



**Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di
Indonesia Menggunakan Algoritma Long Short Term Memory**

TUGAS AKHIR

Juan Antonio
41518110166

UNIVERSITAS
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022



**Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di
Indonesia Menggunakan Algoritma Long Short Term Memory**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
JUAN ANTONIO
41518110166

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518110166

Nama : Juan Antonio

Judul Tugas Akhir : Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 02 Agustus 2022



Juan Antonio

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Juan Antonio
NIM : 41518110166
Judul Tugas Akhir : Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 02 Agustus 2022



Juan Antonio

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Juan Antonio
 NIM : 41518110166
 Judul Tugas Akhir : Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	√
		Jurnal Nasional Terakreditasi	√		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)			
	ISSN	: 2580-0760			
	Link Jurnal	: https://jurnal.iaii.or.id/index.php/RESTI			
	Link File Jurnal Jika Sudah di Publish	:			

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
 Dosen Pembimbing TA



Muhammad Rifqi, S.Kom, M.Kom

Jakarta, 16 Juli 2022



Juan Antonio

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110166
Nama : Juan Antonio
Judul Tugas Akhir : Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 02 Agustus 2022



(Harni Kusniyati, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110166
Nama : Juan Antonio
Judul Tugas Akhir : Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 02 Agustus 2022



(Sabar Rudiarto, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110166
Nama : Juan Antonio
Judul Tugas Akhir : Analisis Sentimen Terhadap Ulasan Pada Aplikasi
Provider Internet di Indonesia Menggunakan
Algoritma Long Short Term Memory

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 02 Agustus 2022



(Eliyani, Dr. Ir.)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41518110166
Nama : Juan Antonio
Judul Tugas Akhir : Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 02 Agustus 2022

Menyetujui,



(Muhammad Rifqi, S.Kom, M.Kom)
Dosen Pembimbing

Mengetahui,



(Wawan Ganawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Ka. Prodi Teknik Informatika

KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah S.W.T. atas berkat serta rahmat-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir dengan judul “Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory” yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Jurusan Teknik Informatika Universitas Mercu Buana.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Muhammad Rifqi, S.Kom, M.Kom, selaku Pembimbing Tugas Akhir yang telah membimbing dalam penyusunan laporan ini.
2. Kedua Orang Tua atas doa serta dukungan moral yang tiada henti kepada penulis dalam menyelesaikan tugas akhir ini.
3. Bapak Achmad Kodar, Drs. MT, selaku dosen pembimbing akademik yang telah membantu dalam hal akademik penulis.
4. Bapak Ir. Emil R. Kaburuan. Ph.D.. IPM., selaku kepala program studi Teknik informatika Universitas Mercubuanya.
5. Semua dosen mata kuliah yang tidak dapat disebutkan satu persatu yang telah memberikan ilmu yang bermanfaat bagi penulis.
6. Teman-teman seperjuangan, serta semua pihak yang tidak bisa disebutkan satu persatu, terima kasih atas dukungan serta kerjasama yang telah diberikan kepada penulis.

Akhir kata, penulis menyadari bahwa penelitian ini jauh dari kesempurnaan. Oleh sebab itu, kritik maupun saran selalu penulis harapkan demi menghasilkan hasil terbaik dari penelitian ini. Besar harapan penulis, semoga penelitian ini dapat memberikan manfaat sekaligus menambah pengetahuan bagi berbagai pihak.

Jakarta, 15 Juli 2022

Juan Antonio

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS.....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR ..	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR	iv
LEMBAR PERSETUJUAN PENGUJI	v
LEMBAR PENGESAHAN	viii
ABSTRAK.....	ix
ABSTRACT.....	x
KATA PENGANTAR	xi
DAFTAR ISI	xii
NASKAH JURNAL	1
KERTAS KERJA	8
BAB 1. LITERATUR REVIEW.....	9
BAB 2. ANALISIS DAN PERANCANGAN	21
BAB 3. SOURCE CODE.....	23
BAB 4. DATASET	31
BAB 5. TAHAPAN EKSPERIMEN	33
BAB 6. HASIL SEMUA EKSPERIMEN	42
DAFTAR PUSTAKA.....	44
LAMPIRAN DOKUMEN HAKI	46
LAMPIRAN KORESPONDENSI.....	49

NASKAH JURNAL

Terakreditasi SINTA Peringkat 2

Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 158/E/KPT/2021
masa berlaku mulai Volume 5 Nomor 2 Tahun 2021 sampai Volume 10 Nomor 1 Tahun 2026Terbit online pada laman web jurnal: <http://jurnal.iaii.or.id>

JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 6 No. x (2022) x - x

ISSN Media Elektronik: 2580-0760

Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma Long Short Term Memory

Juan Antonio¹, Muhammad Rifqi¹^{1,2}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana Jakarta
141518110166@student.mercubuana.ac.id, m.rifqi@mercubuana.ac.id**Abstract**

