

THE APPLICATION OF GATED RECURRENT UNIT IN SOCIAL MEDIA SENTIMENT ANALYSIS

Penerapan Metode Gated Recurrent Unit Pada Analisis Sentimen Media Sosial

Clinton Imanuel Kawatu

Dept. of Electrical Engineering, Sam Ratulangi University Manado, Kampus Bahu St., 95115, Indonesia
e-mails : clintonmanuel73@gmail.com

Abstract — This research is motivated by the increasingly varied opinions from the public about their experiences using a shipping company through social media, which makes it time – consuming for shipping company to manually collect and analyze customer feedback. The purpose of this research is to develop a wae – based text sentiment analysis application to help shipping companies extract and understand the feelings and opinions from the public from text posted on social media, specifically Twitter, with each opinion classified into three categories: positive, negative, and neutral. In this study, the methods used is the Gated Recurrent Unit (GRU) algorithm, which has the ability to capture long – term dependencies in sequential data, such as social media texts, that can process and understand the context of long and complex sentences, thereby providing accurate results. By conducting this analysis, companies can identify areas that need improvement, understand customer satisfaction trends, and design better strategies to enhance service quality. Additionally, the results of this research are expected to make a significant contribution to the field of customer service management and data – driven business development.

Key words : Social Media, Sentiment Analysis, Twitter, Gated Recurrent Unit

Abstrak — Penelitian ini dilatarbelakangi oleh opini dari masyarakat yang semakin bervariasi tentang pengalaman mereka dalam menggunakan suatu perusahaan ekspedisi barang melalui media sosial, yang membuat para perusahaan ekspedisi memakan waktu yang lama dalam mengumpulkan dan menganalisis umpan balik pelanggan secara manual. Tujuan dari penelitian ini adalah membuat aplikasi analisis sentimen teks berbasis web untuk membantu perusahaan ekspedisi dalam mengekstraksi dan memahami perasaan serta opini dari masyarakat dari teks yang diposting dari media sosial *Twitter*, yang tentunya setiap opini akan diklasifikasi menjadi tiga kelas yaitu positif, negatif, dan netral. Dalam penelitian ini, metode yang digunakan adalah algoritma *Gated Recurrent Unit (GRU)* yang memiliki kemampuan untuk menangkap dependensi jangka panjang dalam data sekuensial, seperti teks media sosial serta mampu memproses dan memahami konteks dari kalimat – kalimat yang panjang dan kompleks, sehingga memberikan hasil yang akurat. Dengan melakukan analisis ini, perusahaan dapat mengidentifikasi area yang memerlukan perbaikan, memahami tren kepuasan pelanggan, dan merancang strategi yang lebih baik untuk meningkatkan kualitas layanan. Selain itu, hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang manajemen layanan pelanggan dan pengembangan bisnis berbasis data.

Kata Kunci : Media Sosial, analisis sentimen, Twitter, Gated Recurrent Unit

I. PENDAHULUAN

kepuasan pelanggan adalah salah satu faktor utama. Perusahaan harus fokus pada aspek – aspek yang dianggap penting oleh pelanggan untuk mencapai tingkat kepuasan yang tinggi. Oleh karena itu perusahaan perlu melakukan evaluasi untuk mengidentifikasi faktor – faktor yang mempengaruhi kepuasan pelanggan dengan cara – cara memenuhi harapan mereka. Ketika pelanggan merasa puas dengan layanan yang diberikan, loyalitas pelanggan akan terbentuk. Namun, menjaga pelanggan tetap setia merupakan salah satu tantangan besar yang dihadapi perusahaan.

Twitter adalah platform media sosial yang memungkinkan penggunaannya untuk menulis pesan pendek, mengunggah foto atau video, memperbarui, dan membaca pesan-pesan yang dibuat oleh pengguna yang lainnya, biasanya lebih dikenal dengan istilah *tweet*. Dari platform media sosial ini masyarakat juga dapat membagikan opini mereka tentang pengalaman menggunakan setiap perusahaan ekspedisi barang. Sekarang ini opini masyarakat semakin bervariasi, hal ini dikarenakan sudah banyak masyarakat yang menggunakan media sosial. Analisis sentimen bisa menjadi solusi yang efektif untuk memantau dan memahami tiap pendapat atau opini yang ada dalam bentuk teks bebas di media sosial.

A. Penelitian Terkait

Dalam penelitian ini, terdapat acuan dari studi – studi sebelumnya yang digunakan sebagai referensi. Berikut adalah beberapa penelitian terdahulu yang dijadikan bahan pertimbangan dalam penelitian ini.

Sentiment Analysis on Twitter Using Neural Network: Indonesian Presidential Election 2019 Dataset [1]. Penelitian ini dilakukan oleh Ahmad Fathan Hidayatullah, Siwi Cahyaningtyas, dan Anisa Miladya Hakim. Penelitian ini dilakukan perbandingan terhadap beberapa metode algoritma *machine learning* dan *deep learning* yang berbeda, menggunakan dataset dari twitter menjelang pemilihan presiden 2019. Hasil dari implementasi dari beberapa algoritma tersebut, *bidirectional LSTM* memiliki hasil akurasi yang lebih baik. Nilai yang didapat sebesar 84.60.

Performance Analysis of Deep Approaches on Airbnb Sentiment Reviews [2]. Penelitian ini dilakukan oleh Muhammad Raheel Raza, Walayat Hussain, dan Asaf Varol. Penelitian ini dilakukan untuk menganalisis performa algoritma *deep learning* seperti *Recurrent Neural Network* (RNN), *Long Short – Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) dalam mengklasifikasi ulasan yang diberikan pada *website* penyewaan akomodasi *Airbnb* yang dapat diunduh dari *Kaggle*.

