

Analisis Sentimen Aplikasi Top Chart Google Play Store Pada Kategori Shopping Menggunakan Metode Gated Recurrent Unit (Studi Kasus Aplikasi Shopee)

1st Prayudha Adhithia Libramawan
Teknik Informatika
Universitas Padjadjaran
Jatinangor, Indonesia
prayudha18001@mail.unpad.ac.id

2nd Muhamad Fahrul Azimi
Teknik Informatika
Universitas Padjadjaran
Jatinangor, Indonesia
muhamad18084@mail.unpad.ac.id

3rd Nazmi Muhammad Abkary
Teknik Informatika
Universitas Padjadjaran
Jatinangor, Indonesia
nazmi18001@mail.unpad.ac.id

4th Hanif Dwi Prasetyo
Teknik Informatika
Universitas Padjadjaran
Jatinangor, Indonesia
hanif18005@mail.unpad.ac.id

5th Difa Bagasputra Maulana
Teknik Informatika
Universitas Padjadjaran
Jatinangor, Indonesia
difa18002@mail.unpad.ac.id

Abstrak— Saat ini perkembangan aplikasi mobile sangat pesat membuat banyak aplikasi diciptakan dengan berbagai kegunaan untuk memenuhi kebutuhan pengguna. Setiap aplikasi memungkinkan pengguna untuk memberi ulasan tentang aplikasi tersebut. Tujuan dari ulasan adalah untuk mengevaluasi dan meningkatkan kualitas produk kedepannya. Pada Google Play Store terdapat kategori-kategori yang memudahkan pengguna dalam mencari aplikasi sesuai dengan kategori, salah satunya kategori *shopping*. Penelitian ini ditujukan untuk menganalisa dan mengetahui *review* terhadap kategori *shopping* dengan satu aplikasi yang bernama Shopee di aplikasi Google Play dengan menggunakan metode *Gated Recurrent Unit* (GRU). Penelitian ini menganalisis *review* pengguna pada aplikasi Shopee di aplikasi Google Play Store menggunakan *Word Embedding* dan algoritma *Gated Recurrent Unit*. Hasil dari analisis menggunakan *Gated Recurrent Unit* menghasilkan akurasi sebesar 77.27% dan kata yang paling sering muncul adalah “belanja” yang muncul lebih dari 1100 kali. Adapun kata “kecewa” dan “lambat” yang muncul lebih dari 400 kali yang menandakan bahwa kekecewaan terhadap penjual serta lambatnya pemrosesan oleh penjual yang ada di Aplikasi Shopee.

Kata Kunci— *Shopping, Gated Recurrent Unit, Analisis Sentimen, Shopee, Google Play Store*

I. PENDAHULUAN

Perkembangan bisnis *online* di Indonesia sangat pesat, hal ini menandakan era pemanfaatan teknologi informasi sudah mulai diakui keberadaannya. Bisnis *online* atau biasa yang disebut dengan *e-commerce* semakin banyak di Indonesia hal ini disebabkan perkembangan internet dan adanya perubahan perilaku konsumen [1]. Banyaknya aplikasi yang terdapat di *internet* membuat setiap orang dapat mengemukakan pendapat mereka terhadap aplikasi tersebut. Di zaman modern ini, opini masyarakat menjadi lebih luas dan lebih bebas untuk diungkapkan di sosial media. Sentimen atau opini masyarakat sangat penting melihat perkembangan teknologi yang serba menggunakan internet. Opini juga dapat berpengaruh dan menjadi sebuah potensi besar bagi pengembang sebuah aplikasi untuk mengetahui umpan balik dari pengguna terhadap aplikasi yang mereka kembangkan.

Terdapat beberapa aplikasi komersial yang terdapat di aplikasi Google Play Store, salah satunya adalah Aplikasi Shopee. Shopee merupakan salah satu perusahaan yang bergerak dibidang aplikasi *e-commerce* dan website secara *online*. Shopee mulai masuk ke pasar Indonesia pada akhir bulan Mei 2015 dan Shopee baru mulai beroperasi pada akhir Juni 2015 di Indonesia untuk memenuhi permintaan dan gaya hidup masyarakat di Indonesia. Shopee merupakan anak perusahaan dari Garena yang berbasis di Singapura yang dipimpin oleh CEO bernama Chris Feng. Shopee telah hadir di beberapa negara di kawasan Asia Tenggara seperti Singapura, Malaysia, Vietnam, Thailand, Filipina, dan Indonesia [2].