Internet is a basic need of people in the world. And also due to the COVID-19 outbreak, the internet has become something that people really need for activities such as studying, working, and shopping. In this case, internet providers try to help provide internet services for the community, especially in Indonesia. However, the development of technology is growing very rapidly from time to time. One of them is making an application. This is useful to be able to support the quality of a company and also make it easier for customers. There are many positive and negative reviews related to applications on Google Playstore, especially for internet provider applications. This study will conduct a sentiment analysis of public reviews of the internet provider application. With the aim of helping the company to further improve its applications. The review data that will be taken is the myIndiHome and My Firstmedia applications which are two internet providers in Indonesia and the data taken is 10000 review data. Then, it will be tested with the Long Short Term Memory algorithm method to find out the results of the accuracy of the review. The initial stage in this research is to get review data by scrapping and after that the preprocessing stage is carried out which aims to clean the data before being tested on the algorithm, in order to get good accuracy results. And the last stage is testing using algorithms. And as a result, the best accuracy value is 77% using the Long Short Term Memory algorithm.

Keywords: Sentiment Analysis, Long Short Term Memory, Internet, Reviews, Provider.

Abstrak

Internet merupakan suatu kebutuhan pokok masyarakat di dunia. Dan juga akibat wabah covid 19, internet menjadi hal yang sangat dibutuhkan oleh masyarakat untuk beraktifitas seperti belajar, bekerja, dan belanja. Dalam hal ini, provider internet mencoba untuk membantu memberikan layanan internet untuk masyarakat khususnya di Indonesia. Tapi, perkembangan teknologi berkembang sangat pesat dari waktu ke waktu. Salah satunya pembuatan sebuah aplikasi. Hal ini berguna untuk dapat menunjang kualitas suatu perusahaan dan juga mempermudah pelanggan. Terdapat banyak ulasan positif dan negatif terkait aplikasi yang ada pada *google playstore* khususnya untuk aplikasi provider internet. Penelitian ini akan melakukan analisis sentimen terhadap ulasan masyarakat terhadap aplikasi provider internet tersebut. Dengan tujuan supaya dapat membantu perusahaan tersebut untuk lebih dapat meningkatkan aplikasi yang dimilikinya. Data ulasan yang akan diambil yaitu pada aplikasi myIndiHome dan My Firstmedia yang dimana merupakan dua provider internet yang ada di Indonesia dan data yang diambil adalah sebanyak 10000 data ulasan. Lalu, akan di uji dengan metode algoritma *Long Short Term Memory* untuk mengetahui hasil dari akurasi terhadap ulasan tersebut. Tahap awal dalam penelitian ini adalah mendapatkan data ulasan dengan cara *scrapping* dan setelah itu dilakukan tahap *preprocessing* yang bertujuan untuk membersihkan data sebelum di uji ke algoritma, supaya mendapatkan hasil akurasi yang baik. Dan, tahap terakhir adalah pengujian menggunakan algoritma. Dan hasilnya, didapatkan nilai akurasi terbaik yaitu 77% menggunakan algoritma *Long Short Term Memory*.

Kata kunci : Analisis Sentimen , Long Short Term Memory , Internet, Ulasan, Provider.

1. Pendahuluan

Internet (kependekan dari *interconnection networking*) secara harfiah ialah sistem global dari seluruh jaringan

komputer yang saling terhubung menggunakan standar Internet Protocol Suite (TCP/IP) untuk melayani miliaran pengguna di seluruh dunia. Internet (*Interconnected Network*) merupakan sekumpulan

jaringan yang saling terhubung dimana jaringan tersebut menyediakan sambungan global informasi. Dalam hal ini komputer yang sebelumnya stand-alone kini dapat berhubungan langsung dengan host-host atau komputer-komputer yang lain[1].

Seiring dengan berkembangnya teknologi terutama internet, internet kian menjadi bagian terpenting bagi masyarakat dunia[2]. Dan saat ini sedang terjadi wabah virus covid19 yang dimana pemerintah menghimbau agar semua orang harus berakifitas dari rumah masing – masing guna untuk meminimalisir penyebaran virus covid19. Para provider internet pun tidak tinggal diam dengan situasi ini. Mereka mulai menyediakan layanan internet untuk rumah – rumah agar masyarakat dapat beraktifitas dari rumah. Berikut merupakan persentase penyedia layanan/provider internet di Indonesia dari survey internet pada APJII tahun 2022.