Analisis Sentimen Ulasan Pengguna Aplikasi E-Samsat Provinsi Jawa Barat Menggunakan Metode BiGRU [3]. Penelitian ini dilakukan oleh Rahma Kania Dewi, Bertho Tantular, Jadi Suprijadi, dan Anindya Apriliyanti Ptavitasiari. Penelitian ini menganalisis sentimen terhadap aplikasi Samsat Mobile Jawa Barat (SAMBARA) di Google Play Store dengan mengklasifikasikan ulasan pengguna ke dalam tiga kategori: positif, negatif, dan netral. Metode yang digunakan adalah *BiDirectional Gated Recurrent Unit* (BiGRU). BiGRU dapat memprediksi sentimen ulasan pengguna dengan akurasi hingga 87,37%. Hal ini dirasa cukup dan dapat digunakan untuk mengembangkan aplikasi layanan di Jawa Barat.

Twitter Sentiment Analysis Using Aspect – Based Bidirectional Gated Recurrent Unit with Self – Attention Mechanism [4]. Penelitian ini dilakukan oleh Mohan Kumar Antharasanahalli Venkataramaiah, dan Nandakumar Ambuga Narayana Achar. Dalam penelitian ini digunakan metode *aspect – based BiGRU* pengguna dari twitter yang dikumpulkan untuk mengklasifikasikan *tweet* menjadi positif, negatif, dan netral.

Analyzing Election Sentiments in Tweets with Gated Recurrent Units (GRU) [5]. Penelitian ini dilakukan oleh Agu Edward O., Bako Jeremy Zevini, dan Hambali Moshood Abiola. Penelitian ini berfokus pada mode *Gated Recurrent Unit* (GRU) untuk menganalisis sentimen pada dataset *tweet* pemilu Donald Trump pada tahun 2020. Hasil akurasi yang didapatkan dari penelitian ini adalah sebesar 93% memverifikasi kemampuan model untuk mengkategorikan sikap dengan tepat. Selain itu *recall* dan presisi masing – masing mendapatkan penilaian yang sangat baik sebesar 94% dan 96% menunjukkan keahlian model dalam mengklasifikasikan sentimen menjadi positif dan negatif.

B. Analisis Sentimen

Analisis Sentimen atau penambangan opini adalah studi komputasi atas opini, perasaan, emosi, penilaian, dan sikap masyarakat terhadap entitas seperti produk, layanan, organisasi, penyebab individu, masalah, peristiwa, topik, dan atribut terkait [6]. Analisis sentimen menggunakan teknik pemrosesan bahasa alami (NLP), analisis teks, dan linguistik komputasi untuk mengidentifikasi dan mengekstrak informasi subjektif dari sumber teks [7]. Analisis sentimen bertujuan untuk mengevaluasi emosi, sikap, pendapat, ulasan yang disampaikan oleh seorang pembicara atau penulis tentang suatu produk atau tokoh masyarakat. Oleh karena itu, beberapa penelitian khususnya review produk perlu didahului dengan menentukan elemen produk apa yang dibicarakan sebelum memulai proses pengumpulan opini (*Opinion Mining*) [8]. Ada beberapa tahapan dalam analisis sentimen yaitu.

1. Pengumpulan data

2. *Preprocessing Data*
3. Ekstraksi Fitur
4. Pelatihan Model
5. Evaluasi Model
6. Implementasi

C. Data Mining

Data mining adalah proses yang melibatkan penemuan pola, hubungan, dan pengetahuan yang berguna dari kumpulan data yang besar. Proses ini menggabungkan berbagai teknik dari bidang statistik, pembelajaran mesin, dan basis data untuk mengidentifikasi informasi yang tersembunyi dalam data [9].

Banyaknya data yang terdapat dalam *database* saat ini melebihi kemampuan manusia untuk menganalisis dan mengekstrak informasi yang paling berguna tanpa bantuan teknik analisis otomatis. Penemuan pengetahuan adalah proses yang tidak begitu sederhana dalam mengekstraksi informasi yang tersembunyi, tidak diketahui, dan berpotensi berguna dari database besar. Penambangan data yang digunakan dalam penemuan pengetahuan telah membantu menemukan pola berdasarkan kebutuhan pengguna [10].

D. Machine Learning

Machine Learning adalah cabang dari kecerdasan buatan (Artificial Intelligence) yang berfokus pada pengembangan sistem komputer yang dapat belajar dan meningkatkan kinerjanya dari pengalaman tanpa perlu diprogram secara eksplisit. Secara sederhana, ML memungkinkan komputer untuk belajar dari data yang diberikan dan membuat prediksi atau keputusan berdasarkan pola yang teridentifikasi dalam data tersebut.

Machine Learning adalah studi ilmiah tentang algoritma dan model statistik yang digunakan oleh sistem komputer untuk melakukan tugas tertentu tanpa diprogram secara eksplisit. Algoritma pembelajaran diterapkan dalam banyak aplikasi yang kita gunakan setiap hari. Setiap kali mesin pencari web seperti Google digunakan untuk mencari di internet, salah satu alasan mengapa itu berfungsi dengan baik adalah karena adanya algoritma pembelajaran yang telah belajar bagaimana merangking halaman web. Algoritma ini digunakan untuk berbagai tujuan seperti penambangan data, pemrosesan gambar, analisis prediktif, dan lain-lain. Keuntungan utama menggunakan pembelajaran mesin adalah, setelah sebuah algoritma belajar apa yang harus dilakukan dengan data, ia dapat bekerja secara otomatis [11].

