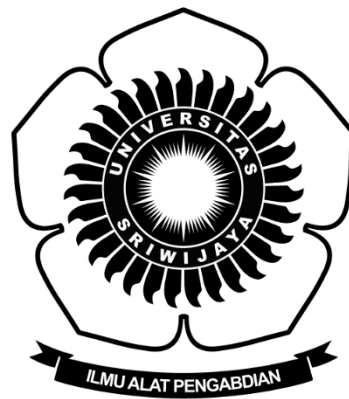


**IDENTIFIKASI KOMENTAR SPAM DI FACEBOOK
MENGUNAKAN *LONG SHORT-TERM MEMORY* (LSTM)**

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh :

Amos Augusto Silangit

NIM : 09021281924051

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024**

LEMBAR PENGESAHAN SKRIPSI

**IDENTIFIKASI KOMENTAR SPAM DI FACEBOOK MENGGUNAKAN
LONG SHORT-TERM MEMORY (LSTM)**

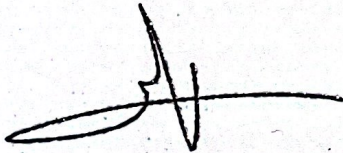
Oleh:

Amos Augusto Silangit

NIM: 09021281924051

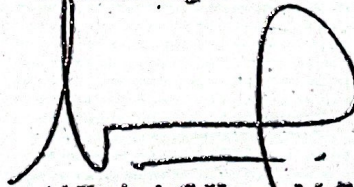
Palembang, 15 Februari 2024

Pembimbing I



Dr. Abdiansah, S.Kom., M.Cs
NIP. 198410012009121005

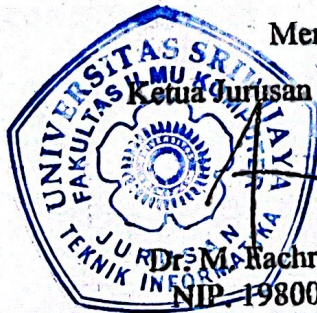
Pembimbing II



Al Farissi, S.Kom., M.Cs.
NIP. 198512152014041001

Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. M. Rachrurrozi, S.Si., M.T.
NIP. 198005222008121002

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Selasa tanggal 19 Desember 2023 telah dilaksanakan ujian komprehensif Skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Amos Augusto Silangit
NIM : 09021281924051
Judul : Identifikasi Komentar Spam di Facebook menggunakan *Long Short-Term Memory* (LSTM)

dan dinyatakan **LULUS**

1. Ketua

Desty Rodiah, M.T.
NIP 198912212020122011

2. Penguji

Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

3. Pembimbing I

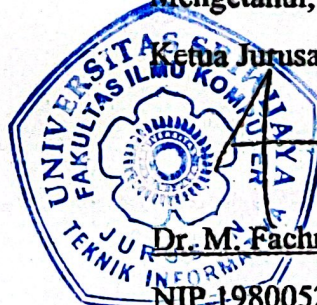
Dr. Abdiansah, S.Kom., M.Cs.
NIP 198410012009121005

4. Pembimbing II

Al Farissi, S.Kom., M.Cs.
NIP 198512152014041001

Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. M. Fachrullozzi, S.Si., M.T.

NIP 198005222008121002

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Amos Augusto Silangit

NIM : 09021281924051

Program Studi : Teknik Informatika

Judul : Identifikasi Komentar Spam di Facebook menggunakan Long Short-Term Memory (LSTM)

Hasil Pengecekan *Software Turnitin* : 13%

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

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari siapapun.



Palembang, 16 Desember 2023



Augusto Silangit

NIM. 09021281924051

MOTTO DAN PERSEMBAHAN

“Great things are not done by impulse, but by a series of small things brought together.”

