

## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

## PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

	FORM PENGAJUAN JUDU	L
Nama	: Talitha Syafiyah	
NIM	: 211402018	
Judul diajukan oleh*	: Dosen	
	✓ Mahasiswa	
Bidang Ilmu (tulis dua bidang)	Data Science and Intelligent System	m
Uji Kelayakan Judul**	: O Diterima O Ditolak	
Hasil Uji Kelayakan Judul :		
Calon Dosen Pembimbing I: Fani	ndia Purnamasari S. T.I. M.IT.	
Calon Dosen Pembimbing I: Fanis		Paraf Calon Dosen Pembimbing I
caron Booth I emonitoring II. Ivui	. vaja 0.01., 111.120111.	
		Medan, 14 Oktober 2024
		Ka. Laboratorium Penelitian,
		,

\*\* Pilih salah satu

Fanindia Purnamasari S.TI., M.IT

NIP. 198908172019032023

<sup>\*</sup> Centang salah satu atau keduanya



Judul / Topik Skripsi

# KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

### UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

### RINGKASAN JUDUL YANG DIAJUKAN

IDENTIFIKASI PROVOKASI DOXING DI MEDIA SOSIAL X

IMPLEMENTASI ALGORITMA GATED RECURRENT UNIT UNTUK

\*Semua kolom di bawah ini diisi oleh mahasiswa yang sudah mendapat judul

Latar Belakang dan Penelitian	Latar Belakang				
Terdahulu	Media sosial telah menjadi <i>platform</i> yang sangat populer bagi masyarakat untuk berinteraksi dan				
	berbagi informasi. Platform media sosial seperti Facebook, Instagram, dan Twitter				
	memungkinkan pengguna untuk berbagi informasi secara luas. Namun, di balik kemudahan				
	tersebut, media sosial juga menjadi ladang subur bagi kejahatan siber, salah satunya adalah				
	doxing. Doxing, yang merupakan kependekan dari "dropping documents," adalah praktik pengungkapan informasi pribadi secara daring tanpa izin, dengan tujuan untuk mengintimidasi,				
	melecehkan, atau merugikan korban. Belakangan ini, doxing tidak hanya dilakukan o				
	individu, tetapi juga melibatkan upaya kolektif dalam bentuk provokasi yang mendorong				
	partisipasi pengguna lain untuk ikut menyebarkan informasi pribadi secara masif.				

Dalam beberapa tahun terakhir, fenomena provokasi *doxing* telah menjadi salah satu tren yang paling mengkhawatirkan di dunia digital. Provokasi *doxing* melibatkan usaha untuk menggerakkan orang lain agar berpartisipasi dalam penyebaran informasi pribadi seseorang secara daring tanpa persetujuan. Tindakan ini sering kali muncul dalam bentuk pesan provokatif di media sosial, yang mengundang pengguna lain untuk ikut menyebarkan informasi sensitif, seperti alamat, nomor telepon, atau data pribadi lainnya. Motivasi provokasi ini sering kali didorong oleh rasa kebencian, dendam, atau keinginan mempermalukan target. Namun, provokasi untuk melakukan *doxing* kerap dianggap sebagai bentuk solidaritas atau aksi sosial.

Menurut penelitian yang dilakukan pada tahun 2021 oleh SAFEnet (Southeast Asia Freedom of Expression Network), *doxing* adalah kejahatan kedua yang paling sering terjadi setelah peretasan, dengan 24 kasus (12,43% dari total kasus). Menurut penelitian tersebut, dari tahun 2017 hingga 2021, jumlah kasus *doxing* di Indonesia meningkat setiap tahunnya. Pada tahun 2017, terdapat 1 kasus, pada tahun 2018, terdapat 2 kasus, pada tahun 2019, terdapat 7 kasus, pada tahun 2020 terdapat 13 kasus, dan pada tahun 2021 terdapat 24 kasus. Jurnalis menyumbang 56% korban *doxing*, 22% aktivis, dan 22% warga sipil (SAFEnet, 2022). Meskipun angka tersebut menggambarkan kasus *doxing* secara umum, provokasi *doxing* turut menjadi fenomena yang berbahaya karena mampu mempercepat penyebaran informasi dan menggerakkan banyak orang untuk terlibat dalam tindakan tersebut.



## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

Bahayanya, mereka yang terlibat dalam provokasi *doxing* sering kali merasa terlepas dari tanggung jawab moral. Partisipasi dianggap sekadar "membantu" atau "ikut-ikutan," tanpa mempertimbangkan dampak besar bagi korban. Akibatnya, korban *doxing* tidak hanya menghadapi risiko privasi yang terancam, tetapi juga ancaman fisik, intimidasi psikologis, hingga kerugian reputasi. Fenomena ini menunjukkan bahwa provokasi *doxing* bukan sekadar tindakan individu, tetapi masalah kolektif yang dapat berdampak luas dan menimbulkan ancaman serius bagi keamanan digital.

