukemi, M.T.**  
**NIP. 196612032006041001**

## HALAMAN PERNYATAAN

Yang Bertanda Tangan Di Bawah Ini :

Nama : Mhd. Haidir Fikri  
NIM : 09011282025064  
Judul : IMPLEMENTASI MODEL *NATURAL LANGUAGE PROCESSING* (NLP) PADA SISTEM REKOMENDASI PEKERJAAN MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORKS* (CNN) DAN *LONG SHORT TERM MEMORY* (LSTM)

### Hasil Pengecekkam Software iThenticate/Turnitin : 2%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya saya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.

Palembang, 21 Desember 2024

Penulis,



Mhd. Haidir Fikri

NIM.09011282025064

## KATA PENGANTAR

Puji syukur Penulis ucapkan kehadirat Allah SWT yang telah memberikan rahmat-nya sehingga penulis dapat menyelesaikan Laporan Tugas Akhir dengan berjudul “Implementasi Model *Natural Language Processing* (NLP) Pada Sistem Rekomendasi Pekerjaan Menggunakan *Convolutional Neural Networks* (CNN) Dan *Long Short Term Memory* (LSTM)”. tepat pada waktunya.

Penulisan Laporan Tugas Akhir ini tidak terlepas dari peran serta semua pihak yang telah membantu, memberikan motivasi, mengingatkan, membimbing, dorongan, semangat, kritik dan saran. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Orang Tua Dan Keluarga Yang Selalu Mendoakan Dan Memberikan Dukungan Moral Maupun Finansial .
2. Bapak Sutarno, S.T.,M.T. selaku Dosen Pembimbing Tugas Akhir
3. Bapak Rahmat Fadli Isnanto, S.Si. M.SC. selaku Dosen Pembimbing Akademik
4. Bapak Prof. Dr. Erwin, S.Si., M.Si. Selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Samsuryadi, M.Kom., Ph.D. selaku Wakil Dekan Bidang Akademik Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Bapak Dr. Ir. Sukemi., M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Bapak Huda Ubaya, S.T., M.T. selaku Sekretaris Jurusan Sistem Komputer.
8. Bapak Angga selaku Admin Jurusan Sistem Komputer yang telah membantu penulis dalam hal-hal administrasi.
9. Teman-teman seperjuangan Jurusan Sistem Komputer Angkatan 2020 , yang sudah menjadi support system saya.
10. Sahabat Karib yaitu : ridho, osama, sugik, dwi dan lain lain yang telah mengisi kekosong dan menghilangkan pusing ketika mengerjakan skripsi .

11. Teman sepembimbing : edwin dan aca yang telah bersama sama berjuang untuk dapat menyelesaikan skripsi.
12. Seluruh pihak yang tidak dapat penulis sebutkan satu persatu, yang telah memberikan semangat serta do'a.
13. Teman-teman seperjuangan Jurusan Sistem Komputer seperti : nurul, liya, indra, edwin, dan lain lain yang telah membantu dari awal perkuliahan hingga sekarang .

Penulis secara penuh menyadari bahwa Laporan Tugas Akhir ini masih sangat jauh dari kata sempurna, oleh karena itu kritik dan saran akan sangat berguna bagi penulis. Semoga apa yang telah penulis kerjakan ini dapat memberikan manfaat kepada para pembaca sehingga dapat memberikan kontribusi dalam meningkatkan mutu pembelajaran maupun penelitian kedepannya.

Palembang, 31 Desember  
2024  
Penulis



**Mhd. Haidir Fikri**  
**NIM. 09011282025064**



**IMPLEMENTASI MODEL *NATURAL LANGUAGE PROCESSING* (NLP)  
PADA SISTEM REKOMENDASI PEKERJAAN MENGGUNAKAN  
*CONVOLUTIONAL NEURAL NETWORKS* ( CNN) DAN *LONG SHORT  
TERM MEMORY* (LSTM)**

**Mhd. Haidir Fikri**

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya  
Email : [haidirfikri13@gmail.com](mailto:haidirfikri13@gmail.com)

**ABSTRAK**


Kemajuan teknologi menghadirkan tantangan dalam mencocokkan keterampilan individu dengan rekomendasi pekerjaan yang sesuai. Penelitian ini mengembangkan sistem rekomendasi pekerjaan berbasis *Natural Language Processing* (NLP) menggunakan model *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM). Data penelitian diperoleh dari dataset publik *Hugging Face* dan *Kaggle*, yang mencakup deskripsi pengalaman kerja dan keterampilan. Data diproses melalui tahapan pra pengolahan seperti penghapusan *stopword*, *regex*, *stemming*, *tokenisasi*, *padding*, serta teknik *balancing* seperti *SMOTE* dan *undersampling*. Dataset dibagi untuk pelatihan dan pengujian model, dengan performa diukur melalui akurasi, *loss*, dan metrik validasi. Hasil menunjukkan bahwa model CNN pada dataset *Hugging face* memberikan hasil terbaik, dengan akurasi 95%, *loss* 0,1, *validation accuracy* 72%, *validation loss* 1,87, *F1-score* 79%, presisi 90%, dan *recall* 72%, mengungguli LSTM yang cenderung *overfitting*. Secara keseluruhan, CNN terbukti lebih stabil dan andal dalam generalisasi, terutama dalam menangani data yang kompleks. Penelitian ini juga mengonfirmasi bahwa CNN lebih efektif dalam memproses data kompleks dibandingkan LSTM, yang cenderung lebih rentan terhadap *overfitting*, terutama pada data teks panjang.