Gambar 1. Survey Provider Internet

Meningkatnya penggunaan internet akan memberikan dampak terhadap kehidupan. Selain mendapatkan informasi, dalam dunia internet pun kita dapat menyampaikan opini atau saran terhadap informasi terkini[3]. Opini atau saran sering digunakan banyak orang sebagai komentar atau ulasan untuk memberikan penilaian atas suatu hal. Komentar atau ulasan yang diberikan dapat bernilai positif, negatif, dan juga netral.

Sentiment analysis merupakan bagian dari *Natural Language Processing* (NLP). Teknik ini sangat baik untuk menentukan hasil ulasan (*review*) dari sebuah opini pengguna (*user*). *Sentiment analysis* adalah salah satu teknik ekstraksi dari sebuah informasi terhadap sebuah isu dan kejadian. Teknik ini digunakan untuk menemukan opini dan paparan terhadap suatu isu atau kejadian dalam bentuk teks[4]. Pada *Sentiment analysis* terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan *sentimen analysis* perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui *feedback* masyarakat atau pun para ahli[5].

Terdapat banyak macam model klasifikasi untuk menentukan suatu sentimen, salah satunya adalah metode Long Short Term Memory (LSTM). LSTM merupakan bagian dari model deep learning yang dapat digunakan untuk melakukan klasifikasi sentimen. Metode ini dapat memproses data secara sekuensial

seperti teks, suara, dan video[6]. Pada penelitian yang dilakukan oleh M. A. Nurrohmat dan Azhari SN. yang berjudul *Sentiment Analysis of Novel Review Using Long Short-Term Memory Method*, LSTM menghasilkan akurasi yang tinggi jika dibandingkan dengan penggunaan metode lainnya dalam proses analisis sentimen[7]. Di sisi lain, Perkembangan teknologi saat ini telah memberikan pengaruh yang sangat besar bagi dunia teknologi informasi dan telekomunikasi. Munculnya beragam aplikasi memberikan pilihan dalam peningkatan kinerja suatu perusahaan atau UMKM, baik yang berbasis desktop, *mobile* maupun web[8]. Karena hal ini para penyedia jasa layanan internet khususnya di Indonesia membuat sebuah aplikasi guna untuk membantu memberikan pelayanan yang lebih baik pada para pelanggannya. Sudah banyak orang yang memberikan ulasan terkait aplikasi para provider internet tersebut.

Dengan menggunakan sentimen analisis, akan dilakukan klasifikasi terhadap ulasan pada aplikasi provider internet tersebut. Yang bertujuan untuk mencari persentase akurasi dan mebandingkan berapa persen ulasan negatif, dan positif. Terdapat beberapa langkah untuk menguji akurasinya. Tahap awal yaitu mengumpulkan data ulasan. Lalu, tahap selanjutnya yaitu *pre-processing* dan labeling. Dan tahap terakhir adalah proses pengujian akurasi dengan metode algoritma *Long Short Term Memory*. Dan hasilnya akan berbentuk nilai persentase.

2. Metode Penelitian

2.1. Literatur Review

Berikut ini merupakan beberapa tinjauan pustaka yang mencakup dari penulisan inti, pembahasan mengenai beberapa istilah dalam tulisan ini.

2.1.1 Analisis Sentimen

Sentiment Analysis adalah merupakan perpaduan dari data mining dan text mining, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi[1]. *Sentiment analysis* merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentimen yang terkandung dalam sebuah opini[1]. Analisis sentimen dilakukan untuk melihat pendapat atau opini yang muncul dari suatu entitas atau objek dan mengolahnya sehingga menghasilkan suatu kecenderungan akan makna dari pendapat tersebut apakah bernilai negatif, positif, atau netral[6].

2.1.2 Text Mining

Text Mining mempunyai pengertian menggali informasi data dari dokumen yang berupa teks, banyaknya data berupa teks yang terdapat pada dokumen bertujuan mencari kata-kata yang dapat mewakili isi dari

dokumen sehingga mendapatkan informasi yang berguna dari sekumpulan dokumen melalui teks dan dapat dilakukan analisa keterhubungan antar dokumen yang bermanfaat untuk tujuan tertentu[9]. *Text mining* juga memiliki definisi menambang data berupa teks dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen[3].

2.1.3 NLP

Natural Language Processing (NLP) adalah salah satu bidang ilmu komputer yang merupakan cabang dari kecerdasan buatan, dan bahasa (linguistik) yang berkaitan dengan interaksi antara komputer dan bahasa alami manusia, seperti bahasa Indonesia atau bahasa Inggris[10]. Tujuan utama dari studi NLP adalah membuat mesin yang mampu mengerti dan memahami makna bahasa manusia lalu memberikan respon yang sesuai[14].