E. Deep Learning

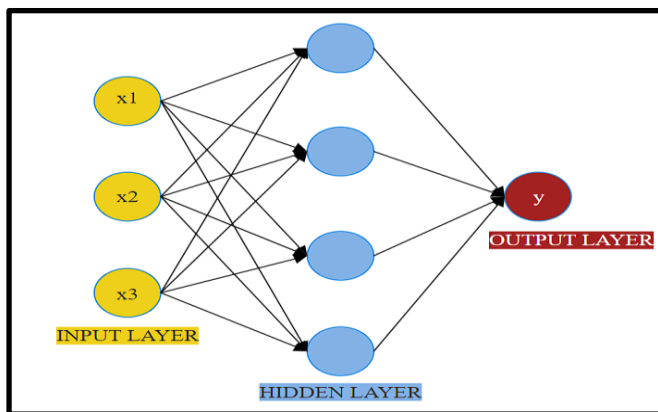
Deep Learning atau pembelajaran mendalam merupakan subbidang dari *Machine Learning* yang berfokus pada pembelajaran representasi data kompleks menggunakan jaringan syaraf tiruan (*deep neural network*) yang terdiri dari banyak lapisan. Pembelajaran mendalam dapat mengotomatiskan ekstraksi fitur dari data mentah. Artinya, model dapat belajar langsung dari data tanpa memerlukan prapemrosesan manual yang ekstensif.

Deep Learning memungkinkan model komputasi yang terdiri dari beberapa lapisan pemrosesan untuk mempelajari

representasi data pada tingkat abstraksi yang berbeda. Metode-metode ini telah meningkatkan secara signifikan teknik-teknik terkini dalam pengenalan suara, pengenalan objek visual, pengenalan objek, dan banyak bidang lainnya seperti penemuan obat dan genomik. Pembelajaran mendalam menggunakan algoritma propagasi mundur untuk mendeteksi struktur kompleks dalam kumpulan data besar, memungkinkan mesin menentukan parameter internal mana yang digunakan untuk menghitung representasi setiap lapisan dari representasi lapisan sebelumnya. Jaringan saraf konvolusional yang dalam telah membawa terobosan dalam pemrosesan gambar, video, audio, dan audio, dan jaringan saraf berulang telah memungkinkan pemahaman yang lebih mendalam tentang data berkelanjutan seperti teks dan audio [12].

F. Artificial Neural Network (ANN)

Artificial Neural Network adalah sebuah pendekatan dalam pengolahan informasi yang terinspirasi oleh fungsi sistem saraf biologis, menerapkan prinsip-prinsip kerja otak manusia untuk memproses data. Melalui kemampuannya dalam memodelkan relasi yang kompleks antara input dan output, ANN memungkinkan untuk mengidentifikasi pola-pola yang ada dalam dataset. Secara umum, ANN terdiri dari tiga lapisan utama: lapisan input, lapisan tersembunyi, dan lapisan output [13].



Gambar 1. Struktur ANN

Gambar 1 merupakan struktur dari ANN. Serupa dengan sistem saraf biologis, *Artificial Neural Network* juga memiliki *neuron – neuron* yang bersifat artifisial dan mereka juga menerima masukan dari elemen lain atau neuron artifisial lainnya. Setelah masukan diberi bobot dan dijumlahkan, hasilnya kemudian diubah oleh fungsi transfer menjadi keluaran.

G. Recurrent Neural Network (RNN)

Recurrent Neural Network merupakan salah satu model arsitektur jaringan syaraf tiruan atau *neural network* yang dirancang khusus untuk memproses data yang bersambung (*sequential data*). Operasi pada RNN dilakukan untuk setiap elemen dari suatu urutan, dengan hasilnya tergantung pada input saat ini dan operasi sebelumnya. RNN berfokus pada sifat data dimana hasil waktu sebelumnya atau saat ini mempengaruhi hasil pada waktu berikutnya. Cara yang dilakukan RNN untuk dapat menyimpan informasi dari masa

lalu adalah dengan melakukan *looping* di dalam arsitekturnya, yang secara otomatis membuat informasi dari masa lalu tetap tersimpan.

Status tersembunyi RNN yang berdimensi tinggi dan evolusi non-linear memberikan daya ekspresif yang besar, memungkinkan status tersembunyi RNN untuk mengintegrasikan informasi selama banyak langkah waktu dan menggunakannya untuk membuat prediksi yang akurat. Meskipun non-linearitas yang digunakan oleh setiap unit cukup sederhana, mengiterasinya seiring waktu menghasilkan dinamika yang sangat kaya. Standar RNN diformalkan sebagai berikut: Diberikan urutan vektor input (x_1, \dots, x_T), RNN menghitung urutan status tersembunyi (h_1, \dots, h_T) dan urutan keluaran (o_1, \dots, o_T) dengan mengiterasi persamaan berikut untuk $t = 1$ hingga T :

$$h_t = \tanh(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

$$o_t = W_{oh}h_t + b_o \quad (2)$$

Berikut penjelasan dalam persamaan (1) dan (2) yang telah ditampilkan di atas. W_{hx} adalah matriks bobot input – ke – tersembunyi, W_{hh} adalah matriks bobot tersembunyi – ke – tersembunyi (atau berulang), W_{oh} adalah matriks bobot tersembunyi – ke – output, dan vektor – vektor b_h dan b_o adalah bias. Ekspresi $W_{hh}h_{t-1}$ yang tidak terdefinisi pada waktu $t = 1$ digantikan dengan vektor bias awal khusus, h_{init} , dan nonlinearitas tanh diterapkan secara koordinat – wise [14].

H. Gated Recurrent Unit

Gated Recurrent Unit adalah salah satu metode *deep learning* yang memiliki Kemampuan untuk mempelajari hubungan informasi dari waktu sebelumnya ke waktu sekarang. Tujuan utama dari pembuatan GRU adalah untuk membuat setiap *recurrent unit* untuk dapat menangkap *dependencies* dalam skala waktu yang berbeda – beda secara adaptif. GRU digunakan untuk memproses data urutan seperti teks, waktu, atau data sekuensial lainnya. GRU dirancang untuk mengatasi masalah pelepasan informasi jangka panjang yang sering terjadi pada mode ANN lainnya seperti simple RNN.