**Kata Kunci:** Sistem rekomendasi, *Natural Language Processing* (NLP), CNN, LSTM, *Text Classification*.

**Ketua Jurusan Sistem Komputer**

Mengetahui,

**Pembimbing Tugas Akhir**

  
**Dr. Ir. Sukemi, M.T.**  
NIP. 196612032006041001

  
**Sutarno, S.T., M.T.**  
NIP. 197811012010121003



**IMPLEMENTATION OF NATURAL LANGUAGE PROCESSING (NLP)  
MODEL ON JOB RECOMMENDATION SYSTEM USING CONVOLUTIONAL  
NEURAL NETWORKS (CNN) AND LONG SHORT TERM MEMORY (LSTM)**

**Mhd. Haidir Fikri**

Department of Computer Systems, Faculty of Computer Science, Sriwijaya University

Email : [haidirfikri13@gmail.com](mailto:haidirfikri13@gmail.com)

**ABSTRACT**

Advances in technology present challenges in matching individual skills with appropriate job recommendations. This research develops a *Natural Language Processing (NLP)* based job recommendation system using *Convolutional Neural Network (CNN)* and *Long Short-Term Memory (LSTM)* models. The research data is obtained from the *Hugging Face* and *Kaggle* public datasets, which include descriptions of work experience and skills. The data was processed through pre-processing stages such as *stopword* removal, *regex*, *stemming*, *tokenization*, *padding*, as well as balancing techniques such as *SMOTE* and *undersampling*. The dataset is split for model training and testing, with performance measured through accuracy, loss, and validation metrics. Results show that the CNN model on the Hugging face dataset provides the best results, with 95% *accuracy*, 0.1 *loss*, 72% *validation accuracy*, 1.87 *validation loss*, 79% *F1-score*, 90% *precision*, and 72% *recall*, outperforming LSTM which tends to *overfitting*. Overall, CNN proved to be more stable and reliable in generalization, especially in handling complex data. This research also confirms that CNN are more effective in processing complex data than LSTM, which tend to be more prone to *overfitting*, especially on long text data.

**Keywords:** Recommendation system, *Natural Language Processing (NLP)*, CNN, LSTM, *Text Classification*.

**Acknowledge,**

**Head Of Computer System  
Departement**



**Dr. Ir. Sukemi, M.T.**  
NIP. 196612032006041001

**Final Project Advisor**



**Sutarno, S.T., M.T.**  
NIP. 197811012010121003

## DAFTAR ISI

<b>HALAMAN PENGESAHAN</b> .....	<b>ii</b>
<b>AUTHENTICATION PAGE</b> .....	<b>iii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iv</b>
<b>HALAMAN PERNYATAAN</b> .....	<b>v</b>
<b>KATA PENGANTAR</b> .....	<b>vi</b>
<b>ABSTRAK</b> .....	<b>viii</b>
<b>ABSTRACT</b> .....	<b>ix</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiv</b>
<b>DAFTAR TABEL</b> .....	<b>xv</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xvi</b>
<b>BAB 1 PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Tujuan Penelitian.....	3
1.3 Manfaat Penelitian.....	4
1.4 Rumusan Masalah .....	4
1.5 Batasan Masalah.....	4
1.6 Metodologi Penelitian .....	5
1.7 Sistematika Penulisan.....	6
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>8</b>
2.1 Penelitian Terdahulu.....	8
2.2 Sistem rekomendasi pekerjaan .....	9
2.3 <i>Natural Language Processing (NLP)</i> .....	10

2.4	<i>Deep learning</i> .....	11
2.5	<i>Machine learning</i> .....	12
2.6	<i>Convolutional Neural Network (CNN)</i> .....	14
2.7	<i>Long Short-Term Memory Network (LSTM)</i> .....	16
2.8	<i>Layer Deep learning</i> .....	19
2.8.1	<i>Batch normalization</i> .....	19
2.8.2	<i>Relu</i> .....	20
2.9	<i>Optimizer</i> .....	20
2.10	<i>Tensorflow</i> .....	21
2.11	<i>Keras</i> .....	21
2.12	<i>Natural language ToolKit (NLTK)</i> .....	22
2.13	<i>Cofusion matrix</i> .....	22
2.13.1	<i>Akurasi</i> .....	23
2.13.2	<i>Presisi</i> .....	23
2.13.3	<i>Recall</i> .....	24
2.13.4	<i>F1-Score</i> .....	24
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>25</b>
3.1	<i>Kerangka Kerja Penelitian</i> .....	25
3.2	<i>Dataset</i> .....	26
3.2.1.	<i>Data set Hugging Face</i> .....	27
3.2.2.	<i>Data set Kaggle</i> .....	30
3.3	<i>Perangkat yang digunakan</i> .....	33
3.3.1	<i>Perangkat keras (Hardware )</i> .....	33
3.3.2	<i>Perangkat Lunak (Software)</i> .....	33
3.4	<i>Preprocessing NLP</i> .....	34
3.5	<i>Balanced Data</i> .....	34