Berdasarkan pada Google Play Store, pada tanggal 19 Desember 2021, Shopee telah diunduh lebih dari 100 juta lebih dan memiliki *rating* 4.6 . Pada aplikasi Google Play Store, tercatat 9 juta lebih ulasan di kolom ulasan aplikasi Shopee [3]. *Review* pengguna berisi dua bagian, yaitu nilai *rating* dan komentar secara tekstual. Nilai *rating* menunjukkan evaluasi keseluruhan pengalaman pengguna menggunakan skala numerik, namun komentar tekstual mampu bercerita lebih mendalam [4]. Menurut Ronal Watrighthos, *review* dari pengguna di Play Store, terdiri dari *review* positif dan negatif seperti keluhan, kritik, atau saran [5].

Menurut Indrayuni, Analisis sentimen ulasan pengguna adalah proses menyelidiki ulasan produk di internet untuk menentukan pendapat atau perasaan tentang suatu produk secara umum [6]. Analisis sentimen adalah teknik untuk melihat pendapat perasaan dan subjektivitas teks [7]. Tugas dasar analisis sentimen adalah mengklasifikasikan polaritas sebuah teks yang ada dalam sebuah dokumen kalimat atau opini. Polaritas berarti bahwa isi suatu dokumen kalimat atau opini memiliki aspek positif atau negatif. Oleh karena itu penelitian ini perlu menganalisis pendapat-pendapat tersebut agar dapat dijadikan sebagai tolak ukur untuk mengetahui apakah aplikasi tersebut bagus atau tidak menurut *review* para pengguna. Tujuan analisis sentimen tidak hanya untuk mengumpulkan informasi tentang emosi yang terkandung dalam teks tetapi juga untuk melihat pendapat teks tentang suatu isu yang cenderung positif atau negatif [8].

Penelitian ini ditujukan untuk menganalisa dan mengetahui *review* terhadap satu aplikasi yang bernama Shopee di aplikasi Google Play dengan menggunakan metode *Gated Recurrent Unit* (GRU). Penggunaan Metode GRU dalam penelitian ini karena lapisan nya membantu jaringan saraf dalam menghafal konteks dalam urutan kalimat yang saling bergantung [9].

II. METODE

Penelitian ini menganalisis *review* pengguna pada aplikasi Shopee di aplikasi Google Play Store menggunakan algoritma *Gated Recurrent Unit*. Berikut penjelasan algoritma yang digunakan:

A. Word2vec

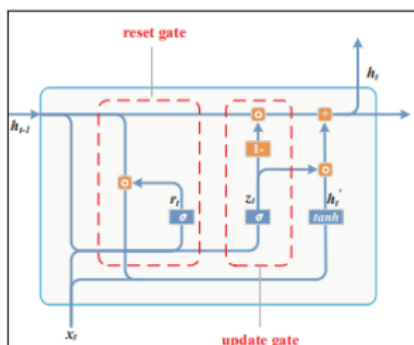
Word Embedding memetakan setiap kata dalam dokumen ke dalam *dense vector*, dimana sebuah vektor merepresentasikan proyeksi kata di dalam ruang vektor. Posisi kata tersebut dipelajari dari teks atau berdasarkan kata – kata disekitarnya. Salah satu algoritma dalam *word embedding* adalah Word2Vec.

Algoritma Word2vec ini diciptakan oleh Mikolov dkk. pada tahun 2013. Word2Vec ini banyak digunakan dalam penelitian NLP. Model ini merupakan salah satu aplikasi *unsupervised learning* menggunakan *neural network* yang terdiri dari *hidden layer* dan *fully connected layer*. Word2Vec mengandalkan informasi lokal dari bahasa semantik yang dipelajari dari kata tertentu dipengaruhi oleh kata – kata sekiranya. Terdapat dua algoritma Word2vec yaitu *Continuous Bag-of-Word* (CBOW) dan *Skip-Gram* [10].