GRU memiliki struktur relatif sederhana dibandingkan dengan model RNN lainnya seperti LSTM (*Long Short-Term Memory*). Di dalam GRU, komponen pengatur alur informasi tersebut disebut sebagai *gate* dan GRU mempunyai dua gate, yaitu *reset gate* dan *update gate*. *Reset gate* pada GRU akan menentukan bagaimana untuk menggabungkan input baru dengan informasi masa lalu dengan kata lain *reset gate* mengontrol beberapa banyak informasi dari masa lalu yang akan digunakan dan yang akan dibuang, sedangkan *update gate* akan menentukan berapa banyak informasi masa lalu yang akan disimpan untuk diteruskan ke status tersembunyi saat ini. Setiap *hidden state* dihitung menggunakan persamaan berikut ini :

$$Z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (3)$$

$$R_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (4)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_{t-1}, x_t] + b_c) \quad (5)$$

$$h_t = (1 - Z_t) \times h_{t-1} + Z_t \times \tilde{h}_t \quad (6)$$

$$h_t = (1 - Z_t) \times h_t + Z_t \times \tilde{h}_t \quad (7)$$

Keterangan:

Z_t	= <i>update reset</i>
rt	= <i>reset gate</i>
\tilde{h}_t	= <i>candidate hidden state</i>
h_t	= <i>final hidden state</i>
h_{t-1}	= <i>state</i>
σ	= lapisan gerbang sigmoid
Tanh	= lapisan jaringan fungsi aktivasi tanh
W_h, W_z, W_r	= matriks bobot

I. Media Sosial

Media sosial adalah platform online yang memungkinkan penggunaannya untuk membuat, berbagi, dan berinteraksi dengan konten serta dengan pengguna lain. Media sosial telah menjadi bagian integral dari kehidupan sehari-hari bagi banyak orang di seluruh dunia, memungkinkan mereka untuk berkomunikasi, berbagi informasi, dan tetap terhubung dengan jaringan sosial mereka.

Media sosial tidak hanya digunakan untuk berkomunikasi saja, saat ini ada beberapa manfaat yang bisa didapatkan Ketika menggunakan media sosial yaitu memungkinkan untuk membangun jaringan, mencari pekerjaan, berbagi pengetahuan industri, edukasi, serta dapat digunakan untuk mempromosikan suatu produk dan layanan. Media sosial telah merevolusi cara kita berkomunikasi dan berbagi informasi. Meskipun menawarkan manfaat, penting untuk menggunakan media sosial dengan bijak dan menyadari tantangan serta risikonya. Dengan pendekatan yang tepat, media sosial dapat menjadi alat yang kuat untuk konektivitas, pembelajaran, dan hiburan.

J. Twitter

Twitter adalah platform media sosial yang memungkinkan penggunaannya untuk mengirim dan membaca pesan berbasis teks yang disebut "tweet". Tweet ini dibatasi hingga 280 karakter (sebelumnya 140 karakter), dan pengguna dapat menyertakan tautan, gambar, video, dan GIF dalam tweet mereka. Twitter telah menjadi salah satu platform media sosial paling populer di dunia sejak tahun 2006. Berikut ini adalah fitur yang ada pada twitter.

II. METODE

A. Kerangka Pikir

Berdasarkan dengan surat keputusan penelitian yang diberlakukan, di Program Studi Teknik Informatika Universitas Sam Ratulangi, Manado, dimulai pada bulan maret 2023 hingga selesai. Berikut ini adalah penjelasan mengenai kerangka pikir dalam penerapan metode *Gated Recurrent Unit* pada analisis sentimen media sosial. Kerangka pikir adalah model konseptual yang digunakan dalam penelitian terhadap teori - teori yang membentuk dasar dari langkah - langkah penelitian yang akan diambil.

1) Identifikasi Masalah

Identifikasi masalah adalah tahap dimana penulis akan mencari suatu masalah dalam penelitian kemudian akan dicari solusi. Permasalahan yang ditemui oleh penulis adalah apakah

metode Gated Recurrent unit dapat diimplementasikan dalam pembuatan aplikasi analisis sentimen media sosial. Data yang digunakan berupa cuitan berbahasa Indonesia dan akan diklasifikasi menjadi tiga kategori yaitu positif, negatif, dan netral.

2) Studi Literatur

Studi literatur adalah tahapan penulis dalam mengumpulkan, mengevaluasi, dan mensintesis literatur yang relevan dengan topik penelitian atau kajian tertentu. Tujuannya adalah untuk memahami penelitian terkait sebelumnya, mengidentifikasi celah pengetahuan, dan membangun landasan teoritis atau dasar pengetahuan dalam penelitian yang akan dibuat. tahapan ini juga akan menentukan keberhasilan penelitian yang akan dibuat, karena jika penulis tidak memahami teori terhadap penelitian yang akan dibuat akan menghadapi kegagalan.

3) Pengambilan Data

Pada tahap ini penulis menyiapkan data cuitan berbahasa Indonesia yang sudah dilakukan penarikan dan juga sudah pernah dipakai dalam penelitian sebelumnya pada tahun 2021 dengan menggunakan metode klasifikasi yang berbeda, selanjutnya akan dilakukan label manual. Setelah data didapatkan, selanjutnya akan dilakukan proses klasifikasi yang terdiri dari tiga kategori yaitu positif, negatif, dan netral.

4) Preproses Data

Pada tahap preproses data terdapat tiga langkah yaitu *cleaning*, *case folding*, *stopword removal*, dan *stemming*. Untuk seluruh tahapan ini akan menggunakan *library* sastrawi Tujuannya yaitu untuk menghindari gangguan dalam data pada tahap berikutnya, yaitu ekstraksi fitur dan pembentukan model klasifikasi.