3.6	Proses <i>Training</i> dan <i>Validation</i> Model CNN dan LSTM.....	34
3.7	Proses Deteksi Rekomendasi Pekerjaan.....	39
3.8	Analisis Hasil Eksperimen Model CNN dan LSTM .....	41
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>42</b>
4.1	<i>Praprocessing</i> Data.....	42
4.1.1.	<i>Remove Stopword</i> .....	42
4.1.2.	<i>Regex (regular expressions)</i> .....	43
4.1.3.	<i>Stemming</i> .....	44
4.2	Balanced Dataset Hugging Face dan Kaggle .....	45
4.3	<i>Tokenization</i> .....	46
4.4	Model Sequential <i>CNN dan LSTM</i> .....	48
4.4.1.	Model Sequential CNN .....	48
4.4.2.	Model Sequential LSTM.....	49
4.5	Kumpulan Model CNN pada Dataset <i>Hugging face</i> .....	49
4.5.1.	CNN Dataset <i>Balanced</i> Manual .....	52
4.5.2.	Model CNN <i>Hugging face</i> kategori dan jumlah data sama .....	55
4.6	Kumpulan Model CNN Dataset <i>Resume Kaggle</i> .....	57
4.6.1	Model CNN Kaggle Kategori dan jumlah data yang sama.....	60
4.7	Kumpulan Model LSTM <i>Hugging Face</i> .....	63
4.7.1.	LSTM Dataset <i>Balanced</i> manual .....	67
4.7.2.	Model LSTM <i>Hugging Face</i> kategori dan jumlah data yang sama .....	70
4.8	Kumpulan Model Lstm data set <i>Kaggle</i> .....	73
4.8.1.	Model LSTM kaggle kategori dan jumlah data yang sama .....	76
4.9	Perbandingan antar model .....	79
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>82</b>
5.1.	Kesimpulan.....	82

5.2. Saran .....	83
<b>DAFTAR PUSTAKA .....</b>	<b>84</b>

## DAFTAR GAMBAR

<b>Gambar 2. 1</b> Arsitektur CNN Text .....	14
<b>Gambar 2. 2</b> Diagram sederhana unit LSTM .....	17
<b>Gambar 2. 3</b> Tipe <i>Gate</i> pada Model LSTM .....	18
<b>Gambar 3. 1</b> Kerangka Kerja Penelitian.....	26
<b>Gambar 3. 2</b> Tampilan Dataset Resume Hugging Face .....	27
<b>Gambar 3. 3</b> Diagram ata Hugging Face manual Balanced .....	30
<b>Gambar 3. 4</b> Tampilan Dataset Resume Kaggle .....	31
<b>Gambar 3. 5</b> Process training model.....	35
<b>Gambar 3. 6</b> Proses deteksi rekomendasi pekerjaan.....	39
<b>Gambar 4. 1</b> Penggunaan Removing Stopword .....	43
<b>Gambar 4. 2</b> Penggunaan Regex.....	43
<b>Gambar 4. 3</b> Penggunaan Stemming .....	45
<b>Gambar 4. 4</b> Diagram Data Hugging Face dan Kaggle Balanced Data .....	46
<b>Gambar 4. 5</b> Dictionary Tokenization Resume .....	47
<b>Gambar 4. 6</b> Sentence to sequence .....	48
<b>Gambar 4. 7</b> Model Sequential CNN.....	48
<b>Gambar 4. 8</b> Model Sequential LSTM .....	49
<b>Gambar 4. 9</b> Grafik Accuracy dan Loss model Terbaik CNN Hugging Face.....	52
<b>Gambar 4. 10</b> Grafik dan Loss Terbaik CNN Balanced manual .....	55
<b>Gambar 4. 11</b> Grafik Accuracy dan Loss model Terbaik CNN Kaggle.....	60
<b>Gambar 4. 12</b> Grafik Accuracy dan Loss model Terbaik LSTM Hugging Face	66
<b>Gambar 4. 13</b> Grafik Accuracy dan Loss Terbaik LSTM Balanced Manual .....	69
<b>Gambar 4. 14</b> Grafik Accuracy dan Loss model Terbaik LSTM Kaggle.....	76

