

**ANALISIS PERBANDINGAN ARSITEKTUR LONG SHORT-  
TERM MEMORY UNTUK KLASIFIKASI TEKS**

**JALUR SCIENTIST**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**MOH. FAJAR ADBILLAH**

**19.11.3237**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2024**

**ANALISIS PERBANDINGAN ARSITEKTUR LONG SHORT-  
TERM MEMORY UNTUK KLASIFIKASI TEKS**

**JALUR SCIENTIST**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**MOH. FAJAR ABDILLAH**

**19.11.3237**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA**

**YOGYAKARTA**

**2024**

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS PERBANDINGAN ARSITEKTUR LONG SHORT-TERM  
MEMORY UNTUK KLASIFIKASI TEKS

yang disusun dan diajukan oleh

**Moh. Fajar Abdillah**

19.11.3237

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 20 Desember 2023

Dosen Pembimbing,



Kusnawi, M. Eng  
NIK. 190302112

**HALAMAN PENGESAHAN**  
**JALUR SCIENTIST**  
**ANALISIS PERBANDINGAN ARSITEKTUR LONG SHORT-TERM**  
**MEMORY UNTUK KLASIFIKASI TEKS**

yang disusun dan diajukan oleh

**Moh. Fajar Abdillah**

19.11.3237

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 20 Desember 2023

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Anna Baita, M.Kom  
NIK. 190302290

Muhammad Ainul Fikri, S. T. M. Eng  
NIK. 190302680

Kusnawi, M.Eng  
NIK. 190302112



Paper ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 20 Desember 2023

**DEKAN FAKULTAS ILMU KOMPUTER**



Hanif Al Fatta, S.Kom., M.Kom.  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Moh. Fajar Abdillah

NIM : 19.11.3237

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Kusnawi, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Desember 2023