1. Cleaning

Pada proses cleaning, teks pada cuitan yang dianggap tidak perlu akan dilakukan penghapusan. Beberapa bagian yang akan dihapus yaitu:

- Tanda baca
- Karakter kosong (Whitespace)
- Karakter Non ASCII
- Link
- Single Character
- Karakter Angka

2. Case Folding

Pada tahap ini akan dilakukan proses menghilangkan perbedaan bentuk huruf, karena dalam setiap cuitan huruf yang digunakan tidak seragam. Langkah ini akan membuat semua huruf dalam data cuitan menjadi kecil.

3. Stopword Removal

Tahap ini dilakukan untuk menghapus kata - kata yang tidak berhubungan atau tidak relevan dengan topik yang sedang diteliti.

5) Pembentukan Model

Dalam Penelitian ini, data yang telah melalui tahap pemrosesan akan dipisahkan menjadi data latih dan data uji menggunakan teknik train test split. Implementasi tahap pembagian ini akan menggunakan library scikit-learn yang tersedia dalam pemrograman python.

Setelah dilakukan proses split data, dilakukan ekstraksi fitur menggunakan metode Tokenizer. Tujuan dari proses ini adalah untuk mengkonversi kata - kata menjadi representasi numerik yang akan digunakan sebagai data latih dan data uji. Implementasi ini akan memanfaatkan modul tensorflow keras dalam pemrograman python. konversi kata dilakukan berdasarkan jumlah kata dalam dataset.

Setelah proses ekstraksi fitur selesai, langkah selanjutnya adalah melakukan proses pelatihan dan pengujian menggunakan algoritma Gated Recurrent Unit (GRU). Setelah mendapatkan model dan akurasi terbaik, model tersebut akan disimpan dalam format .keras.

6) Implementasi

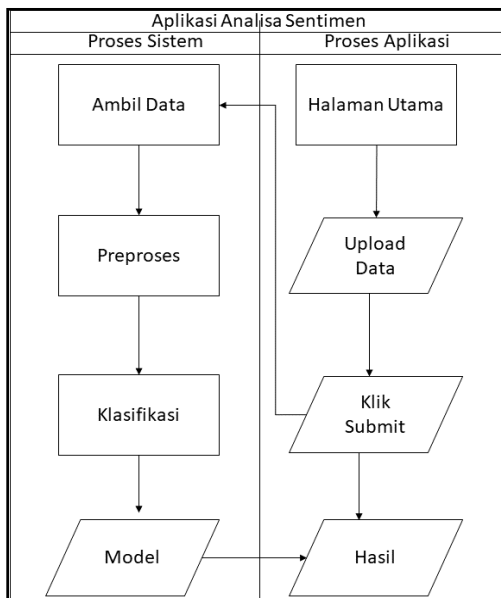
Tahapan ini bertujuan untuk menerapkan metode Gated Recurrent Unit (GRU) dalam analisis sentimen media sosial melalui program terstruktur yang mencakup fungsi - fungsi yang diperlukan dalam aplikasi yang dibuat.

7) Evaluasi

Akan dilakukan pengujian pada aplikasi dengan memasukkan kata kunci ke dalam kotak dialog yang sudah disediakan, aplikasi yang telah dibuat akan memastikan bahwa hasil penarikan data cuitan baru akan melalui proses pre proses, dan klasifikasi secara otomatis.

B. Prinsip Kerja Sistem

Pada bagian ini akan ditampilkan diagram mengenai bagaimana alur aplikasi berjalan dari awal hingga akhirnya mendapatkan hasil yang diinginkan.



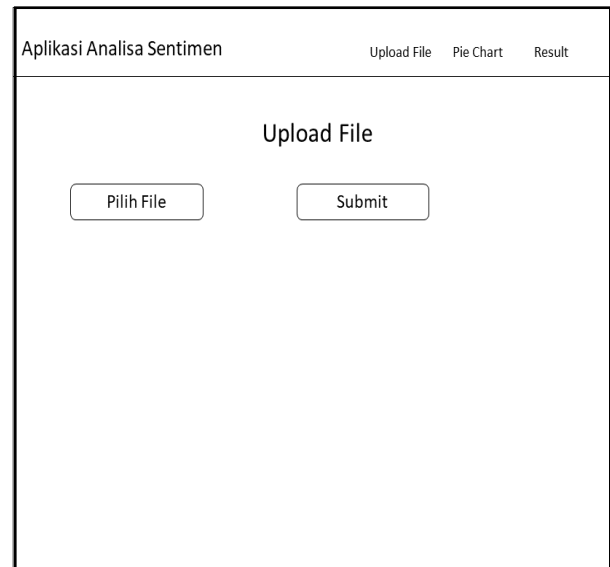
Gambar 2. Diagram Prinsip Kerja Sistem

Pada gambar 2 diperlihatkan alur proses dari aplikasi analisis sentimen dengan metode Gated Recurrent Unit (GRU). Pada