2.1.4 LSTM

Long Short Term Memory (LSTM) merupakan sebuah evolusi dari arsitektur *Recurrent Neural Network* (RNN), di mana pertama kali diperkenalkan oleh Hochreiter & Schmidhuber (1997)[15]. Pada LSTM, nilai eror saat proses *back propagation* dilakukan melalui waktu dan layer akan disimpan, sehingga memungkinkan LSTM untuk menyimpan informasi lebih dari 1000 langkah waktu tanpa mengalami permasalahan *vanishing gradient*[6]. Jaringan LSTM memiliki ukuran yang besar, pada jaringan ini terdapat LSTM cell yang menggantikan posisi hidden layer pada RNN, gunanya untuk menyimpan teks sebelumnya [17]. Pada jaringan LSTM terdapat *forget gate*, *input gate*, *output gate* dan *memory cell* yang akan menghitung nilai keluaran sebagai hidden layer untuk jaringan selanjutnya [16]. Berikut merupakan beberapa rumus dari setiap gate di LSTM :

A. Forget Gate (f_t)

Pada gerbang ini nilai output sebelumnya dengan input saat ini digabung lalu melewati fungsi aktivasi *sigmoid*. Gerbang inilah yang menentukan apakah informasi sebelumnya akan dilupakan atau tidak. Kemudian informasi ini dilanjutkan ke *memory cell* atau *cell state* [16].

$$f_t = \sigma(W_f \times [x_t + h_{t-1}] + b_f) \quad (2)$$

B. Input Gate (i_t)

Pada gerbang ini nilai output sebelumnya dengan input saat ini digabung, lalu ada dua fungsi aktivasi yang akan dilewatinya. Jalur satu melewati fungsi aktivasi *sigmoid* untuk nilai input, jalur lainnya melewati fungsi aktivasi *tanh* untuk nilai *candidate memory cell* [2].

$$i_t = \sigma(W_i \times [x_t + h_{t-1}] + b_i) \quad (2)$$

C. Cell State (C_t)

Pada tahap ini ada penggabungan dari dua nilai. Nilai pertama adalah nilai dari *forget gate* akan dikalikan dengan nilai dari *cell state* sebelumnya. Nilai kedua adalah nilai dari input *gate* dikalikan dengan nilai dari *candidate memory cell* [2].

$$C_t = f_t \times C_{t-1} + i_t \times C^t \quad (2)$$

D. Output Gate (o_t)

Gerbang ini menghasilkan nilai *output*, dimana nilai ini berasal dari gabungan nilai sebelumnya dengan nilai saat ini yang telah melalui fungsi aktivasi *sigmoid* [2].

$$o_t = \sigma(W_o \times [x_t + h_{t-1}] + b_o) \quad (2)$$

E. Hidden Layer (h_t)

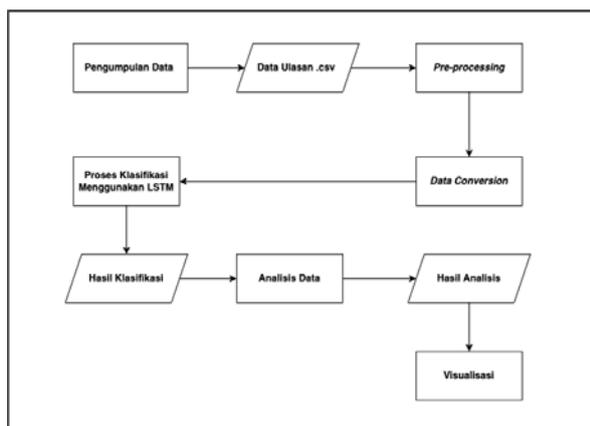
Hidden layer berpengaruh untuk nilai di proses selanjutnya, nilai dari layer ini berasal dari nilai *output* yang dikalikan dengan nilai dari *cell state* atau *memory cell* yang telah diaktivasi dengan fungsi *tangen* [2].

$$h_t = o_t \times \tanh(C_t) \quad (2)$$

3. Hasil dan Pembahasan

Pada penelitian ini akan menjelaskan mengenai alur perancangan dari penelitian yang diambil oleh penulis. Dan tujuan dari penelitian ini yaitu melihat seberapa baik akurasi dari penggunaan algoritma LSTM terhadap ulasan pada aplikasi provider internet di Indonesia. Data yang akan dianalisa adalah data ulasan dari aplikasi provider internet di Indonesia. Disini, penulis mengambil 2 aplikasi provider internet di Indonesia yang bisa dibilang memiliki ulasan terbanyak pada *google playstore* yaitu *myIndiHome* dan *My Firstmedia*. Dan selanjutnya, data tersebut akan dilakukan proses *pre-processing* data sebelum diuji menggunakan algoritma. Lalu, hasil dari pengujian pada penelitian ini akan divisualisasikan ke dalam bentuk tabel dan juga diagram dengan menggunakan data nilai persentase. Data nilai persentase diperoleh dari hasil perhitungan algoritma *Long Short Term Memory*.