Yang Menyatakan,

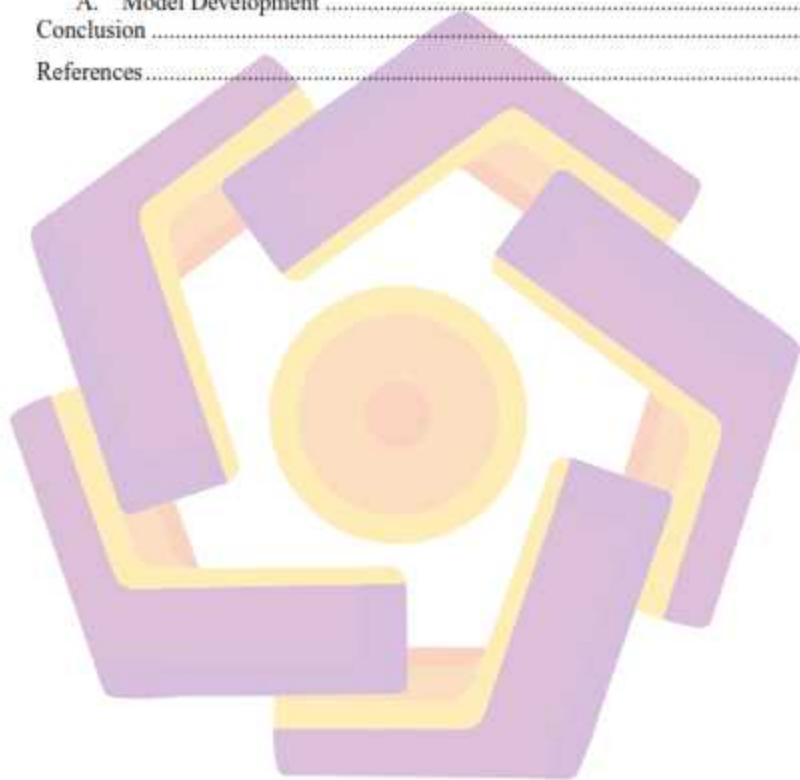


Moh. Fajar Abdillah

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
DAFTAR ISI.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
BAB I PENDAHULUAN.....	9
BAB II LANDASAN TEORI.....	11
A. <i>Natural Language Processing (NLP)</i> .....	11
B. <i>Klasifikasi Teks</i> .....	11
C. <i>Data Preprocessing</i> .....	11
D. <i>Pre- Trained Word Embedding</i> .....	12
E. <i>Recurrent Neural Network (RNN)</i> .....	13
F. <i>Long Short-Term Memory (LSTM)</i> .....	13
G. <i>SoftMax Layer</i> .....	13
H. <i>Batch Normalization</i> .....	14
BAB III METODE PENELITIAN.....	15
A. <i>Pengumpulan Data</i> .....	16
B. <i>Data Preprocessing</i> .....	16
C. <i>Convert to Numerical Format</i> .....	17
D. <i>Pre- Trained Word Embedding</i> .....	18
E. <i>Developing Model</i> .....	18
F. <i>Validation and Testing</i> .....	23
BAB IV PEMBAHASAN.....	24
A. <i>Model Development</i> .....	24
BAB V KESIMPULAN.....	27
REFERENSI.....	28
LAMPIRAN.....	33
A. <i>Letter of Acceptance (LOA)</i> .....	33
B. <i>Lembar Review dari Penyelenggara</i> .....	33
C. <i>BUKTI KARYA ILMIAH</i> .....	34
Abstract (Times 10pt).....	34
Introduction.....	34
Method.....	35

A. Data Collection .....	36
B. Data Preprocessing .....	36
C. Convert to Numerical Format .....	38
D. Pre- Trained Word Embedding.....	38
E. Developing Model .....	39
F. Validation and Testing.....	44
Results and Discussion .....	44
A. Model Development .....	44
Conclusion .....	46
References.....	46



## ABSTRAK

Klasifikasi Teks yang merupakan bagian dari NLP adalah pengelompokan objek berupa teks yang berdasarkan ciri – ciri tertentu yang menunjukkan kesamaan antara satu dokumen dengan yang lainnya . Salah satu metode yang digunakan dalam Klasifikasi Teks yaitu LSTM. Performa dari metode LSTM sendiri dipengaruhi oleh beberapa hal seperti dataset, arsitektur dan tools yang dipakai untuk mengklasifikasikan teks. Pada kesempatan kali ini peneliti akan menganalisis pengaruh jumlah layer pada arsitektur LSTM terhadap performa yang dihasilkan oleh metode LSTM. Dalam penelitian kali ini menggunakan data IMDB Movie Reviews dengan jumlah data sebanyak 50.000 data. Data terdiri dari data positif, negatif dan terdapat data yang belum memiliki label. Data IMDB Movie Reviews akan melalui beberapa tahapan sebagai berikut: Pengumpulan data, Pre-processing data, Convert to numerical format (konversi teks ke numerik), Embedding Teks menggunakan model Pre – trained word embedding : Fastext , melatih dan menguji klasifikasi model menggunakan LSTM, terakhir melakukan validasi dan menguji model sehingga didapatkanlah hasil dari tahapan penelitian ini. Hasil dari penelitian ini menunjukkan bahwa arsitektur LSTM satu layer memiliki akurasi terbaik dibandingkan LSTM dua layer dan tiga layer dengan training accuracy dan testing accuracy dari LSTM satu layer yaitu 0,856 dan 0,867. Sedangkan training accuracy dan testing accuracy pada LSTM dua layer adalah 0,846 dan 0,854, dan training accuracy dan testing accuracy pada tiga layer yaitu 0,848 dan 864.

**Kata kunci:** Klasifikasi Teks, NLP, LSTM, FastText



## ABSTRACT

*Text Classification which is part of NLP is a grouping of objects in the form of text based on certain characteristics that show similarities between one document and another. One of the methods used in Text Classification is LSTM. The performance of the LSTM method itself is influenced by several things such as datasets, architecture and tools used to classify text. On this occasion, researchers will analyze the effect of the number of layers in the LSTM architecture on the performance generated by the LSTM method. This research uses IMDB Movie Reviews data with a total of 50,000 data. The data consists of positive, negative data and there is data that does not yet have a label. IMDB Movie Reviews data will go through several stages as follows: Data collection, data pre-processing, convert to numerical format, text embedding using the pre-trained word embedding model: Fasttext, train and test the classification model using LSTM, finally validate and test the model so that the results are obtained from the stages of this research. The results of this study show that the one-layer LSTM architecture has the best accuracy compared to two-layer and three-layer LSTM with training accuracy and testing accuracy of one-layer LSTM which are 0.856 and 0.867. While the training accuracy and testing accuracy on two-layer LSTM are 0.846 and 0.854, the training accuracy and testing accuracy on three layers are 0.848 and 864.*

**Keyword:** Text Classification, NLP, LSTM, FastText