-Vincent Van Gogh

Kupersembahkan karya tulis ini kepada:

- Tuhan YME
- Orang Tua dan Keluargaku
- Dosen Pembimbing
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Spam comments on social media platforms, especially Facebook, have become an escalating issue with the growth of users and online interactions. This research aims to develop a spam comment identification using the Long Short-Term Memory (LSTM) approach, a recurrent neural network model capable of understanding the temporal context in textual data. The study involves collecting comment data from Facebook, and LSTM is employed to extract temporal patterns embedded in each comment. Evaluation is conducted to assess the effectiveness of this method in distinguishing between spam and non-spam comments, considering performance aspects such as accuracy, precision, recall, and F-measure. The data utilized consists of 800 comments from a Facebook post. Testing involves 6 experimental scenarios, and the best results are obtained with hyperparameters: layer dropout value of 0, 64 LSTM neurons, a learning rate of 10^{-3} , LSTM dropout value of 0.5, batch size of 8, and epoch value of 10. The results show an accuracy of 87.50%, precision of 86.98%, recall of 87.55%, and an F-measure of 87.21%.

Keyword : *Spam, Identification, Facebook, Long Short-Term Memory*

ABSTRAK

Komentar spam di platform sosial media, khususnya Facebook, menjadi permasalahan yang semakin meruncing seiring dengan pertumbuhan pengguna dan interaksi online. Penelitian ini bertujuan untuk mengembangkan identifikasi komentar spam menggunakan pendekatan Long Short-Term Memory (LSTM), sebuah model jaringan saraf rekuren yang mampu memahami konteks temporal dalam data teks. Penelitian ini melibatkan pengumpulan data komentar dari Facebook, dan LSTM digunakan untuk mengekstrak pola temporal yang terkandung dalam setiap komentar. Evaluasi dilakukan terhadap keefektifan metode ini dalam membedakan antara komentar spam dan non-spam, dengan memperhatikan aspek-aspek kinerja seperti akurasi, presisi, recall, dan F-measure. Data yang digunakan didapatkan dari komentar-komentar suatu post di Facebook sebanyak 800 data komentar. Pengujian menggunakan 6 skenario percobaan hasil pengujian terbaik diperoleh dengan *hyper parameter* nilai layer *dropout* sebesar 0, jumlah neuron LSTM sebesar 64, nilai *learning rate* sebesar 10^{-3} , nilai *dropout* LSTM sebesar 0.5, nilai *batch size* sebesar 8, dan nilai *epoch* sebesar 10. Hasil didapatkan dengan nilai *accuracy* sebesar 87.50%, *precision* sebesar 86.98%, *recall* sebesar 87.55%, dan *f-measure* sebesar 87.21%.

Kata Kunci : *Spam*, Identifikasi, Facebook, *Long Short-Term Memory*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan YME, atas berkat dan karunia-Nya sehingga penulisan skripsi dengan judul “Identifikasi Komentar Spam di Facebook menggunakan Long Short-Term Memory (LSTM)” dapat diselesaikan dengan baik sebagai salah satu syarat dalam menyelesaikan pendidikan Program Sarjana Strata-1 Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari bahwa tugas akhir ini tidak mungkin terselesaikan tanpa adanya dukungan dan bantuan baik itu secara materil maupun moril dari berbagai pihak selama penyusunan tugas akhir ini. Penulis ingin menyampaikan ucapan terima kasih kepada :

1. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Bapak Dr. M. Fachrurozzi, S.Si., M.T. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Abdiansah, S.Kom., M.CS. dan Bapak Al Farissi. S.Kom., M.Comp.Sc selaku Dosen Pembimbing Skripsi, atas bimbingan dan arahnya selama pengerjaan Tugas Akhir penulis.
4. Bapak Osvari Arsalan. S.Kom. M.T. selaku Dosen Pembimbing Akademik, atas bimbingan yang telah diberikan selama masa perkuliahan.
5. Seluruh dosen dan staf Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

6. Orang Tua, keluarga, dan teman-teman atas dukungan yang selalu diberikan.
7. Anros Monica Hutagalung, yang telah menemani, membantu, dan menjadi tempat berkeluh kesah serta penyemangat dari awal penulisan Tugas Akhir hingga saat ini.
8. Josie, Aldi, Abdul, Alfath, Asabil, dan Dori yang telah kebersamai selama perkuliahan serta teman-teman seperjuangan IF Reg C 2019 yang telah membantu dalam pengerjaan proses Tugas Akhir ini.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman. Oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Palembang, 11 Oktober 2023

Penulis



Amos Augusto Silangit

09021281924051

DAFTAR ISI

Halaman

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI	iii
HALAMAN PERNYATAAN.....	iv
MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xvi
DAFTAR LAMPIRAN.....	xviii
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan.....	II-1
2.2 Landasan Teori	II-1
2.2.1 Spam.....	II-1
2.2.2 Text Preprocessing	II-2