Proses Analisis Sentimen Terhadap Ulasan Pada Aplikasi Provider Internet di Indonesia Menggunakan Algoritma *Long Short Term Memory* dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Pengumpulan Data

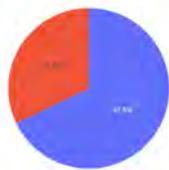
Proses pengambilan data menggunakan teknik *scraping*. Data diambil dari aplikasi *google playstore*. Data yang digunakan pada penelitian ini adalah data ulasan terhadap aplikasi provider internet khususnya di Indonesia. Dan data tersebut diubah ke dalam bentuk *csv* agar dapat di proses lebih lanjut. Terdapat dua data ulasan dari aplikasi provider internet yang diambil oleh penulis, yaitu aplikasi *myIndiHome* dan aplikasi *My Firstmedia*. Penulis mengambil 5000 data ulasan dari masing – masing aplikasi. Terdapat ulasan positif dan negatif di setiap dataset.

Tabel 1. Data Ulasan

Provider	Ulasan Positif	Ulasan Negatif
myIndiHome	83,8%	16,2%
My Firstmedia	50,7%	49,3%

Tabel 1. merupakan hasil persentase dari ulasan positif dan negatif pada dataset *myIndiHome* dan *My Firstmedia*. Dan pada Gambar 2, itu merupakan hasil persentase dari keseluruhan ulasan. Dan didapatkan hasil 67,8% untuk ulasan positif dan 32,2% untuk ulasan negatif.

Percentage of Positive and Negative Review



Gambar 2. Diagram Persentase Ulasan

3.2. Pre-Processing Data

Pre-processing merupakan salah satu langkah penting dalam analisis sentimen. Data ulasan yang telah dikumpulkan masuk ke proses *pre-processing* untuk mendapatkan data yang bersih sehingga proses pembuatan vektor kata dan klasifikasi sentimen lebih

akurat. Tahap *Pre-processing* data yang pertama yaitu *case folding*. *Case folding* merupakan tahap perubahan huruf dari huruf kapital menjadi huruf kecil [9]. Pada tahap ini semua huruf yang ada disetarakan bentuknya yaitu menjadi huruf kecil atau huruf besar semua, tujuannya untuk mempermudah proses setelahnya [2]. Berikut ini merupakan contoh proses *case folding* yang terdapat pada Tabel 2.

Tabel 2. Case folding

Sebelum	Sesudah
Setiap kali saya hendak melakukan pengaduan layanan selalu muncul menu "Mohon Maaf, Kami sedang melakukan peningkatan layanan" Dan ini terjadi sejak 2021. Setiap kali update baru selalu ada kendala.	setiap kali saya hendak melakukan pengaduan layanan selalu muncul menu "mohon maaf, kami sedang melakukan peningkatan layanan" dan ini terjadi sejak 2021. setiap kali update baru selalu ada kendala.

Lalu tahap *pre-processing* data selanjutnya adalah data *cleansing*. Pada tahap ini, data teks yang sudah ada dibersihkan dengan menghilangkan simbol, angka, tanda baca, spasi yang berlebihan maupun karakter yang tidak ada di alfabet [2]. Berikut ini merupakan contoh proses data *cleansing* yang terdapat pada Tabel 3.

Tabel 3. Data Cleansing

Sebelum	Sesudah
setiap kali saya hendak melakukan pengaduan layanan selalu muncul menu "mohon maaf, kami sedang melakukan peningkatan layanan" dan ini terjadi sejak 2021. setiap kali update baru selalu ada kendala.	setiap kali saya hendak melakukan pengaduan layanan selalu muncul menu mohon maaf kami sedang melakukan peningkatan layanan dan ini terjadi sejak setiap kali update baru selalu ada kendala

Tahap *pre-processing* data selanjutnya adalah *tokenization*, yaitu proses tahapan memecah dokumen menjadi kumpulan kata perkata, juga berfungsi untuk menghilangkan tanda baca, spesial karakter dan memisahkannya per spasi [9]. Tujuannya agar dapat dianalisis dengan mudah dari setiap kata. Berikut ini merupakan contoh proses *Tokenization* yang terdapat pada Tabel 4.