## DAFTAR TABEL

<b>Tabel 1.</b> Penelitian terdahulu .....	8
<b>Tabel 2.</b> Confusion Matrix .....	23
<b>Tabel 3</b> Jenis dan sumber data .....	26
<b>Tabel 4.</b> Jumlah Data Per Kategori Hugging Face Dataset .....	28
<b>Tabel 5</b> Jumlah Data Per Kategori Kaggle Dataset .....	31
<b>Tabel 6.</b> Perangkat keras (Hardware) .....	33
<b>Tabel 7.</b> Perangkat Lunak (Software).....	33
<b>Tabel 8.</b> Hasil Eksperimen Model CNN Dataset <i>Hugging Face</i> .....	50
<b>Tabel 9</b> Hasil Eksperimen Model CNN Dataset Hugging Face Balanced Manual .....	53
<b>Tabel 10</b> Hasil Model CNN <i>Hugging face</i> kategori dan jumlah data sama .....	56
<b>Tabel 11.</b> Hasil Eksperimen Model CNN Dataset Kaggle .....	58
<b>Tabel 12</b> Hasil Model CNN Kaggle kategori dan jumlah data sama .....	61
<b>Tabel 13.</b> Hasil Eksperimen Model LSTM Dataset Hugging Face.....	64
<b>Tabel 14</b> Hasil Eksperimen Model LSTM Hugging Face Balanced manual .....	67
<b>Tabel 15</b> Hasil Model LSTM Hugging face kategori dan jumlah data sama .....	70
<b>Tabel 16.</b> Hasil Eksperimen Model LSTM Dataset <i>Kaggle</i> .....	73
<b>Tabel 17</b> Hasil Model LSTM Kaggle kategori dan jumlah data sama .....	76
<b>Tabel 18.</b> Hasil Perbandingan Model Terbaik.....	80



## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Uji coba Model .....	89
<b>Lampiran 2</b> Form perbaikan .....	90
<b>Lampiran 3</b> Hasil Cek Plagiat .....	92

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan teknologi dalam era *digital* telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk dunia kerja. *Transformasi digital* telah mengubah cara kita bekerja, mencari pekerjaan, dan merekrut tenaga kerja [1]. Teknologi telah membuka banyak peluang baru, namun juga menghadirkan tantangan-tantangan baru. Salah satu tantangan utama yang dihadapi oleh pencari kerja dan perusahaan adalah menemukan kesesuaian antara kualifikasi individu dan kebutuhan pekerjaan [2]

Persoalan *mismatch* di pasar kerja dan ketidaksesuaian pekerjaan dengan kualifikasi merupakan tantangan umum yang dihadapi oleh banyak negara di berbagai tingkatan pengembangan ekonomi [3]. masalah yang terus berkembang di pasar tenaga kerja global ini dapat menyebabkan penurunan kepuasan kerja bagi karyawan yang merasa keterampilannya tidak sepenuhnya dimanfaatkan dalam pekerjaannya [4]. Studi lebih lanjut mengindikasikan bahwa ketidakcocokan keterampilan tidak hanya mempengaruhi kepuasan kerja tetapi juga memiliki implikasi ekonomi yang lebih luas, termasuk upah yang lebih rendah dan peluang karir yang terbatas bagi individu yang mengalami ketidakcocokan keterampilan [5].

Tantangan ini muncul dari dinamika kompleks antara ketersediaan peluang pekerjaan dan kebingungan para pencari kerja dalam menemukan pekerjaan yang sesuai dengan kualifikasi mereka [6]. Ketidakcocokan antara tawaran pekerjaan yang tersedia dan keterampilan yang dimiliki oleh pencari kerja menjadi akar permasalahan yang perlu diatasi dengan serius. Hal ini menciptakan ketidakharmonisan antara pasokan dan permintaan tenaga kerja di pasar [3].

Perkembangan teknologi, khususnya dalam *Natural Language Processing* (NLP), memberikan peluang untuk menghadapi tantangan ini secara inovatif. Penelitian ini hadir sebagai langkah proaktif untuk mengusulkan penerapan teknologi (NLP) sebagai solusi yang efektif untuk membantu para pencari kerja

menemukan pekerjaan yang sesuai dengan kualifikasi mereka [7]. Terdapat dua metode dalam menyelesaikan masalah ini, yaitu pendekatan statistik dan machine learning

Metode statistik tradisional seperti *Simple Moving Average* (SMA), *Weighted Moving Average* (WMA), dan *Exponential Smoothing*, umumnya digunakan untuk mengidentifikasi tren dalam data. Namun, model-model ini memiliki keterbatasan dalam menangkap kompleksitas dan *non-linearitas* data pasar tenaga kerja yang dinamis. Akibatnya, prediksi yang dihasilkan cenderung kurang akurat dan relevan, terutama dalam menghadapi perubahan yang cepat dan tidak terduga [8].

Sebaliknya, model pembelajaran mendalam seperti *Convolutional Neural Networks* (CNN) dan *Long Short-Term Memory Networks* (LSTM) dirancang khusus untuk menangani data yang kompleks dan berurutan. CNN mampu mengekstrak fitur semantik dari teks deskripsi pekerjaan [9], sementara LSTM dapat menangkap konteks yang lebih luas dalam profil pencari kerja. Kemampuan ini memungkinkan model pembelajaran mendalam menghasilkan rekomendasi yang lebih *Presisi* dan relevan[10], sehingga meningkatkan peluang pencari kerja untuk menemukan pekerjaan yang sesuai dengan kualifikasi mereka dan bagi Perusahaan mempercepat proses *requirements* karyawan .