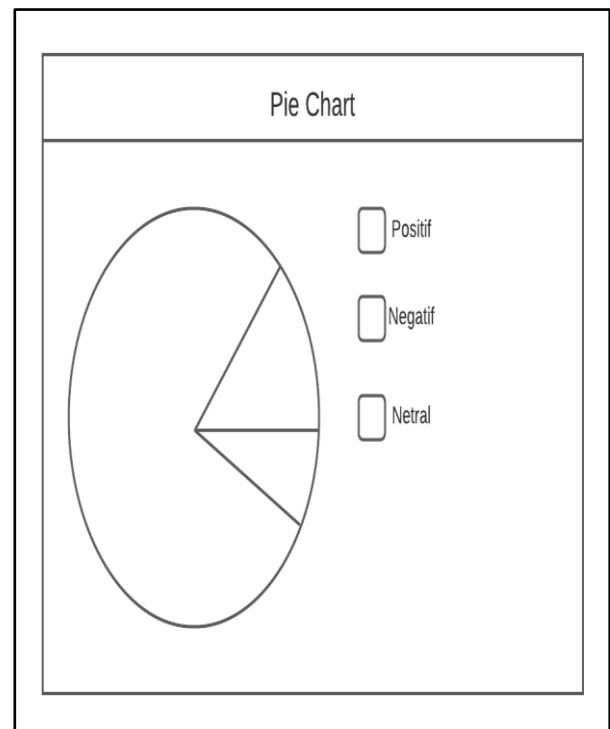
awalnya, user akan mengupload data yang diinginkan kemudian menekan tombol submit. Sistem akan segera menarik data dari Twitter untuk melewati tahap preproses. Setelah itu, model yang telah dibuat sebelumnya akan diaplikasikan pada data yang baru saja ditarik dan dipreproses untuk tahap klasifikasi.

C. Perancangan Antarmuka

Antarmuka adalah mekanisme komunikasi antara User dengan sistem dimana User dapat memberi dan menerima informasi dari sistem. Perancangan tampilan ini adalah konsep antarmuka aplikasi yang pada nantinya akan diterapkan suatu proses sistem di dalamnya. Perancangan antarmuka aplikasi analisis sentimen sosial media ini berbasis website.



Gambar 3. Tampilan awal



Gambar 4. Tampilan Pie Chart

Hasil		
No.	Tweet	Sentimen
1.		100.

Gambar 5. Tampilan Halaman Hasil

Gambar 3 adalah antarmuka awal aplikasi analisis sentimen sosial media menampilkan 1 kotak dialog untuk dan tombol submit yang akan terhubung dengan sistem penarikan data.

Gambar 4 adalah antarmuka aplikasi analisis sentimen menampilkan *pie chart* yang menggambarkan perbandingan data cuitan yang telah diklasifikasikan menjadi positif, negatif, dan netral.

Gambar 5 adalah antarmuka aplikasi menampilkan tabel yang berisi cuitan sesuai dengan data yang di upload pada tampilan awal, lengkap dengan sentimen yang diklasifikasi.

III. HASIL DAN PEMBAHASAN

A. Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data cuitan berbahasa Indonesia tentang pengalaman masyarakat dalam menggunakan suatu perusahaan jasa ekspedisi sicepat, yang sudah dilakukan penarikan dan juga sudah pernah dipakai dalam penelitian sebelumnya pada pada tahun 2021 dengan menggunakan metode klasifikasi berbeda, yang terdiri dari 2677 cuitan. selanjutnya dilakukan label atau klasifikasi manual pada data dengan hasil yang didapatkan dari klasifikasi tersebut terdapat 841 cuitan positif, 842 negatif, dan 974 netral. Setelah itu sata akan dibagi menjadi dua sampel data latih dan data uji dengan ukuran sebesar 33% untuk sampel data latih dari keseluruhan data, sementara 67% sisanya akan digunakan untuk pelatihan pada data dengan menggunakan metode *train test split*.

TABEL I
PEMBAGIAN DATASET

Data Cuitan	Jumlah cuitan	Presentase
Sampel data latih	1773	67%
Sampel data uji	884	33%
Total	2677	100%

	A	B
330	Sebelumnya, awal 2122, SiCepat Ekspres telah berkolaborasi dengan BPBD Ka	1
331	Antisipasi Bencana Banjir, SiCepat Ekspres Salurkan Bantuan Perahu Karet ke	1
332	GRATIS ONGKIR! Beli buku favorit kamu di Tokopedia Haru Semesta dan dapat	1
333	/amn jd custku resel mots DVD total 7kgan, ada yg tau ga pengiriman kargo lain	1
334	Berikut cara untuk cek resi SiCepat untuk dapat melacak pengiriman paket Anda	1
335	help minta nomor kurir jnt sama sicepat dong yang di bekasi sekitar summarecor	1
336	Pppp @sicepat_ekspres	1
337	hallo min @sicepat_ekspres selamat siang, tolong info min paket status paketny	1
338	Buat kamu para Seller di Tokopedia, Mimin punya info yang bisa bikin kamu SEM	1
339	Lowongan Kerja Sicepat Ekspres Sanggau Oktober 2122	1
340	jnt & sicepat halu lama bgt ya :)	0
341	Gila kali ni sicepat pengiriman ampe malem	2
342	HALLO... @sicepat_ekspres ...cek DM yes	1
343	Sumpah ngeri banget yaa :(aku kirim SYB kemarin masa paket yang awalnya ra	0
344	OPEN JASA ADMIN PACKING GO - shopee udh freeong xtra -dom solo	1
345	Motor Besutan SiCepat dan Mitra Ramaikan Indonesia E-Vehicle Expo 2122	1
346	halo @sicepat_ekspres sudah jam nih, tapi paket saya tidak ada infonya sudah	0
347	moots wta,kalo misalnya nih kalian ke jnt/sicepat tuh kan biasanya ada cewe/cov	1

Gambar 6. Potongan data cuitan

Gambar 6 adalah potongan data yang akan digunakan untuk proses klasifikasi yang sudah diberi label pada tiap cuitan.

B. Konfigurasi Hyper – parameter

Pada tahap ini, teks yang sudah melalui tahap *preprocessing* akan dilakukan proses pembangunan model dari *Gated Recurrent Unit* yang memerlukan hyper – parameter yang akan ditampilkan pada tabel II di bawah ini.