Tabel 4 Tokenization

Sebelum	Sesudah
setiap kali saya hendak melakukan pengaduan layanan selalu muncul menu "mohon maaf, kami sedang melakukan peningkatan layanan" dan ini terjadi sejak 2021. setiap kali update baru selalu ada kendala.	['setiap' 'kali' 'saya' 'hendak' 'melakukan' 'pengaduan' 'layanan' 'selalu' 'muncul' 'menu' 'mohon' 'maaf' 'kami' 'sedang' 'melakukan' 'peningkatan' 'layanan' 'dan' 'ini' 'terjadi' 'sejak' 'setiap' 'kali' 'update' 'baru' 'selalu' 'ada' 'kendala']

Tahap *pre-processing* data selanjutnya adalah *stopwords/filtering*, yaitu merupakan proses penghilangan kata tidak penting pada deskripsi melalui pengecekan kata - kata hasil parsing deskripsi apakah

termasuk di dalam daftar kata tidak penting (stoplist) atau tidak [9]. Berikut ini adalah contoh proses *Tokenization* yang ditunjukkan pada Tabel 5.

Tabel 5. *Stopwords*

Sebelum	Sesudah
setiap kali saya hendak melakukan pengaduan layanan selalu muncul menu "mohon maaf, kami sedang melakukan peningkatan layanan" dan ini terjadi sejak 2021. setiap kali update baru selalu ada kendala.	['kali' 'pengaduan' 'layanan' 'muncul' 'menu' 'mohon' 'maaf' 'peningkatan' 'layanan' 'kali' 'update' 'kendala']

Tahap *pre-processing* data yang terakhir yaitu *Stemming*, yaitu dilakukan penghapusan imbuhan baik di awal (*prefixes*), di akhir (*suffixes*), kata sisipan (*infixes*) maupun imbuhan di awal kata dan di akhir (*confixes*) pada setiap kata sehingga kata tersebut menjadi kata dasar [13]. Berikut ini adalah contoh proses *Stemming* yang ditunjukkan pada Tabel 6.

Tabel 6. *Stemming*

Sebelum	Sesudah
setiap kali saya hendak melakukan pengaduan layanan selalu muncul menu "mohon maaf, kami sedang melakukan peningkatan layanan" dan ini terjadi sejak 2021. setiap kali update baru selalu ada kendala.	['kali' 'adu' 'layan' 'muncul' 'menu' 'mohon' 'maaf' 'tingkat' 'layan' 'kali' 'update' 'kendala']

3.3. Data Conversion

Dalam tahap ini, konversi kalimat dilakukan dalam beberapa langkah: membuat kamus kata, mengonversi kata menjadi angka, dan *padding* kalimat. Proses ini dilakukan untuk digunakan sebagai *input* untuk sistem. Prosesnya dimulai dengan membuat kamus kata untuk memberikan id kata yang terdapat di kalimat dalam data ulasan yang telah melalui proses *pre-processing*. Kedua adalah untuk menghapus kata-kata duplikat dalam ulasan. Yang ketiga memberikan id untuk setiap kata secara berurutan berdasarkan kata-kata yang sering muncul. Langkah selanjutnya adalah mengubah kalimat menjadi angka, mengubah kata itu menyusun kalimat ulasan menjadi angka berdasarkan kamus kata-kata yang telah dibuat pada langkah sebelumnya. Langkah terakhir adalah memberikan *padding* pada kalimat tersebut sehingga kalimat tersebut panjangnya sama. *Padding* dilakukan dengan menentukan panjang maksimum kata dalam kalimat atau mencari jumlah kalimat terpanjang dalam data ulasan. Pada tahap ini juga akan dilakukan konversi vector, yang dimana itu

dilakukan dengan menggunakan kamus kata yang telah dibuat dan direpresentasikan dalam model *word2vec* yang telah dibuat. Model *word2vec* telah dibuat menggunakan teks Wikipedia bahasa Indonesia. Jika kata ditemukan dalam model *word2vec* (memiliki representasi vektor untuk sebuah kata), maka vektor *word2vec* digunakan untuk mewakili kata tersebut. Tetapi jika kata itu tidak ditemukan dalam model, itu diganti dengan nilai vektor acak dari semua vektor nilai pada model. Nilai vektor dari kata kamus vektor kemudian diperoleh berdasarkan pada model *word2vec*. Masukan tersebut kemudian menjadi vektor berdimensi $n*m$ dimana n adalah jumlah maksimum ulasan dan m menggunakan dimensi dalam model *word2vec*.

