



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

- a. Pengutipan hara untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Klasifikasi Abstrak Artikel Ilmiah pada *Cora Dataset* Menggunakan *Long Short Term Memory*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Januari 2025

Daffa Fikri
G640120108

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

- a. Pengutipan hara untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

DAFFA FIKRI. Klasifikasi Abstrak Artikel Ilmiah pada *Cora Dataset* Menggunakan *Long Short Term Memory*. Dibimbing oleh AHMAD RIDHA.

Penelitian ini bertujuan untuk mengklasifikasikan topik artikel ilmiah berdasarkan abstraknya dengan menggunakan metode *Long Short-Term Memory* (LSTM) yang dipadukan dengan FastText sebagai word embedding. Klasifikasi teks pada artikel ilmiah menjadi penting karena memudahkan peneliti dalam menemukan referensi yang relevan. Meskipun banyak penelitian terkait klasifikasi teks, masih terdapat tantangan dalam mengembangkan model klasifikasi yang efisien dan akurat, terutama pada *dataset* ilmiah seperti CORA. Penelitian ini melalui beberapa tahapan, yaitu pengumpulan data, praproses data, *word embedding*, pembagian data, pembangunan model, serta evaluasi dan perbandingan model. Sebanyak dua belas model diuji dengan lima kali pelatihan pada setiap model, menghasilkan 60 hasil pengujian dengan menggunakan lima *fold cross-validation*. Hasil penelitian menunjukkan bahwa model kedelapan memiliki performa terbaik dengan rata-rata akurasi 0,9355 dan rata-rata *loss* 0,3108, menggunakan ukuran vektor 16, *batch size* 64, dan *epoch* 50. Namun, jika mempertimbangkan kecepatan pengujian tanpa memperhatikan akurasi tertinggi, model pertama menjadi pilihan terbaik dengan waktu pengujian 47,2 menit.

Kata Kunci: Artikel Ilmiah, FastText, LSTM, Topik

ABSTRACT

DAFFA FIKRI. Classification of Scientific Article Abstracts on Cora Dataset Using Long Short Term Memory. Supervised by AHMAD RIDHA.

This study aims to classify the topics of scientific articles based on their abstracts using the Long Short-Term Memory (LSTM) method combined with FastText as a word embedding technique. Text classification in scientific articles is crucial as it aids researchers in finding relevant references efficiently. Despite numerous studies on text classification, developing efficient and accurate models for scientific datasets like CORA remains challenging.. The research involves several stages, including data collection, data preprocessing, word embedding, data splitting, model construction, and model evaluation and comparison. Twelve models were tested with five training runs for each model, resulting in 60 evaluation outcomes using five-fold cross-validation. The results indicate that Model 8 performed the best, with an average accuracy of 0.9355 and an average loss of 0.3108, using a vector size of 16, batch size of 64, and 50 epochs. However, when considering testing speed without prioritizing the highest accuracy, Model 1 is the optimal choice, with a testing time of 47.2 minutes.

Keywords: FastText, LSTM, Papers, Topics.



@Hak cipta milik IPB University

IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

- a. Pengutipan hara untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



**KLASIFIKASI ABSTRAK ARTIKEL ILMIAH
PADA CORA DATASET MENGGUNAKAN
*LONG SHORT TERM MEMORY***

DAFFA FIKRI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Sarjana Ilmu Komputer

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta mitik IPB University

PER University



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah

b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:

1 Dr. Toto Haryanto, S.Si., M.Si

2 Hari Agung Adrianto, S.Kom., M.Si, Ph. D



@Hak cipta milik IPB University

IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

- a. Pengutipan hara untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

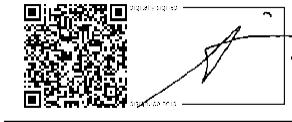
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Klasifikasi Abstrak Artikel Ilmiah pada Cora Dataset
Menggunakan *Long Short Term Memory*

Nama : Daffa Fikri
NIM : G6401201086

Disetujui oleh



Pembimbing 1:
Ahmad Ridha, S.Kom., M.S

Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 2008121002



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanaahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Mei 2024 sampai bulan Agustus 2024 ini ialah *machine learning*, dengan judul “Klasifikasi Abstrak Artikel Ilmiah pada Cora Dataset Menggunakan Long Short Term Memory”.