TABEL II

KONFIGURASI HYPER – PARAMETER

Hyper – parameter	Jenis dan jumlah hyper – parameter
Model	Sequential
Embedding Size	100
Optimizer	Adam
Aktivasi	Softmax
Layer	1-GRU
Learning rate	0.0001

Tabel 2 adalah jenis hyper – parameter yang digunakan dalam pemrosesan teks dengan membangun model GRU. Pertama dibuat fungsi sequential, ini adalah model linier dari library keras dimana lapisan – lapisan dibangun satu per satu dalam urutan. Embedding membantu model memahami konteks kata berdasarkan distribusi kata dalam pelatihan. Berikut ini ada beberapa parameter yang akan dimasukkan nilai tetap dalam lapisan embedding, pertama jumlah maksimal kata yang sudah ditentukan sebelumnya dengan tokenizer dengan jumlah 10.000 kata akan dimasukkan dalam parameter *input_dim*. Ukuran dari vektor embedding yang akan dihasilkan untuk setiap kata dalam kasus ini 100 akan dimasukkan dalam parameter *output_dim*.

Pada Pada parameter *input length*, semua urutan akan dipotong atau dipadatkan sesuai dengan panjang maksimum

token yang telah ditentukan pada `max_tokens`. Kemudian masuk pada lapisan GRU dengan jumlah unit yang dipakai yaitu sebesar 64 dengan tingkat dropout 0.9 untuk regulasi, mengurangi overfitting dengan mengabaikan 80% unit secara acak selama pelatihan. GRU memproses urutan data teks dan menangkap dependensi temporal atau sekuensial dalam data.

Selanjutnya masuk pada lapisan dense yang merupakan lapisan fully connected atau lapisan neuron biasa yang digunakan untuk menghasilkan output prediksi GRU yang kemudian diubah menjadi probabilitas untuk setiap kelas dengan menggunakan fungsi aktivasi softmax sesuai dengan jumlah kelas yang diprediksi dalam penelitian ini yaitu klasifikasi 3 kelas.

Terakhir model kompilasi untuk menentukan fungsi loss, optimizer, dan metrics yang akan digunakan untuk melatih model. Fungsi loss yang digunakan adalah `categorical_crossentropy` untuk klasifikasi multi kelas, fungsi optimizer yang digunakan adalah optimizer Adam yang digunakan untuk memperbaiki bobot model selama pelatihan dengan learning rate sebesar 0.0001, dan fungsi metrics accuracy digunakan untuk mengevaluasi performa model selama pelatihan dan pengujian data.

C. Hasil Akurasi dan Nilai Metrik

Pada bagian ini akan dihitung nilai metrik seperti *precision*, *recall*, dan *F1 - score* pada proses klasifikasi data menggunakan algoritma *Gated Recurrent Unit* berikut ini hasil metrik yang didapatkan akan ditampilkan pada tabel III berikut ini.

TABEL III
NILAI METRIK GRU

Label	Precision	Recall	F1-score	Support
Netral	0.63	0.79	0.70	314
Possitif	0.74	0.60	0.67	285
Negatif	0.75	0.69	0.72	285

Tabel III adalah jumlah nilai metrik klasifikasi tiap label pada sampel data uji. Berikut penjelasan dari tabel evaluasi metrik diatas.

1) Precision

Jumlah prediksi benar untuk sebuah kelas dibagi dengan total jumlah sampel yang diprediksi untuk kelas tersebut.

2) Recall

Jumlah prediksi benar untuk sebuah kelas dibagi dengan jumlah total sampel actual dari kelas tersebut.

3) F1 - score

Harmonic rata - rata dari *precision* dan *recall*. Mengukur keseimbangan antara keduanya.

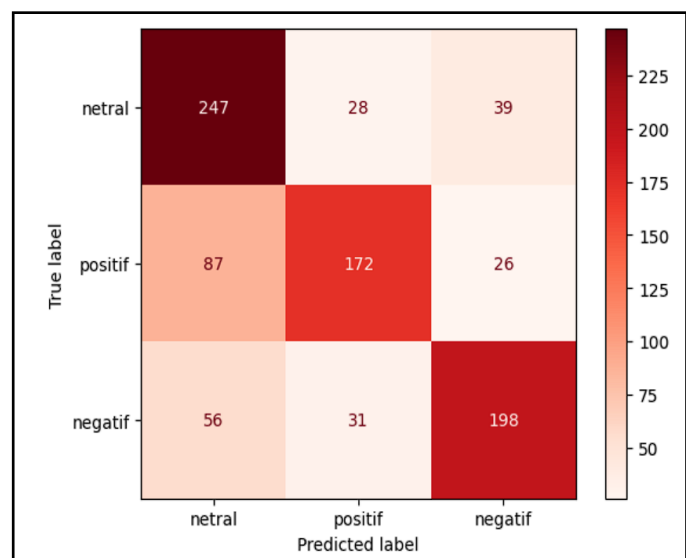
4) Support

Jumlah aktual sampe dari tiap kelas.

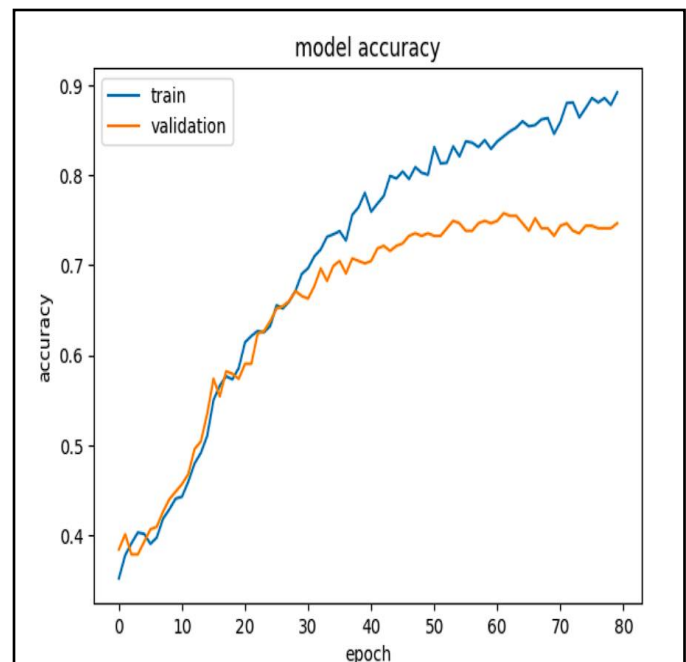
Dari hasil yang sudah didapatkan model ini memiliki akurasi keseluruhan 70%.