3.4. Algoritma *Long Short Term Memory*

Pada tahap penelitian ini, algoritma LSTM digunakan sebagai algoritma utama untuk menguji keakuratan terhadap data ulasan yang telah diberikan label positif dan negatif. Data yang telah melalui proses *pre-processing* dan telah dikonversi akan dilanjutkan proses klasifikasi dengan menggunakan algoritma *Long Short Term Memory* (LSTM). Penulis menggunakan algoritma LSTM karena memiliki memori yang sangat besar dan cocok digunakan untuk data yang berbentuk *sequence*[14]. Cara kerja dari algoritma LSTM yaitu, pertama membagi dataset terlebih dahulu. Dataset yang telah diberikan label, dibagi menjadi data *train* dan data *test*. Disini penulis membagi data menjadi 70% data *train* dan 30% data *test* dari total keseluruhan 7247 dataset yang akan digunakan untuk tahap pengujian menggunakan LSTM. Pada metode *Long Short Term Memory*, akan dilakukan beberapa pengujian parameter seperti pengujian jumlah *neuron*, *learning rate* ADAM, *word embedding* dan *dropout* pada lapisan LSTM untuk diketahui parameter yang memiliki performa terbaik [6].

```

Memasukkan Model

| | import tensorflow as tf
| |
| | input_dim = 12
| | output_dim = 10
| |
| | # max_length = 100
| | # x = np.random.random((100, input_dim))
| |
| | model = Sequential()
| | model.add(LSTM(input_dim, output_dim, return_sequences=True))
| | model.add(Dense(output_dim))
| | model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
| | # model.fit_generator(generator, data_loader, validation_data_loader, validation_monitor='accuracy')
| | print(model.summary())

```

Gambar 3. Model

Lalu, setelah kita uji data tersebut menggunakan algoritma LSTM. Didapatkan nilai akurasi terbaik dari beberapa percobaan yaitu sebesar 77%. Dan Gambar 4. merupakan hasil akurasi yang didapat.

```

Testing Dataset dan mendapatkan Akurasi

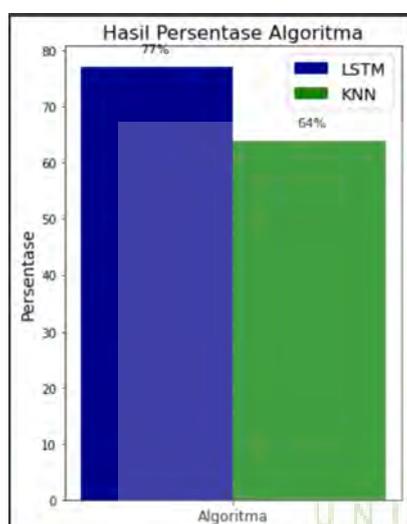
{ } validation_size = 1500
X_validata = X_test[:validation_size]
Y_validata = Y_test[:validation_size]
X_test = X_test[:validation_size]
Y_test = Y_test[:validation_size]
score,acc = model.evaluate(X_test, Y_test, verbose = 2, batch_size = batch_size)
print("score: %.2f" % (score))
print("acc: %.2f" % (acc))

22/22 - 1s - loss: 0.3824 - accuracy: 0.7704 - 1s/wpooh - 67ms/step
score: 0.56
acc: 0.77

```

Gambar 4. Hasil Akurasi

Pada tahap ini, penulis juga menggunakan salah satu algoritma klasifikasi yaitu *K-Nearest Neighbor* dengan tujuan untuk mengetahui apakah nilai akurasi dari algoritma LSTM dapat dikatakan baik atau tidak. Dan dari hasil algoritma tersebut didapatkan nilai akurasi terbaik yaitu 64%.



Gambar 5 Visualisasi hasil persentase akurasi algoritma

Dan Gambar 5. merupakan hasil dari penelitian ini, algoritma *Long Short Term Memory* (LSTM) memiliki nilai akurasi 77%. Nilai tersebut lebih baik dari algoritma *K-Nearest Neighbor* (KNN) yang mendapatkan nilai akurasi sebesar 64%.

4. Kesimpulan

Berdasarkan hasil dari pengujian yang telah dilakukan, dapat ditarik kesimpulan bahwa penerapan metode algoritma *Long Short Term Memory* dapat melakukan klasifikasi sentimen dari dataset ulasan terhadap aplikasi provider di Indonesia secara otomatis dengan tingkat akurasi 77% saat menggunakan data sebanyak 7247 yang dimana dibagi menjadi 70% data *training* dan 30% data *testing*. Dan juga harus diperhatikan yaitu pada tahap *preprocessing*, yang dimana dilakukan pembersihan data sebelum diuji menggunakan algoritma. Karena, jika dataset tidak melewati *preprocessing* dengan benar maka dipastikan keakuratan data itupun bisa sangat rendah. Lalu, nilai akurasi dengan menggunakan metode algoritma LSTM dapat lebih tinggi lagi jika menggunakan data yang

jumlahnya lebih banyak. Karena seperti yang sudah dijelaskan bahwa algoritma LSTM cocok digunakan untuk mengolah data yang besar.