CNN yang awalnya dikembangkan untuk *pemrosesan citra* telah berhasil diterapkan dalam *pemrosesan teks* [11]. Secara dasar, CNN menggunakan *filter* (kernel) untuk mengekstrak fitur dari *input*, dan konsep ini dapat diadopsi pada teks melalui proses yang dikenal sebagai *konvolusi* teks. Untuk mereduksi atau memperkecil hasil *konvolusi* serta mencegah *overfitting*, CNN menggunakan *Pooling layer*. *Pooling layer* memiliki dua jenis utama, yaitu *max pooling* dan *average pooling*. *Max pooling* digunakan untuk mengekstraksi fitur yang paling dominan dari suatu wilayah, sementara *Average Pooling* berfungsi untuk meratakan informasi dengan mengambil nilai rata-rata dari wilayah tersebut [12].

LSTM merupakan jenis arsitektur *Recurrent Neural Network* (RNN) yang dikembangkan khusus untuk menangani masalah ketergantungan jarak jauh dalam data berurutan, seperti teks [13]. Dengan sel memori yang dapat menyimpan dan mengakses informasi dari masa lalu, LSTM mampu mempertahankan informasi

penting dalam jangka waktu yang lebih lama. Selain itu, LSTM efektif dalam mengatasi masalah gradien yang menghilang atau meledak, yang sering dialami oleh RNN konvensional. Karena kemampuannya ini, LSTM banyak digunakan berbagai tugas pemrosesan teks yang melibatkan pemahaman konteks jarak jauh, seperti terjemahan mesin, analisis sentimen, dan pembangkitan teks kreatif [14].

Melalui pelatihan model menggunakan data yang telah diproses dengan teknologi NLP, penelitian ini bertujuan untuk memberikan pemahaman mendalam tentang efektivitas kedua model tersebut. Dengan judul lengkap "**Implementasi Model *Natural Language Processing* (NLP) pada Sistem Rekomendasi Pekerjaan menggunakan *Convolutional Neural Networks* (CNN) dan *Long Short-Term Memory Network* (LSTM)**" diharapkan bahwa penelitian ini akan memberikan informasi bagi pencari kerja berdasarkan keahlian yang sesuai dengan yang dimiliki, khususnya dengan memanfaatkan kemajuan teknologi NLP dan juga dengan menggunakan model CNN dan LSTM.

Langkah awal penelitian ini melibatkan teknik *Preprocessing data*, seperti penggunaan *tokenizer* dan *eliminasi stopwords*, guna mempersiapkan data dengan cermat dan mengurangi hambatan dalam proses analisis [15]. Dalam konteks teknik pemodelan, penelitian ini memfokuskan diri pada CNN dan LSTM sebagai model yang memiliki potensi signifikan dalam menentukan rekomendasi pencarian pekerjaan yang sesuai.

## 1.2 Tujuan Penelitian

Melalui latar belakang tersebut ditetapkan tujuan penelitian sebagai berikut:

1. Mengetahui proses *Preprocessing data* bidang NLP yang mencakup penggunaan *Stemming*, *regular expressions* dan *eliminasi stopwords*.
2. Mengetahui implementasi pada bidang NLP dalam model CNN dan LSTM dalam sistem rekomendasi pekerjaan.
3. Mengukur kinerja dan mendapatkan model terbaik dari model CNN dan LSTM pada sistem rekomendasi pekerjaan.

### 1.3 Manfaat Penelitian

Adapun manfaat dari penelitian tugas akhir yang dilakukan ini, antara lain:

1. Memberikan pengembangan pengetahuan mengenai bidang NLP terkait penggunaan model CNN dan LSTM dalam Sistem rekomendasi pekerjaan .
2. Memberikan Solusi Perusahaan dalam seleksi calon pekerja berdasarkan CV sesuai dengan keahlian dan kriteria yang di inginkan.
3. Memberikan panduan bagi pencari kerja dalam menentukan pilihan pekerjaan yang sesuai.

### 1.4 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan, permasalahan utama yang akan dijelaskan dalam penelitian ini adalah :

1. Apa tahapan proses *preprocessing* data dalam bidang NLP yang diperlukan dalam implementasi model CNN dan LSTM ?
2. Bagaimana implementasi model CNN dan LSTM dalam konteks sistem rekomendasi pekerjaan ?
3. Bagaimana perbandingan kinerja antara model CNN dan LSTM dalam konteks sistem rekomendasi pekerjaan ?