2.2.3	Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency)	II-6
2.2.4	Long-Short Term Memory	II-7
2.2.5	Confusion Matrix	II-11
2.3	<i>Rational Unified Process</i> (RUP)	II-13
2.4	Penelitian Lain yang Relevan	II-15
2.5	Kesimpulan	II-16
BAB III METODOLOGI PENELITIAN		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.3	Tahapan Penelitian	III-2
3.3.1	Kerangka Kerja	III-4
3.3.2	Kriteria Pengujian	III-7
3.3.3	Format Data Pengujian	III-7
3.3.4	Alat yang Digunakan dalam Pelaksanaan Penelitian	III-8
3.3.5	Pengujian Penelitian	III-9
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan	III-9
3.4	Metode Pengembangan Perangkat Lunak	III-10
3.4.1	Fase Insepsi (<i>Inception</i>)	III-10
3.4.2	Fase Elaborasi (<i>Elaboration</i>)	III-11
3.4.3	Fase Konstruksi (<i>Construction</i>)	III-11
3.4.4	Fase Transisi (<i>Transition</i>)	III-12
3.5	Kesimpulan	III-12
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Kebutuhan Sistem	IV-2
4.2.2	Analisis dan Perancangan	IV-3
4.2.3	Implementasi	IV-13
4.3	Fase Elaborasi	IV-19
4.3.1	Pemodelan Bisnis	IV-19
4.3.2	Perancangan Data	IV-19

4.3.3	Perancangan Antarmuka	IV-19
4.3.4	Kebutuhan Sistem	IV-23
4.3.5	Diagram Aktivitas	IV-23
4.3.6	Diagram <i>Sequence</i>	IV-27
4.4	Fase Konstruksi	IV-30
4.4.1	Diagram Kelas.....	IV-30
4.4.2	Implementasi	IV-31
4.5	Fase Transisi.....	IV-35
4.5.1	Pemodelan Bisnis	IV-35
4.5.2	Kebutuhan	IV-35
4.5.3	Analisis dan Perancangan	IV-36
4.5.4	Implementasi	IV-37
4.6	Kesimpulan.....	IV-39
BAB V HASIL DAN PEMBAHASAN		V-1
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Hasil Konfigurasi 1	V-5
5.2.3	Hasil Konfigurasi 2	V-6
5.2.4	Hasil Konfigurasi 3	V-7
5.2.5	Hasil Konfigurasi 4	V-8
5.2.6	Hasil Konfigurasi 5	V-9
5.2.7	Hasil Konfigurasi 6	V-10
5.2.8	Hasil Konfigurasi 7	V-11
5.2.9	Hasil Konfigurasi 8	V-12
5.2.10	Hasil Konfigurasi 9	V-13
5.2.11	Hasil Konfigurasi 10	V-14
5.2.12	Hasil Konfigurasi 11	V-15
5.2.13	Hasil Konfigurasi 12	V-16
5.3	Analisis Hasil Penelitian	V-17
5.4	Kesimpulan.....	V-22
BAB VI KESIMPULAN DAN SARAN.....		VI-1