Daftar Rujukan

- [1] C. Raymond, MC.LeodAhmad, E. S. Rini, and I. W. A. Wiratama, "Kota Pagar Alam Berbasis Web," *Ekspelora Inform.*, vol. 07, no. April, pp. 36–49, 2016, [Online]. Available: <https://eksplora.stikom-bali.ac.id/index.php/eksplora/article/view/66/52>.
- [2] J. Nurvania and K. M. Lhaksamana, "Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM)," vol. 8, no. 4, pp. 4124–4135, 2021.
- [3] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 64–69, 2018, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626>.
- [4] Warnia Nengsih, M. Mahrus Zein, and Nazifa Hayati, "Coarse-Grained Sentiment Analysis Berbasis Natural Language Processing – Ulasan Hotel," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 1, pp. 41–48, 2021, doi: 10.22146/jnteti.v10i1.548.
- [5] E. M. Sipayung, H. Maharani, and I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *J. Sist. Inf.*, vol. 8, no. 1, pp. 958–965, 2016, [Online]. Available: <https://ejournal.unsri.ac.id/index.php/jsi/article/view/3250/1907>.
- [6] C. K. N. Papatungan and A. Jacobus, "Sentiment Analysis of Social Media Users Using Long-Short Term Memory Method," *J. Tek. Elektro dan Komput.*, vol. 10, no. 2, pp. 99–106, 2021.
- [7] M. A. Nurrohmat and A. SN, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 3, p. 209, 2019, doi: 10.22146/ijccs.41236.
- [8] A. Y. Chandra, D. Kurniawan, and R. Musa, "Perancangan Chatbot Menggunakan Dialogflow Natural Language Processing (Studi Kasus: Sistem Pemesanan pada Coffee Shop)," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 208, 2020, doi: 10.30865/mib.v4i1.1505.
- [9] M. Yusuf and A. Cherid, "Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar," vol. 9, pp. 8–16, 2020.
- [10] Y. Yunefri and Y. Ersan Fadrial, "Chatbot Pada Smart Cooperative Oriented Problem Menggunakan Natural Language Processing Dan Naive Bayes Classifier Chatbot on Smart Cooperative Oriented Problems Using Natural Language Processing and Naive Bayes Classifier," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2, p. 2021, 2021.
- [11] M. Fuad Muttaqin, G. A. Mutiara, and R. Handayani, "RFID SISTEM PARKIR DENGAN USER PROFILER."
- [12] G. Al Azhar, T. Winarno, S. Izza, and P. N. Malang, "Implementasi g-h Filter Pada Sensor Kompas Sebagai Peningkatan Akurasi Trajectory Tracking Robot Differential Drive," *J. Mech. Electr. Technol.*, vol. 1, no. 1, 2022.
- [13] "Journal of Computer Science IJCSIS June 2018 Full Volume."
- [14] A. S. Talita and A. Wiguna, "Implementasi Algoritma Long Short-Term Memory (LSTM) Untuk Mendeteksi Ujaran Kebencian (Hate Speech) Pada Kasus Pilpres 2019," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 37–44, Nov. 2019, doi: 10.30812/matrik.v19i1.495.
- [15] Alamanda, R., Suhery, C., Brianorman, Y., & Komputer, J. S. (2016). Jurnal Coding , Sistem Komputer Untan APLIKASI PENDETEKSI PLAGIAT TERHADAP KARYA TULIS BERBASIS WEB MENGGUNAKAN NATURAL LANGUAGE PROCESSING DAN Jurnal Coding , Sistem Komputer Untan ISSN: 2338-493x. Jurnal Coding, Sistem Komputer Untan, 04(1).
- [16] S. Hochreiter And J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, Vol. 9, No. 8, Pp. 1735–1780, 1997, Doi: 10.1162/Neco.1997.9.8.1735.
- [17] Li, D., & Qian, J. (2016, October). Text sentiment analysis based on long short-term memory. In 2016 First IEEE International

- Conference on Computer Communication and the Internet (ICCCI) (pp. 471-475). IEEE.
- [18] Miedema, F. (2018). Sentiment analysis with long short-term memory networks. Research Paper Business Analytics Vrije Universiteit Amsterdam.
- [19] Bisong, E. (2019). Google Colaboratory BT - Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners (E. Bisong (ed.); pp. 59–64). Apress. https://doi.org/10.1007/978-1-4842-4470-8_7



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “ANALISIS SENTIMEN TERHADAP ULASAN PADA APLIKASI PROVIDER INTERNET DI INDONESIA MENGGUNAKAN ALGORITMA LONG SHORT TERM MEMORY”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, Analisa perancangan, dataset yang digunakan, tahapan eksperimen , dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan Analisa perancangan dan alur sistem rekomendasi. Bagian III menjelaskan mengenai source code yang digunakan pada penelitian ini. Bagian VI menjelaskan mengenai dataset yang digunakan. Bagian V memuat tahapan eksperimen yang disajikan dalam gambar dan table dengan penjelasan dari setiap tahapan. Bagian VI merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.