### 1.5 Batasan Masalah

Setelah merumuskan masalah dari latar belakang penelitian, berikut cakupan atau batasan permasalahan pada tugas akhir ini, antara lain:

1. Model yang akan digunakan berasal dari *machine learning* yaitu CNN dan LSTM .
2. Tidak akan membahas secara mendalam tentang model NLP .
3. Data yang digunakan dalam penelitian ini menggunakan Bahasa Inggris.
4. Implementasi dilakukan menggunakan Bahasa *Python*.
5. Jumlah data masing masing data set adalah *Hugging Face* 32.481 dan *Kaggle* 3.421.
6. Data terdiri dari 52 kategori pekerjaan untuk *Hugging Face* dan 46 kategori untuk *Kaggle* .

## **1.6 Metodologi Penelitian**

Metodologi yang diterapkan dalam penyusunan Tugas Akhir ini melalui beberapa tahapan sebagai berikut :

### **1. Tahap Pertama (Studi Pustaka)**

Tahap pertama yang akan dilakukan dalam penelitian ini adalah mencari sumber informasi, memahami, serta mempelajari kajian literatur dan referensi seperti buku, artikel yang terkait, maupun jurnal ilmiah yang berkaitan dengan NLP dalam sistem rekomendasi kerja dan juga CNN dan LSTM. Sehingga dapat menjadi penunjang pada metodologi yang akan diterapkan dalam penelitian ini.

### **2. Tahap Kedua (Perancangan Sistem)**

Pada tahap kedua membahas tentang perancangan sistem yang akan dilakukan dalam penelitian ini. Dimana, tahap ini akan membuat suatu perancangan pemodelan simulasi dengan program berbahasa python.

### **3. Tahap Ketiga (Pengujian)**

Tahap ketiga ini adalah melakukan pengujian data. tahap ini dilakukan jika semua sistem telah dibuat dan dikonfigurasi, maka akan dilakukan pengujian sesuai batasan masalah pada penelitian agar mendapatkan hasil yang optimal.

### **4. Tahap Keempat (Analisa)**

Tahap ini akan dilakukan analisa dari hasil tahap pengujian. dimana pada tahap ini akan dilakukan analisa dari proses penelitian yang tujuannya untuk mengetahui model yang terbaik dan mengetahui kekurangan pada hasil perancangan dan factor apa saja yang menjadi penyebabnya. Sehingga, bisa dilakukan pengembangan lagi untuk peneliti selanjutnya.

### **5. Tahap Kelima (Kesimpulan dan saran )**

Tahap ini merupakan tahapan terakhir, dimana penulis melakukan penarikan kesimpulan yang berdasarkan studi pustaka, hasil perancangan sistem, serta hasil analisa penelitian. Dan memberikan saran untuk peneliti selanjutnya sebagai bahan referensi.

## **1.7 Sistematika Penulisan**

Dalam mempermudah penyusunan tugas akhir ini dan juga membuat rangkaian isi dari setiap bab lebih jelas, maka disusun sistematika penulisan sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini membahas mengenai latar belakang masalah penelitian, tujuan dan manfaat penelitian, perumusan masalah, batasan masalah, metodologi penelitian, serta sistematika penulisan tugas akhir.

### **BAB II TINJAUAN PUSTAKA**

Bab 2 berisi tentang landasan teori yang sesuai dengan penelitian dan berasal dari sumber-sumber teori yang relevan, serta referensi dari hasil penelitian sebelumnya dan beberapa referensi untuk memecahkan masalah.

### **BAB III METODOLOGI**

Pada Bab 3 ini mengemukakan tentang tahapan-tahapan (metodologi) yang digunakan untuk mencari, mengumpulkan dan menganalisa permasalahan penelitian serta pemecahan masalah pada penelitian tugas akhir ini.

### **BAB IV HASIL DAN ANALISIS**

Pada bab 4 ini menjelaskan mengenai hasil pengujian yang telah dilakukan, data-data yang didapat dari pengujian akan dianalisa menggunakan metode yang menjadi acuan dalam tugas akhir ini, serta sebagai pembuktian dari sistem yang telah dibuat yang kemudian dilakukan pengujian untuk mendapatkan model yang terbaik serta mendapatkan keakuratan dan analisa dari penelitian tugas akhir yang dilakukan.



## **BAB V KESIMPULAN DAN SARAN**

Pada tahap ini akan dirumuskan suatu kesimpulan yang diperoleh dari permasalahan, studi pustaka, metodologi penelitian dan analisis hasil pengujian. Kemudian beberapa saran yang diharapkan dapat dijadikan landasan untuk penelitian lanjutan.