6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA		xix
LAMPIRAN.....		xxii

DAFTAR TABEL

	Halaman
Tabel II-1. Confusion Matrix.....	II-11
Tabel III-2. Rancangan Tabel Confusion Matrix Hasil Klasifikasi	III-8
Tabel III-3. Rancangan Tabel Hasil Pengujian	III-8
Tabel III-4. Format Hasil Pengujian	III-10
Tabel IV-1. Kebutuhan Fungsional	IV-2
Tabel IV-2. Kebutuhan Non Fungsional	IV-2
Tabel IV-3. Post di Facebook.....	IV-4
Tabel IV-4. Data Komentar di Facebook	IV-5
Tabel IV-5. Hasil Proses Cleaning Data Komentar.....	IV-6
Tabel IV- 6. Hasil Proses Case Folding Data Komentar	IV-7
Tabel IV-7. Hasil Proses Tokenizing Data Komentar	IV-8
Tabel IV-8. Hasil Proses Normalisasi Data Komentar.....	IV-9
Tabel IV-9. Hasil Proses Stopword Removal Data Komentar	IV-10
Tabel IV-10. Hasil Proses Stemming Data Komentar.....	IV-11
Tabel IV-11. Definisi Aktor pada Use Case.....	IV-14
Tabel IV-12. Tabel Definisi Use Case.....	IV-14
Tabel IV-13. Skenario klasifikasi spam pada komentar	IV-15
Tabel IV-14. Skenario Pra-pengolahan Dataset	IV-16
Tabel IV-15. Skenario input dataset komentar	IV-17
Tabel IV-16. Skenario klasifikasi dataset dan kinerja model	IV-18
Tabel IV- 17. Perancangan Data.....	IV-19
Tabel IV-18. Daftar kelas	IV-32
Tabel IV-19. Daftar file antarmuka	IV-33
Tabel IV-20. Rencana Pengujian Use Case Klasifikasi Spam Komentar	IV-36
Tabel IV-21. Rencana Pengujian Use Case Pra-pengolahan Dataset.....	IV-36
Tabel IV-22. Rencana Pengujian Use Case Input Dataset Komentar	IV-37
Tabel IV-23. Rencana Pengujian Use Case Klasifikasi Dataset dan Kinerja Model	IV-37

Tabel IV-24. Pengujian Use Case Klasifikasi Spam Komentar	IV-37
Tabel IV-25. Pengujian Use Case Pra-pengolahan Dataset.....	IV-38
Tabel IV-26. Pengujian Use Case Input Dataset Komentar	IV-38
Tabel IV-27. Pengujian Use Case Input Dataset Komentar	IV-39
Tabel V-1. Kofigurasi parameter konstan	V-1
Tabel V-2. Konfigurasi pada Skenario 1	V-2
Tabel V-3. Konfigurasi pada Skenario 2	V-2
Tabel V-4. Konfigurasi pada Skenario 3	V-3
Tabel V-5. Konfigurasi pada Skenario 4	V-3
Tabel V-6. Konfigurasi pada Skenario 5	V-4
Tabel V-7. Konfigurasi pada Skenario 6	V-4
Tabel V-8. Hasil Kinerja Pengujian Model 1	V-5
Tabel V-9. Hasil Kinerja Pengujian Model 2	V-6
Tabel V-10. Hasil Kinerja Pengujian Model 3	V-7
Tabel V-11. Hasil Kinerja Pengujian Model 4	V-8
Tabel V-12. Hasil Kinerja Pengujian Model 5	V-9
Tabel V-13. Hasil Kinerja Pengujian Model 6	V-10
Tabel V-14. Hasil Kinerja Pengujian Model 7	V-11
Tabel V-15. Hasil Kinerja Pengujian Model 8	V-12
Tabel V-16. Hasil Kinerja Pengujian Model 9	V-13
Tabel V-17. Hasil Kinerja Pengujian Model 10	V-14
Tabel V-18. Hasil Kinerja Pengujian Model 11	V-15
Tabel V-19. Hasil Kinerja Pengujian Model 12	V-16
Tabel V-20. Perbandingan kinerja model berdasarkan pengujian data uji	V-21

DAFTAR GAMBAR

	Halaman
Gambar II-1. Proses Cleaning	II-3
Gambar II-2. Proses Case Folding.....	II-3
Gambar II-3. Proses Tokenizing.....	II-4
Gambar II-4. Proses Normalisasi.....	II-4
Gambar II-5. Proses Filtering	II-5
Gambar II-6. Proses Stemming	II-5
Gambar II-7. Unit LSTM (Olah, 2015)	II-8
Gambar II- 8. Penjelasan Operasi LSTM (Olah, 2015).....	II-8
Gambar II- 9. Ilustrasi Arsitektur RUP (Anwar, 2014).....	II-13
Gambar III- 1. Tahapan Penelitian.....	III-2
Gambar III-2. Kerangka Kerja Penelitian	III-5
Gambar IV-1. Diagram Use Case.....	IV-13
Gambar IV-2. Rancangan Antarmuka Halaman Awal	IV-20
Gambar IV-3. Rancangan Antarmuka Halaman Prediksi Komentar.....	IV-21
Gambar IV-4. Rancangan Antarmuka Hasil Halaman Prediksi Komentar	IV-21
Gambar IV-5. Rancangan Antarmuka Halaman Preprocessing Dataset	IV-22
Gambar IV- 6. Rancangan Antarmuka Halaman Train Dataset	IV-22
Gambar IV-7. Diagram Aktivitas Halaman Klasifikasi Input Komentar	IV-24
Gambar IV-8. Diagram Aktivitas Halaman Proses Data.....	IV-25
Gambar IV-9. Diagram Aktivitas Halaman Input Dataset	IV-25
Gambar IV-10. Diagram Aktivitas Halaman Kinerja Model	IV-26
Gambar IV-11. Diagram Sequence Klasifikasi Input Komentar.....	IV-27
Gambar IV-12. Diagram Sequence Input Dataset	IV-28
Gambar IV-13. Diagram Sequence Proses Data.....	IV-29
Gambar IV-14. Diagram Sequence Kinerja.....	IV-30
Gambar IV-15. Kelas Diagram.....	IV-31
Gambar IV-16. Antarmuka Halaman Utama.....	IV-33
Gambar IV-17. Antarmuka Halaman Prediksi Komentar	IV-34