B. Gated Recurrent Unit

Gated Recurrent Unit adalah sebuah pengembangan lanjutan dari arsitektur RNN. GRU merupakan pengembangan yang relatif baru, yang diusulkan oleh K. Cho pada tahun 2014. GRU jauh lebih sederhana dan lebih praktis daripada LSTM. Mirip dengan LSTM, dimana GRU memiliki *unit gated* yang memodulasi aliran informasi dalam unit, tetapi tanpa sel memori yang terpisah. [11].

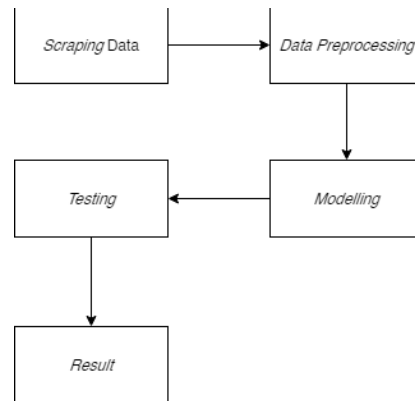
GRU memiliki dua gerbang yang disebut gerbang *reset* dan gerbang *update*. Gerbang *reset* sama dengan gerbang pelupa dan gerbang *input* LSTM, yang akan lebih memilih untuk menggabungkan informasi untuk disimpan atau dibuang dengan input baru ke memori sebelumnya. Gerbang *Update* digunakan untuk menentukan jumlah data yang akan dibuang [12].



Gambar 1. Gerbang pada GRU

III. METODOLOGI

Penelitian ini dilakukan dengan tahapan penelitian yang dapat dilihat pada gambar 2.



Gambar 2. Flowchart alur metodologi

A. Data Collection

Menurut Moraes dkk (2013), salah satu teknik yang cocok dalam pengumpulan data informasi yakni menggunakan teknik *scraping* [13]. Teknik *scraping* merupakan proses pengambilan sebuah dokumen semi-terstruktur dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman di *internet*. Data di-*scraping* menggunakan python yang dilakukan pada tanggal 19 Desember 2021, sejumlah 20,000 buah data yang masing-masing terdiri dari *review* bintang 1 sampai bintang 5, kecuali bintang 3. Bintang 3 tidak diambil karena dianggap netral sedangkan pada penelitian ini, hanya digunakan dua label yaitu positif dan negatif dengan mengasumsikan bahwa bintang 1 dan 2 itu negatif serta bintang 4 dan 5 itu positif.

B. Variabel Atribut

Variabel yang terdapat pada dataset ini berisikan *reviewId*, *userName*, *userImage*, *content*, *score*, *thumbsUpCount*, *reviewCreatedVersion*, *at*, *replyContent*, *repliedAt*, *sortOrder*, dan *appld*. Kemudian dari variabel atribut tersebut yang akan digunakan pada penelitian kali ini adalah *content* dan *score*.

C. Data Collection

Proses pengumpulan data dalam penelitian ini menggunakan teknik *scraping* digunakan untuk mengumpulkan data *review* atau ulasan yang diberikan oleh pengguna pada aplikasi Shopee di Google Play Store (<https://play.google.com/store/apps/details?id=com.shopee.id>).

D. Data Preprocessing

Pada *data preprocessing*, dilakukan beberapa tahapan agar memiliki data yang layak untuk digunakan pada proses berikutnya. Berikut tahapan yang dilakukan untuk mendapatkan data yang layak.

1. Case Folding

Case folding merupakan proses pengubahan kalimat menjadi sama ke bentuk *lower-case*. Proses ini dilakukan agar proses perhitungan menjadi optimal.

2. Noise Removal

Noise removal merupakan sebuah proses untuk membersihkan kata-kata berupa simbol, angka, tanda baca, kata yang diikuti dengan angka hingga URL akan

dihapus pada proses ini agar menghasilkan data yang lebih bersih.

3. Tokenizing

Tokenizing merupakan proses pemisahan atau pemotongan kalimat menjadi kata. Proses ini dilakukan agar memudahkan pada saat mengubah data menjadi bentuk vektor.