## DAFTAR PUSTAKA

- [1] P. M. Gilch and J. Sieweke, "Recruiting digital talent: The strategic role of recruitment in organisations' digital transformation," *Ger. J. Hum. Resour. Manag.*, vol. 35, no. 1, pp. 53–82, 2021, doi: 10.1177/2397002220952734.
- [2] O. Allal-Chérif, A. Yela Aránega, and R. Castaño Sánchez, "Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence," *Technol. Forecast. Soc. Change*, vol. 169, no. April, 2021, doi: 10.1016/j.techfore.2021.120822.
- [3] I. Hana Kaharudin, M. S. Ab Rahman, and W. N. Syamimi Wan Zin, "The Literature Review on Job Mismatch and Unemployment Among Graduates," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 13, no. 11, pp. 26–32, 2023, doi: 10.6007/ijarbss/v13-i11/19149.
- [4] S. Bischof, "Mismatched, but not aware of it? How subjective and objective skill mismatch affects employee job satisfaction," *Soc. Sci.*, vol. 10, no. 10, 2021, doi: 10.3390/socsci10100389.
- [5] P. J. Sloane and K. Mavromaras, "Overeducation, skill mismatches, and labor market outcomes for college graduates," *IZA World Labor Evidence-Based Policy Mak.*, vol. 88, no. 2, pp. 1–10, 2020, doi: 10.15185/izawol.88.v2.
- [6] N. Bhoir *et al.*, "Resume Parser using hybrid approach to enhance the efficiency of Automated Recruitment Processes," *Authorea*, pp. 1–12, 2023, doi: DOI: 10.22541/au.168170278.82268853/v1.
- [7] M. Soni, S. Gomathi, and Y. Bhupendra Kumar Adhyaru, "Natural language processing for the job portal enhancement," *2020 7th Int. Conf. Smart Struct. Syst. ICSSS 2020*, pp. 14–17, 2020, doi: 10.1109/ICSSS49621.2020.9202046.
- [8] I. Bhattacharjee and P. Bhattacharja, "Stock Price Prediction: A Comparative Study between Traditional Statistical Approach and Machine Learning Approach," *2019 4th Int. Conf. Electr. Inf. Commun. Technol. EICT 2019*, no. December, pp. 1–6, 2019, doi:

10.1109/EICT48899.2019.9068850.

- [9] Y. Li, G. Ning, and H. Zhang, “Research on Position Recommendation System Based on Convolutional Neural Network,” *J. Phys. Conf. Ser.*, vol. 2171, no. 1, 2022, doi: 10.1088/1742-6596/2171/1/012065.
- [10] A. Nigam, A. Roy, H. Singh, and H. Waila, “Job recommendation through progression of job selection,” *Proc. 2019 6th IEEE Int. Conf. Cloud Comput. Intell. Syst. CCIS 2019*, pp. 212–216, 2019, doi: 10.1109/CCIS48116.2019.9073723.
- [11] J. Deng, L. Cheng, and Z. Wang, “Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification,” *Comput. Speech Lang.*, vol. 68, p. 101182, 2021, doi: 10.1016/j.csl.2020.101182.
- [12] A. Zafar *et al.*, “A Comparison of Pooling Methods for Convolutional Neural Networks,” *Appl. Sci.*, vol. 12, no. 17, pp. 1–21, 2022, doi: 10.3390/app12178643.
- [13] Y. Luan and S. Lin, “Research on Text Classification Based on CNN and LSTM,” *Proc. 2019 IEEE Int. Conf. Artif. Intell. Comput. Appl. ICAICA 2019*, pp. 352–355, 2019, doi: 10.1109/ICAICA.2019.8873454.
- [14] G. Liu and J. Guo, “Bidirectional LSTM with attention mechanism and convolutional layer for text classification,” *Neurocomputing*, vol. 337, pp. 325–338, 2019, doi: 10.1016/j.neucom.2019.01.078.
- [15] A. Tabassum and R. R. Patil, “A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing,” *Int. Res. J. Eng. Technol.*, no. June, pp. 4864–4867, 2020, [Online]. Available: [www.irjet.net](http://www.irjet.net)
- [16] M. Tikhonova and A. Gavrishchuk, “NLP methods for automatic candidate’s CV segmentation,” *2019 Int. Conf. Eng. Telecommun. EnT 2019*, pp. 1–5, 2019, doi: 10.1109/EnT47717.2019.9030535.
- [17] P. Senarathne, M. Silva, A. Methmini, D. Kavinda, and S. Thelijjagoda, “Automate Traditional Interviewing Process Using Natural Language Processing and Machine Learning,” *2021 6th Int. Conf. Converg. Technol. I2CT 2021*, pp. 1–6, 2021, doi: 10.1109/I2CT51068.2021.9418115.