Gambar IV-18. Antarmuka Halaman Hasil Prediksi Komentar	IV-34
Gambar IV- 19. Antarmuka Halaman Train Dataset.....	IV-35
Gambar V-1. Confusion Matrix Model 1	V-5
Gambar V-2. Confusion Matrix Model 2	V-6
Gambar V-3. Confusion Matrix Model 3	V-7
Gambar V-4. Confusion Matrix Model 4	V-8
Gambar V-5. Confusion Matrix Model 5	V-9
Gambar V-6. Confusion Matrix Model 6	V-10
Gambar V-7. Confusion Matrix Model 7	V-11
Gambar V-8. Confusion Matrix Model 8	V-12
Gambar V-9. Confusion Matrix Model 9	V-13
Gambar V-10. Confusion Matrix Model 10	V-14
Gambar V-11. Confusion Matrix Model 11	V-15
Gambar V-12. Confusion Matrix Model 12	V-16
Gambar V-13. Perbandingan Hasil Skenario 1.....	V-17
Gambar V-14. Perbandingan Hasil Skenario 2.....	V-18
Gambar V-15. Perbandingan Hasil Skenario 3.....	V-18
Gambar V-16. Perbandingan Hasil Skenario 4.....	V-19
Gambar V-17. Perbandingan Hasil Skenario 5.....	V-20
Gambar V-18. Perbandingan Hasil Skenario 6.....	V-20
Gambar V-19. Perbandingan Akurasi Pengujian Data	V-22

DAFTAR LAMPIRAN

Lampiran 1. Kode Program	xxii
---------------------------------------	------

BAB I

PENDAHULUAN

1.1 Pendahuluan

Dalam Bab pendahuluan ini akan membahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang diuraikan secara singkat. Topik yang akan dibahas dalam penelitian ini adalah “Identifikasi Komentar Spam di Facebook menggunakan *Long Short-Term Memory (LSTM)*”.

1.2 Latar Belakang Masalah

Penggunaan internet kini menjadi suatu kebutuhan yang penting dalam kehidupan untuk berkomunikasi. Di antara banyak fasilitas yang tersedia di internet, terdapat media sosial sebagai salah satu bentuknya (Wibisono et al., 2020). Dibalik kemajuan teknologi tersebut, tentunya cara orang berkomunikasi banyak mengalami perubahan, baik dalam mengkomunikasikan hal positif, maupun hal negatif (Djohan, 2022). Seiring majunya teknologi dengan sangat cepat, manusia menciptakan sebuah ide untuk membuat sebuah *platform* dimana sesama manusia dapat berinteraksi dan berkomunikasi dengan orang yang jauh hanya dengan menggunakan *internet* (Wolf et al., 2018).

Pada saat ini, masyarakat Indonesia banyak yang menggunakan Internet untuk mengakses media sosial salah satunya yaitu Facebook, sebagai wadah untuk tetap terhubung dan berinteraksi dengan sesama mereka di dunia maya terutama

bagi publik figur seperti politisi dan selebriti di Indonesia (Rachmat & Lukito, 2017).