4. Stopwords Removal

Stopwords removal dilakukan untuk membuang kata-kata yang tidak memiliki arti yang relevan. Dalam sebuah kalimat, pasti terdapat banyak kata yang dimaksud seperti, kata penghubung, serta kata keterangan yang tidak memiliki makna yang berarti

5. Normalisasi

Normalisasi merupakan proses mengubah kata-kata yang semula tidak baku menjadi kata baku. Kata-kata tidak baku meliputi kata-kata singkatan, kata-kata sehari-hari, serta kata dengan komposisi huruf yang tidak seharusnya.

E. Model

Proses pembentukan model dibentuk dari proses *training* dataset dan dengan penggunaan Word2vec dan algoritma GRU. Penggunaan Word2vec ditujukan untuk mengubah dataset yang awalnya berupa kata-kata menjadi sebuah vektor. Proses pada saat menggunakan algoritma GRU dilakukan pada saat setelah perubahan dataset menjadi vektor untuk membuat sebuah model yang nantinya dapat digunakan untuk menunjang proses analisis sentimen ini.

F. Testing

Proses *testing* pada model dilakukan menggunakan inputan untuk menentukan apakah model yang telah dibentuk ini layak untuk digunakan atau tidak. Hasil yang keluar akan berupa persentase dan prediksi pada kata-kata yang dimasukkan.

G. Result

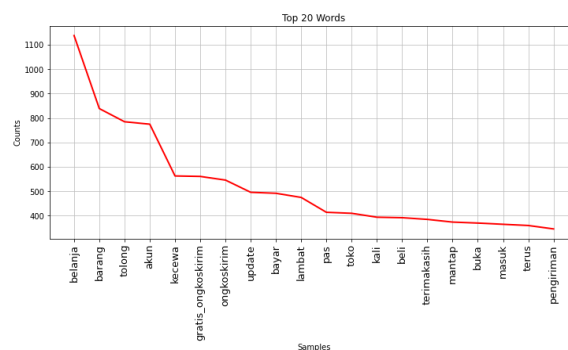
Setelah semua proses dilakukan yang dimulai dari *scraping* data, *preprocessing* data, pembentukan model, *testing*, dan sebuah proses dalam *deep learning*, didapatkan sebuah model yang layak untuk digunakan dalam menentukan sebuah sentimen yang ada di tiap halaman ulasan Google Play Store. Hasil yang berupa model tersebut sesuai dengan apa yang direncanakan untuk mendapatkan sebuah informasi pengetahuan yang nantinya dapat berguna untuk tiap pemilik aplikasi.

IV. HASIL DAN ANALISIS

Proses *scraping* dataset dilakukan terhadap aplikasi Shopee yang ada di Google Play Store menggunakan Python dengan *library google-play-scraper*. Setelah proses *scraping* selesai, lanjut ke proses pelabelan. Pada proses pelabelan data, terdapat dua buah label yang terdiri dari angka 0 dan 1. Angka 0 merepresentasikan bintang 1 dan 2 sedangkan untuk angka 1 merepresentasikan bintang 4 dan 5. Kemudian dilanjutkan ke proses *preprocessing*. Pada proses ini, dilakukan seluruh

rangkaian yang telah dituliskan pada bagian *Data Preprocessing*. Lalu proses selanjutnya adalah pembagian data yang terdiri dari $\frac{1}{3}$ data test dan $\frac{2}{3}$ adalah data train.

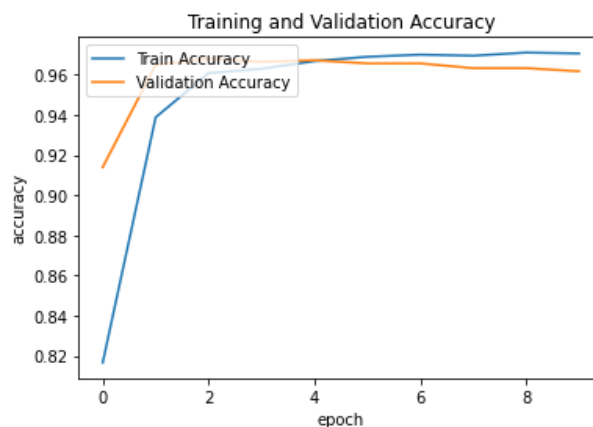
Selanjutnya, proses dilanjutkan dengan *word embedding*, yaitu merubah teks menjadi vektor yang digunakan untuk pembentukan model. Algoritma *word embedding* yang digunakan adalah Word2vec dengan *library gensim*. Dalam proses *word embedding* ini dilakukan pengurutan kata-kata menjadi 20 kata teratas yang divisualisasikan dalam grafik berikut :