Penulis menyadari bahwa skripsi ini tidak mungkin selesai tanpa adanya dukungan, bantuan, bimbingan, dan saran dari berbagai pihak selama penyusunan. Pada kesempatan ini, penulis menyampaikan rasa terima kasih kepada:

- a. Ayah dan Ibu serta kedua adik yang selalu memberikan dukungan tanpa henti setiap kondisi yang saya alami. Terima kasih atas doa, kasih sayang serta pengorbanan yang tiada tara.
- b. Bapak Ahmad Ridha, S.Kom., M.S. selaku dosen pembimbing skripsi yang senantiasa memberikan bimbingan, saran, dan mendorong penulis agar dapat segera menyelesaikan penelitian.
- c. Bapak Dr. Toto Haryanto, S.Si., M.Si. dan Bapak Hari Agung Adrianto, S.Kom., M.Si., Ph.D selaku dosen penguji ujian akhir penelitian yang telah memberikan saran sehingga penelitian ini menjadi lebih baik.
- d. Ahmad Bintang Arif selaku teman yang selalu memberikan ide, motivasi, saran, semangat, dan waktu luangnya kepada penulis agar segera menyelesaikan penelitian.
- e. Azriel Izwanie, Syabil Rofilah, dan Zulfa Meilida yang selalu memberikan dukungan kepada penulis serta menemani penulis dalam melewati masa-masa krisis dalam mengerjakan skripsi.
- f. Ahmed Siddiq, Dzaki Fahreza, Fakhri Givan, M Rafli, Qais Jabbar, Azzam Dzulhilmie, Pulung Rafi dan teman-teman *discord* yang selalu menemani penulis dalam bermain yang telah memberikan motivasi kepada penulis untuk segera menyelesaikan penelitian.
- g. Mahasiswa program studi S1 Ilmu Komputer angkatan 57 yang telah memberikan bantuan dan semangat kepada penulis selama perkuliahan, dan
- h. Teman-teman penulis lainnya baik di dalam maupun di luar kampus yang telah memberikan semangat.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, November 2024

Daffa Fikri

**DAFTAR ISI**

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	ix
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup Penelitian	2
METODE	3
2.1 Data Penelitian	3
2.2 Tahapan Penelitian	3
2.3 Lingkungan Pengembangan Sistem	10
III. HASIL DAN PEMBAHASAN	11
3.1 Persiapan Data	11
3.2 Pembentukan Model	14
3.3 Evaluasi Model	15
IV. SIMPULAN DAN SARAN	21
4.1 Simpulan	21
4.2 Saran	21
DAFTAR PUSTAKA	22
LAMPIRAN	24
RIWAYAT HIDUP	26

Hak Cipta Dilindungi Undang-undang
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



1	Perbandingan berbagai versi <i>dataset</i> Cora	9
2	Kombinasi parameter pembangunan model LSTM	9
3	Contoh data artikel ilmiah pada <i>dataset</i> akhir	11
4	Perbandingan abstrak sebelum dan setelah praproses	13
5	Hasil pembagian <i>dataset</i> dengan K-Fold	13
6	Konfigurasi <i>hyperparameter</i> FastText	14
7	Hasil matriks <i>embedding</i>	14
8	Detail arsitektur model LSTM	15
9	Performa model setiap kombinasi	16
10	Hasil performa <i>recall</i> , <i>precision</i> , dan <i>f1-score</i> model	20

DAFTAR GAMBAR

1	Diagram alur penelitian	3
2	Diagram alur pengumpulan	4
3	Diagram proses FastText <i>Embedding</i>	7
4	Diagram alur pembagian <i>k-folding</i>	8
5	Diagram proses pembentukan model LSTM (Murthy <i>et al.</i> 2020)	8
6	Sebaran kata unik dengan range 1-20	12
7	Grafik kinerja model pada <i>fold</i> 1 hingga 3	16
8	Grafik kinerja model pada <i>fold</i> 4 hingga 5	17
9	<i>Confusion matrix</i> model pada <i>fold</i> 1 dan 2	17
10	<i>Confusion matrix</i> model pada <i>fold</i> 3 dan 4	18
11	<i>Confusion matrix</i> model pada <i>fold</i> 5	18
12	<i>Confusion matrix</i> model dari keseluruhan <i>fold</i>	19
13	Hasil sebaran nilai TP, FN, dan FP model dari keseluruhan <i>fold</i>	19

DAFTAR LAMPIRAN

1	Sampel fail <i>extraction</i>	24
2	Step1_bacaabstrak.ipynb	24

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah

b. Pengutipan tidak merugikan kepentingan yang wajar IPB University

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.