Alasan utama yang membuat Facebook menjadi salah sosial media yang ramai digunakan oleh orang-orang yaitu dikarenakan Facebook memberikan kemudahan bagi pengguna aplikasi mereka dalam membagikan status, baik dalam bentuk teks maupun foto secara langsung dengan menggunakan *smartphone*. Hal ini sangat diperhatikan mengingat mayoritas pengguna media sosial adalah kalangan muda yang aktif di *platform* tersebut. Walaupun demikian, di samping keunggulan tersebut, terdapat kendala yang cukup mengganggu, yakni tingginya jumlah komentar yang dapat dianggap sebagai spam pada suatu *post*, baik itu berupa teks atau foto yang diunggah di Facebook. Jumlah komentar spam ini cenderung meningkat terutama pada akun Facebook milik selebriti atau orang terkenal karena jumlah pengikutnya juga semakin besar (Rachmat & Lukito, 2017). Pengguna yang mengikuti publik figur yang mereka sukai cenderung memberikan dukungan berupa *like* serta dapat melakukan komentar pada setiap *post* atau status terbaru yang diunggah oleh publik figur tersebut. Namun, tidak semua komentar yang ditemukan pada status yang diunggah bersifat relevan dengan topik yang disajikan. Contohnya seperti komentar yang bertujuan untuk mempromosikan bisnis mereka, seperti promosi atau penjualan, menyertakan tautan spam, serta berbagai hal lain yang tidak berkaitan dengan konten status yang dibagikan dan secara nyata dapat menyebabkan ketidaknyamanan bagi pengguna lain (Rachmat & Lukito, 2017).

Pengguna mempunyai pilihan atau opsi untuk mengatasi permasalahan tersebut dengan cara menghapus komentar yang dianggap *spam* secara manual dan menggunakan fitur dari Facebook untuk melaporkan suatu komentar apabila komentar tersebut bersifat *spam*. Kedua hal tersebut hanya dapat dilakukan secara manual hingga memakan waktu yang cukup lama serta perlu dilakukan pemeriksaan satu per satu. Upaya terakhir untuk mengatasi komentar *spam* adalah dengan menggunakan pengaturan yang memberikan akses kepada Facebook untuk secara otomatis menghapus komentar yang mengandung kata-kata tertentu yang telah ditentukan pengguna sebagai *spam*. Namun, sangat disayangkan semua upaya tersebut hanya berlaku untuk penggunaan dalam bahasa Inggris dan tidak dapat diterapkan pada bahasa Indonesia.

Dengan dasar informasi tersebut, penelitian ini akan mengembangkan sebuah sistem yang mampu mengenali komentar spam dalam bahasa Indonesia. Pengumpulan data diambil dari komentar-komentar pada Facebook, khususnya dari akun dengan jumlah pengikut terbanyak di Indonesia. Metode pengklasifikasian yang diterapkan dalam penelitian ini adalah Long Short-Term Memory (LSTM).

Metode LSTM merupakan variasi dari *Recurrent Neural Network* (RNN), dan keunggulan metode LSTM dibandingkan dengan metode lainnya terletak pada kemampuannya untuk memahami konteks jangka panjang, penanganan urutan dan ketergantungan kontekstual, kemampuan menangkap hubungan semantik, pengelolaan waktu, pencegahan *overfitting* dengan Dropout, fleksibilitas dalam desain arsitektur, pengelolaan data teks yang panjang, dan efektivitas pada tugas pemrosesan bahasa alami (NLP) (Isnain et al., 2022).

1.3 Rumusan Masalah

Berdasarkan permasalahan yang dijelaskan pada latar belakang maka rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana melakukan identifikasi komentar *spam* di Facebook menggunakan Long Short-Term Memory (LSTM).
2. Bagaimana mengukur kinerja metode LSTM dalam mengidentifikasi komentar *spam* pada aplikasi Facebook?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menghasilkan perangkat lunak yang dapat mengidentifikasi komentar *spam* di Facebook menggunakan *Long Short-Term Memory* (LSTM).
2. Mengetahui kinerja metode LSTM dalam mengidentifikasi komentar *spam* di Facebook.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah :

1. Hasil penelitian ini dapat digunakan untuk mengidentifikasi komentar *spam* di Facebook.
2. Hasil laporan penelitian dapat menjadi sumber referensi bagi peneliti lainnya untuk mengidentifikasi komentar *spam* menggunakan metode *Deep Learning*.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah :

1. Menggunakan data dari akun yang memiliki banyak pengikut.
2. Komentar yang digunakan merupakan postingan 3 bulan dari *post* terbaru.
3. Dalam penelitian ini *dataset* yang digunakan berbahasa Indonesia.
4. Format *dataset* yang digunakan adalah csv.
5. Label kelas komentar dalam penelitian ini yaitu menggunakan dua kelas yaitu *spam* dan *non spam*.
6. Sistem yang dibangun merupakan sistem berbasis *Website* dan hanya mengidentifikasi komentar *spam* dan *non spam*.