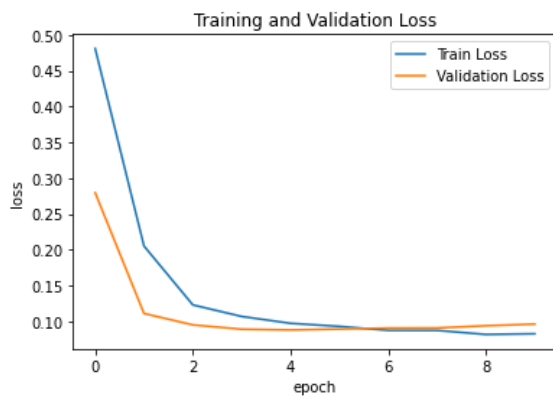
Gambar 3. Grafik 20 Kata Teratas yang Sering Muncul

Dari grafik tersebut dapat disimpulkan bahwa kata yang sering muncul pada dataset yang penulis uji adalah tentang “belanja” yang muncul lebih dari 1100 kali dan kata selanjutnya yaitu kata “barang” yang muncul lebih dari 800 kali. Setelah itu data diproses dan diperoleh model yang akan dilanjutkan dalam proses algoritma *Gated Recurrent Unit*.

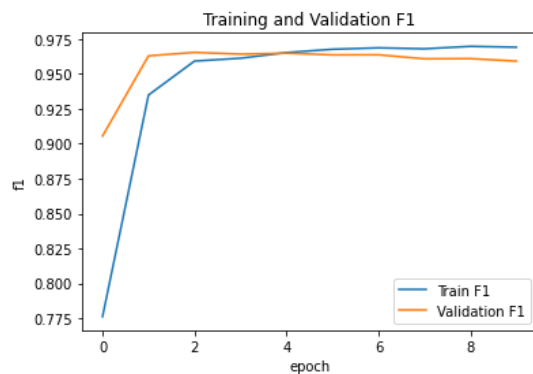
Penggunaan algoritma *Gated Recurrent Unit* dilakukan untuk membuat model *Neural Network*. Setelah model selesai dibentuk, dilanjutkan dengan proses testing dengan perhitungan akurasi, *precision*, *recall*, *F1 Score*, dan *Confusion Matrix*. Setelah perhitungan sebelumnya dilakukan. Model divisualisasikan dalam bentuk grafik seperti pada dibawah ini :



Gambar 4. Grafik Akurasi dari data Training dan Validasi



Gambar 5. Grafik Loss dari data Training dan Validasi



Gambar 6. Grafik Loss dari data Training dan Validasi

Berdasarkan visualisasi tersebut dapat dilihat bahwa model yang dihasilkan pada proses K-Fold ke 10 menghasilkan model yang cukup baik dikarenakan tidak terlalu banyak terjadi *overfitting* antara data Training dan Validasi. Selain itu tingkat akurasi pada visualisasi tersebut lebih tinggi dan tingkat *Loss* nya pun rendah.

TABLE I. HASIL PERBANDINGAN FOLD

Fold	Akurasi (%)	Loss (%)	F1 (%)
1	80.37	49.28	78.94
2	79.23	20.97	78.14
3	78.61	17.25	77.41
4	78.10	15.34	77.08
5	77.35	12.46	76.56
6	77.26	11.34	75.90
7	77.32	10.08	76.23
8	77.08	11.63	76.09
9	77.19	9.76	76.17
10	77.27	9.62	76.16

Penulis juga mencoba model yang dibentuk menggunakan input kalimat yang akan menghasilkan antara *positive* atau *negative* beserta persentase *confidence*-nya. Berikut hasil yang diberikan:

TABLE II. HASIL PERSENTASE *CONFIDENCE*

Kata	Label	Confidence
Ongkir gratis memang terbaik	Positive	99.83%
Kecewa Sekali	Negative	83.07%
Baik saya kecewa	Negative	85.34%
Benci banget buka fitur shopee	Negative	94.40%
Belanja di Shopee lebih murah	Positive	96.53%

V. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan dan ditampilkan pada Gambar 3, disimpulkan bahwa kata yang sering muncul pada dataset yang penulis uji adalah tentang “belanja” yang muncul lebih dari 1100 kali dan disusul oleh kata “barang” yang muncul lebih dari 800 kali. Adapun kata “kecewa” dan “lambat” yang muncul lebih dari 400 kali yang menandakan bahwa kekecewaan terhadap penjual serta lambatnya pemrosesan oleh penjual yang ada di Aplikasi Shopee.