	precision	recall	f1-score	support
netral	0.63	0.79	0.70	314
positif	0.74	0.60	0.67	285
negatif	0.75	0.69	0.72	285
accuracy			0.70	884
macro avg	0.71	0.69	0.70	884
weighted avg	0.71	0.70	0.70	884

Gambar 7. Laporan Evaluasi Metrik



Gambar 8. Plot Kinerja Model Klasifikasi



Gambar 9. Grafik model akurasi

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

. berdasarkan hasil penelitian, Semua fungsi dalam preproses yang dilakukan pada 2677 data cuitan Bahasa Indonesia berhasil dilakukan dengan baik. Semua simbol – symbol, karakter non-ASCII, dan tautan yang tidak diperlukan dalam proses klasifikasi berhasil dibersihkan, dan semua huruf kapital pada data berhasil dinormalisasikan menjadi huruf kecil. Proporsi dari prediksi yang benar dalam tiap sampel kelas pada data mendapatkan nilai akurasi sebesar 0.7 atau 70%, yang menunjukkan kinerja dari model seimbang.

B. Saran

Berdasarkan kesimpulan diatas, dalam pembuatan skripsi ini masih diperlukan pengembangan lebih dalam penggunaan algoritma, terutama dalam membangun jumlah unit pada GRU, ataupun dalam pembersihan dataset agar mendapatkan hasil akurasi yang lebih baik dan menghindari *overfitting*.

V. KUTIPAN

- [1] Ahmad Fathan Hidayatullah dkk. 2020. *IOP Conference Series: Materials Science and Engineering* 2020. “Sentiment analysis on twitter using neural network: Indonesian presidential election 2019 dataset”.
- [2] Muhammad Raheel Raza dkk. 2022 *IEEE*. “Performance analysis of Deep Approaches on Airbnb Sentiment Reviews”.
- [3] Dimas Ananda dkk. “Application of Bidirectional Gated Recurrent Unit (BiGRU) in sentiment analysis of tokopedia application users”. *Research Center for Artificial Intelligence and Big Data, Universitas Padjajaran Bandung, Indonesia* 2021.
- [4] Mohan Venkataramaiah dan Nandakumar Achar. *International Journal of Intelligent Engineering and Systems, Vol.13, No.5, 2020*. “Twitter Sentiment Analysis using Aspect-based Bidirectional Gated Recurrent Unit with Self-Attention Mechanism”.
- [5] Agu Edward O. dkk. *Asian Journal of Research in Computer Science* 2023. “Analyzing Election Sentiment in Tweets with Gated recurrent Units”.
- [6] Noor Afiza Mat Razali dkk. *Journal of Big Data (2021)* 8:150. “Opinion Mining for National Security: techniques, domain applications, challenges and research opportunities”.
- [7] Farhan Afthab. “A Comprehensive survey on Sentiment Analysis Techniques”. *International Journal of Technology* 14(6) 1288-1298 (2023)
- [8] Liu, Bing. 2012. “Sentiment Analysis and Opinion Mining”. Chicago: Morgan and Claypool. Publisher. <http://www.dcc.ufrj.br/~valeriab/DTMSentimentAnalysisAndOpinionMining-BingLiu.pdf>.
- [9] Han, J., Kamber, M., dan Pei, J. (2012). “Data mining: concepts and techniques”. (3rd ed.). Elsevier.
- [10] Abdulmoshen Algarni. 2016. “data mining in education”. (*IJACSA*) *International Journal of Advanced Computer Science and Applications. Vol. 7, No. 6, 2016. College of*

Computer Science King Khalid University Abha 61421, Saudi Arabia.

- [11] Batta Mahesh. *International Journal of Science and Research (IJSR) ISSN: 2319-7064* 2018. “Machine Learning Algorithms – A Review”.
- [12] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. *HAL Open Science: hal-04206682* 2015. “Deep Learning”
- [13] Vidushi Sharma, Sachin Rai, Anurag Dev. *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, Issue. 10, October 2012.
- [14] Ilya Sutskever, James Martens, Geoffrey Hinton. *University of Toronto, 6 King’s College Rd., Toronto, ON M5S 3G4 CANADA*. “Generating Text with Recurrent Neural Networks.



Clinton Imanuel Kawatu. Lahir di Kotamobagu, 13 Agustus 1999 dari pasangan suami istri Desmonnt Royke Kawatu dan Hesty Krestin Londa. Penulis merupakan anak pertama dari tiga bersaudara. Alamat tempat tinggal penulis berada di Kelurahan Woloan Dua, Kecamatan Tomohon Barat, Kota Tomohon.

Penulis menempuh Pendidikan Sekolah Dasar di SD Santo Yoseph Denpasar (2006 – 2009), lalu pindah ke SD Katolik Santa Clara Tomohon (2009 – 2011). Kemudian melanjutkan Pendidikan di SMP Katolik Stella Maris Tomohon (2011 – 2014). Selanjutnya penulis melanjutkan Pendidikan di SMK Kristen 1 Tomohon (2014 – 2017). Setelah lulus dari Pendidikan SMK, penulis melanjutkan Pendidikan tingkat Sarjana (S1) di salah satu perguruan tinggi yang adadi Sulawesi Utara yaitu Universitas Sam Ratulangi Manado Jurusan Teknik Elektro Program Studi Teknik Informatika.