1.7 Sistematika Penulisan

Dibawah ini adalah tata cara penulisan yang digunakan dalam penyusunan tugas akhir ini:

BAB 1. PENDAHULUAN

Bab ini akan membahas tentang latar belakang penelitian, rumusan masalah penelitian, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penelitian.

BAB II. KAJIAN LITERATUR

Bab ini akan membahas teori-teori dasar yang digunakan dalam penelitian, seperti definisi Spam, *Text Preprocessing*, Pembobotan TF-IDF, LSTM, dan

Confusion Matrix, serta memaparkan beberapa literatur penelitian terdahulu yang relevan dengan topik penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini akan membahas tentang tahapan-tahapan yang dikerjakan pada penelitian ini, seperti pengumpulan data, analisis data dan perancangan perangkat lunak. Tahapan yang dijelaskan disusun secara rinci berdasarkan kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini menjelaskan tentang perencanaan dan pelaksanaan perangkat lunak, termasuk analisis kebutuhan perangkat lunak, perancangan serta pembuatan perangkat lunak, dan pengujian sesuai dengan rencana dan kebutuhan yang telah ditetapkan dalam penelitian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menguraikan hasil penelitian sesuai dengan format yang telah ditetapkan. Selain itu, bab ini merinci analisis hasil penelitian yang menjadi dasar untuk menyimpulkan temuan dalam penelitian ini.

BAB VI. HASIL DAN ANALISIS PENELITIAN

Bagian ini menampilkan rangkuman dari seluruh paparan pada bab sebelumnya dan mengandung saran-saran yang diharapkan dapat bermanfaat bagi penelitian dengan topik yang serupa.

1.8 Kesimpulan

Pada Bab ini telah dijelaskan mengenai dasar penelitian yang melibatkan latar belakang permasalahan, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

DAFTAR PUSTAKA

- Agastya, I. M. A. (2018). Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film. *Jurnal Tekno Kompak*.
<http://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/70>
- Albab, M. U., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*.
<https://journals.usm.ac.id/index.php/transformatika/article/view/5374>
- Djohan, S. D. (2022). *Implementasi Convolutional Neural Network untuk Klasifikasi Jenis Cyberbullying dalam Komentar Unggahan Instagram*. Universitas Multimedia Nusantara.
- Houdt, G. Van. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(8), 5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>
- Hussain, N., Mirza, H. T., Rasool, G., Hussain, I., & ... (2019). Spam review detection techniques: A systematic literature review. *Applied Sciences*.
<https://www.mdpi.com/2076-3417/9/5/987>
- Isnain, A. R., Sulistiani, H., Hurohman, B. M., & ... (2022). Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen. ... (*Jurnal Edukasi Dan ...*).
<https://jurnal.untan.ac.id/index.php/jepin/article/view/54704>
- Kaushal, A., White, L., Innes, M., & Kumar, R. (2020). WordTokenizers. jl: Basic tools for tokenizing natural language in Julia. *Journal of Open Source ...*.
<https://doi.org/10.21105/joss.01956>
- Ladani, D. J., & Desai, N. P. (2020). Stopword identification and removal techniques on tc and ir applications: A survey. *2020 6th International Conference on ...*. <https://ieeexplore.ieee.org/abstract/document/9074166/>
- Larasati, R. (2015). *Klasifikasi Teks Dengan Menggunakan Algoritma K-Nearest Neighbor Pada Dokumen Tugas Akhir*. Universitas Widyatama.
- Muhammad, J. G., Herdiani, A., & ... (2019). Identifikasi Spam Tweet Komentar Pada Twitter Berbasis Ontologi (studi Kasus: Tweet/Caption Di Twitter Dengan Tema "pilkpres 2019"). *EProceedings ...*.
<https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/viewFile/9881/9740>
- Muhammad, P. F., Kusumaningrum, R., & ... (2021). Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel

- reviews. *Procedia Computer ...*
<https://www.sciencedirect.com/science/article/pii/S1877050921000752>
- Nuraliza, H., Pratiwi, O. N., & ... (2022). Analisis Sentimen IMDb Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance. *Jurnal Mirai ...*
<https://www.journal.stieamkop.ac.id/index.php/mirai/article/view/2222>
- Pakdeetrakulwong, U. (2018). Semantic web-based approach to support rational unified process software development. ... & *Communications, Cloud & Big Data ...* <https://ieeexplore.ieee.org/abstract/document/8560080/>
- Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. ... *Conference on Machine ...*
<https://proceedings.mlr.press/v28/pascanu13.html>
- Prasanti, A. A., Fauzi, M. A., & Furqon, M. T. (2018). Klasifikasi teks pengaduan pada sambat online menggunakan metode n-gram dan neighbor weighted k-nearest neighbor (NW-KNN). *Jurnal Pengembangan Teknologi ...* <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/885>
- Prayitno, E., Suprawoto, T., & ... (2021). OPTIMASI HASIL PENCARIAN PADA WEB SCRAPPING MENGGUNAKAN PEMBOBOTAN KATA TF-IDF. *Journal of Innovation ...*
<https://bajangjournal.com/index.php/JIRK/article/view/822>
- Qaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*.
https://www.researchgate.net/profile/Shahzad-Qaiser/publication/326425709_Text_Mining_Use_of_TF-IDF_to_Examine_the_Relevance_of_Words_to_Documents/links/5b4cd57fa6fdcc8dae245aa3/Text-Mining-Use-of-TF-IDF-to-Examine-the-Relevance-of-Words-to-Documents.pdf
- Rachmat, A., & Lukito, Y. (2017). Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes. *Ultimatics - Jurnal Teknik Informatika, IX*, 50–58. <https://doi.org/10.31937/ti.v9i1.564>
- Rehmer, A., & Kroll, A. (2020). On the vanishing and exploding gradient problem in Gated Recurrent Units. *IFAC-PapersOnLine*.
<https://www.sciencedirect.com/science/article/pii/S2405896320317481>
- Sahoo, S. R., Gupta, B. B., Peraković, D., & ... (2022). Spammer detection approaches in online social network (OSNs): a survey. ... *Systems in Industry 4.0*. https://doi.org/10.1007/978-3-030-90462-3_11

- Shafiee, S., Wautelet, Y., Hvam, L., Sandrin, E., & ... (2020). Scrum versus Rational Unified Process in facing the main challenges of product configuration systems development. *Journal of Systems and ...*
<https://www.sciencedirect.com/science/article/pii/S0164121220301643>
- Shankar, R., Sarojini, B. K., Mehraj, H., & ... (2023). Impact of the learning rate and batch size on NOMA system using LSTM-based deep neural network. *The Journal of ...* <https://doi.org/10.1177/15485129211049782>
- Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1—learning rate, batch size, momentum, and weight decay. *ArXiv Preprint ArXiv:1803.09820*. <https://arxiv.org/abs/1803.09820>
- Srividya, K., & Sowjanya, A. M. (2019). Aspect based sentiment analysis using POS tagging and TFIDF. *International Journal of Engineering ...*
https://www.researchgate.net/profile/Mary-Sowjanya/publication/336084735_Aspect_Based_Sentiment_Anaysis_using_POS_Tagging_and_TFIDF/links/5d8d9290a6fdcc25549e8231/Aspect-Based-Sentiment-Analysis-using-POS-Tagging-and-TFIDF.pdf
- Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM—a tutorial into long short-term memory recurrent neural networks. *ArXiv Preprint ArXiv:1909.09586*. <https://arxiv.org/abs/1909.09586>
- Wibisono, A. D., Rizkiono, S. D., & ... (2020). Filtering Spam Email Menggunakan Metode Naive Bayes. *Telefortech: Journal Of ...*
<https://ejurnal.teknokrat.ac.id/index.php/telefortech/article/view/685>
- Wolf, M., Sims, J., & Yang, H. (2018). *Social media? What social media?* [aisel.aisnet.org. https://aisel.aisnet.org/ukais2018/3/](https://aisel.aisnet.org/aisel.aisnet.org/ukais2018/3/)