- [18] N. Vanetik, A. Kolesnev, and G. Kogan, “NLP-based Screening for IT Job Vacancies: A Case Study,” *CEUR Workshop Proc.*, vol. 3477, pp. 13–22, 2023.
- [19] C. H. Ayishathahira, C. Sreejith, and C. Raseek, “Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing,” *2018 Int. CET Conf. Control. Commun. Comput. IC4 2018*, pp. 388–393, 2018, doi: 10.1109/CETIC4.2018.8530883.
- [20] A. Mukherjee, S. Mukhopadhyay, P. K. Panigrahi, and S. Goswami, “Utilization of Oversampling for multiclass sentiment analysis on Amazon Review Dataset,” *2019 IEEE 10th Int. Conf. Aware. Sci. Technol. iCAST 2019 - Proc.*, pp. 1–6, 2019, doi: 10.1109/ICAwST.2019.8923260.
- [21] R. Mishra and S. Rathi, *Efficient and Scalable Job Recommender System Using Collaborative Filtering*. Singapore: Springer Singapore, 2020.
- [22] R. Taufiq and A. A. Permana, “Sistem Pendukung Keputusan Penerimaan Karyawan Menggunakan Simple Additive Weighting Studi Kasus PT. Trafoindo Prima Perkasa,” *J. Al-AZHAR Indones. SERI SAINS DAN Teknol.*, vol. 4, no. 4, p. 186, 2018, doi: 10.36722/sst.v4i4.309.
- [23] T. Young, D. Hazarika, S. Poria, and E. Cambria, “Recent trends in deep learning based natural language processing [Review Article],” *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, 2018, doi: 10.1109/MCI.2018.2840738.
- [24] A. N. Rohman, R. Luviana Musyarofah, E. Utami, and S. Raharjo, “Natural Language Processing on Marketplace Product Review Sentiment Analysis,” *2020 2nd Int. Conf. Cybern. Intell. Syst. ICORIS 2020*, 2020, doi: 10.1109/ICORIS50180.2020.9320827.
- [25] E. Le Penneç, “Foundations of Machine Learning Book of Proofs,” 2023. [Online]. Available: <https://api.semanticscholar.org/CorpusID:264045639>
- [26] S. Moraboena, G. Ketepalli, and P. Ragam, “A deep learning approach to network intrusion detection using deep autoencoder,” *Rev. d’Intelligence Artif.*, vol. 34, no. 4, pp. 457–463, 2020, doi: 10.18280/ria.340410.
- [27] S. Setiowati, Zulfanahri, E. L. Franita, and I. Ardiyanto, “A review of optimization method in face recognition: Comparison deep learning and

- non-deep learning methods,” *2017 9th Int. Conf. Inf. Technol. Electr. Eng. ICITEE 2017*, vol. 2018-Janua, pp. 1–6, 2017, doi: 10.1109/ICITEED.2017.8250484.
- [28] D. Alsaleh and S. Larabi-Marie-Sainte, “Arabic Text Classification Using Convolutional Neural Network and Genetic Algorithms,” *IEEE Access*, vol. 9, pp. 91670–91685, 2021, doi: 10.1109/ACCESS.2021.3091376.
- [29] S. Selva Birunda and R. Kanniga Devi, “A Review on Word *Embedding* Techniques for Text Classification,” in *Innovative Data Communication Technologies and Application*, 2021, pp. 267–281.
- [30] W. Haitao, H. Jie, Z. Xiaohong, and L. Shufen, “A short text classification method based on n-gram and cnn,” *Chinese J. Electron.*, vol. 29, no. 2, pp. 248–254, 2020, doi: 10.1049/cje.2020.01.001.
- [31] Y. Zhang, “Research on text classification method based on lstm neural network model,” in *Proceedings of IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2021*, 2021, pp. 1019–1022. doi: 10.1109/IPEC51340.2021.9421225.
- [32] N. Pattanajak and H. Malekmohamadi, “Improving a 3-D convolutional neural network model reinvented from VGG16 with batch normalization,” *Proc. - 2019 IEEE SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Internet People Smart City Innov. SmartWorld/UIC/ATC/SCALCOM/IOP/SCI 2019*, pp. 45–50, 2019, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00052.
- [33] J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, “Reluplex made more practical: Leaky ReLU,” *Proc. - IEEE Symp. Comput. Commun.*, vol. 2020-July, 2020, doi: 10.1109/ISCC50000.2020.9219587.
- [34] Chitra Desai, “Comparative Analysis of Optimizers in Deep Neural Networks,” *Int. J. Innov. Sci. Res. Technol.*, vol. 5, no. 10, pp. 959–962, 2020.
- [35] B. Pang, E. Nijkamp, and Y. N. Wu, “Deep Learning With TensorFlow: A Review,” *J. Educ. Behav. Stat.*, vol. 45, no. 2, pp. 227–248, 2020, doi: 10.3102/1076998619872761.
- [36] R. S. Kathuria, S. Gautam, A. Singh, S. Khatri, and N. Yadav, “Real Time

- Sentiment Analysis on Twitter Data Using Deep Learning(Keras),” *Proc. - 2019 Int. Conf. Comput. Commun. Intell. Syst. ICCIS 2019*, vol. 2019-Janua, pp. 69–73, 2019, doi: 10.1109/ICCIS48478.2019.8974557.
- [37] F. Alshani, A. Apon, A. Herzog, I. Safro, and J. Sybrandt, “Accelerating Text Mining Using Domain-Specific Stop Word Lists,” *Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020*, pp. 2639–2648, 2020, doi: 10.1109/BigData50022.2020.9378226.
- [38] M. Heydarian and T. E. Doyle, “MLCM : Multi-Label Confusion Matrix,” pp. 19083–19095, 2022.
- [39] M. Shorfuzzaman, “Detection of cyber attacks in IoT using tree-based ensemble and feedforward neural network,” *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2020-Octob, pp. 2601–2606, 2020, doi: 10.1109/SMC42975.2020.9283443.