Hasil dari proses analisis sentimen terhadap $\frac{1}{3}$ data dari dataset yang menggunakan nilai pengujian fold=10 pada metode *Gated Recurrent Unit*(GRU) diperoleh akurasi sebesar 77.27% serta nilai F1 sebesar 76.16%.

Pada penelitian ini, data ulasan yang digunakan adalah berkategori relevan dan yang terbaru. Di penelitian ini juga terdapat bentuk *test* yang memungkinkan penulis menginputkan kata-kata atau kalimat yang nantinya dapat dianalisis dan mengeluarkan hasil antara *positive* atau *negative* beserta nilai *confidence*-nya.

REFERENSI

- [1] I. Sidharta and B. Suzanto, “Pengaruh Store Atmosphere, Inovasi Produk Dan Customer Experience Terhadap Kepuasan Pelanggan Cafe Warunk Upnormal Di Banjarmasin Kalimantan Selatan,” *J. Bissmart*, vol. 1, no. 1, pp. 1–23, 2020, [Online]. Available: <http://jurnalmahasiswa.stesia.ac.id/index.php/jirm/article/download/1753/1763>.
- [2] F. Margaretha, “Analisis Hubungan Antara Motif Dengan Tingkat Kepuasan Pengguna Aplikasi Shopee Sebagai Media Berbelanja Online Pada Shopeeholics Di Kota Samarinda,” *eJournal Ilmu Komun.*, vol. 5, no. 4, pp. 26–40, 2017.
- [3] Google Play, “Shopee: Online Shopping,” *Google Play*, 2021. <https://play.google.com/store/apps/details?id=com.shopee.id>.
- [4] L. B. Ilmawan and E. Winarko, “Aplikasi Mobile untuk Analisis Sentimen pada Google Play,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 9, no. 1, p. 53, 2015, doi: 10.22146/ijccs.6640.
- [5] R. Watrionthos, S. Suryadi, D. Irmayani, M. Nasution, and E. F. S. Simanjorang, “Sentiment analysis of traveloka app using naïve bayes classifier method,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 7, pp. 786–788, 2019, doi: 10.31227/osf.io/2dbe4.
- [6] E. Indrayuni, “Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization,” *J. Evolusi Vol. 4 Nomor 2 - 2016*, vol. 4, no. 2, pp. 20–27, 2016.
- [7] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.
- [8] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi:

10.31294/ji.v8i2.9681.

- [9] S. S. and R. D. Seenaiiah Pedipina, "Sentiment Analysis of Twitter Political Data using GRU Neural Network," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 06, pp. 5307–5320, 2020, [Online]. Available: <http://sersc.org/journals/index.php/IJAST/article/view/19592>.
- [10] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, and Z. Abidin, "Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks," *J. Tekno Kompak*, vol. 14, no. 2, p. 74, 2020, doi: 10.33365/jtk.v14i2.732.
- [11] S. Biswas, E. Chadda, and F. Ahmad, "Sentiment analysis with gated recurrent units," *Adv. Comput. Sci. Inf. Technol.*, vol. 2, no. 11, pp. 59–63, 2015.
- [12] R. R. Nuli Giarsyani, Ahmad Fathan Hidayatullah, "KOMPARASI ALGORITMA MACHINE LEARNING DAN DEEP LEARNING UNTUK NAMED ENTITY RECOGNITION: STUDI KASUS DATA KEBENCANAAN," *JIRE (Jurnal Inform. Rekayasa Elektron.*, vol. 3, no. 1, pp. 48–57, 2020, doi: <https://doi.org/10.36595/jire.v3i1.222>.
- [13] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, 2013, doi: 10.1016/j.eswa.2012.07.059.