Penelitian sebelumnya telah dilakukan untuk mendeteksi *doxing* di media sosial. (Karimi et al., 2022) dalam "Automated Detection of *Doxing* on Twitter" mengembangkan algoritma pendeteksian *doxing* otomatis yang menggunakan algoritma Support Vector Machine (SVM) untuk membedakan pengungkapan informasi pribadi oleh pihak ketiga dan pengungkapan yang dilakukan oleh individu sendiri. Hasil penelitian ini menunjukkan akurasi deteksi sebesar 96,86% dan recall 97,37%, memberikan kontribusi signifikan dalam pengembangan sistem deteksi *doxing* yang lebih efektif. Namun, fokus penelitian tersebut hanya pada deteksi umum *doxing*, tanpa mempertimbangkan provokasi *doxing* sebagai bentuk ancaman spesifik yang dapat memobilisasi tindakan berbahaya oleh pengguna lain. Selain itu, penelitian ini hanya meneliti dua jenis informasi pribadi, yaitu SSN dan IP address, yang membuat cakupan deteksinya menjadi terbatas.

Penelitian terkait lainnya dilakukan oleh (Patihullah and Winarko, 2019) dalam "Hate Speech Detection for Indonesia Tweets Using Word Embedding And Gated Recurrent Unit". Penelitian ini bertujuan mendeteksi ujaran kebencian menggunakan dataset cuitan *platform* Twitter. Metode yang digunakan melibatkan Word2vec untuk ekstraksi fitur dan Gated Recurrent Unit (GRU) sebagai algoritma klasifikasi. Kombinasi Word2vec dan GRU berhasil mencapai akurasi 92,96% dalam mendeteksi ujaran kebencian di *platform* tersebut. Selain itu, penelitian ini juga mengeksplorasi pengaruh *learning rate* dan jumlah neuron pada *hidden layers*, kemudian menyimpulkan bahwa kombinasi *learning rate* sebesar 0,001 dan 200 neuron menghasilkan kinerja terbaik.

Dalam penelitian ini, penulis mengajukan penelitian menggunakan algoritma Gated Recurrent Unit (GRU). GRU telah menjadi salah satu model paling populer dalam pemrosesan bahasa alami (NLP) dan data sekuensial sejak diperkenalkan sebagai varian yang lebih sederhana dari Long Short-Term Memory (LSTM). GRU menunjukkan potensi besar dalam menangani data sekuensial yang panjang tanpa mengalami masalah *vanishing gradient*, yang sering terjadi pada model RNN konvensional. Dibandingkan dengan LSTM, GRU menawarkan



## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

kecepatan pelatihan yang lebih cepat dan struktur yang lebih sederhana karena hanya menggunakan dua gerbang, yaitu *update gate* dan *reset gate*, namun tetap mempertahankan kemampuan untuk menangkap dependensi jangka panjang dalam data. Keunggulan ini menjadikan GRU sebagai alternatif yang lebih efisien dalam berbagai aplikasi *real-time* yang memerlukan model yang ringan dan cepat, tanpa mengorbankan akurasi prediksi.

Berdasarkan latar belakang masalah yang telah dijelaskan, provokasi *doxing* atau dorongan untuk mengungkap informasi pribadi secara publik merupakan ancaman serius terhadap keamanan digital yang dapat memicu tindakan *doxing* sesungguhnya dan menimbulkan risiko signifikan terhadap privasi dan keselamatan individu. Oleh karena itu, penulis mengusulkan penelitian berjudul "IMPLEMENTASI ALGORITMA *GATED RECURRENT UNIT* UNTUK IDENTIFIKASI PROVOKASI *DOXING* DI MEDIA SOSIAL X". Penelitian ini bertujuan mengembangkan sistem berbasis web yang dapat mengidentifikasi cuitan di media sosial X yang mengandung provokasi untuk melakukan *doxing* dengan mengimplementasikan algoritma *Gated Recurrent Unit* (GRU).

#### Penelitian Terdahulu

No.	Penulis	Judul	Tahun
1.	Younes Karimi, Anna Squicciarini, Shomir Wilson	Automated Detection of <i>Doxing</i> on Twitter	2022
2.	Max Aliapoulios, Kejsi Take, Prashanth Ramakrishna, Daniel Borkan, Beth Goldberg, Jeffrey Sorensen, Anna Turner, Rachel Greenstadt, Tobias Lauinger, Damon McCoy	A Large-Scale Characterization of Online Incitements to Harassment Across <i>Platforms</i>	2021
3.	A K M Nuhil Mehdy & Hoda Mehrpouyan	A Multi-Input Multi-Output Transformer-Based Hybrid Neural Network for Multiclass Privacy Disclosure Detection	2021



# UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

## PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

4.	A K M Nuhil Mehdy & Hoda Mehrpouyan	A User-Centric and Sentiment Aware Privacy-Disclosure Detection Framework Based on Multi-Input Neural Network	2020
5.	Chandan Akiti, Anna Squicciarini & Sarah Rajtmajer	A Semantics-based Approach to Disclosure Classification in User- Generated Online Content	2020
6.	Awwaliyah Aliyah, Nailah Azzahra, Aliffia Isma Putri & Nur Aini Rakhmawati	Analisis Sentimen Twitter terhadap Tren Penyebaran Informasi Pelaku Kejahatan Menggunakan Algoritma Naïve Bayes	2024
7.	Junanda Patihullah & Edi Winarko	Hate Speech Detection for Indonesia Tweets Using Word Embedding and Gated Recurrent Unit	2019
8.	Angela Marpaung, Rita Rismala & Hani Nurrahmi	Hate Speech Detection in Indonesian Twitter Texts using Bidirectional Gated Recurrent Unit	2021



## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

	9.	Amrutha B R & Bindu K R	Detecting Hate Speech in Tweets Using Different Deep Neural Network Architectures	2019
--	----	-------------------------	---	------

#### Rumusan Masalah

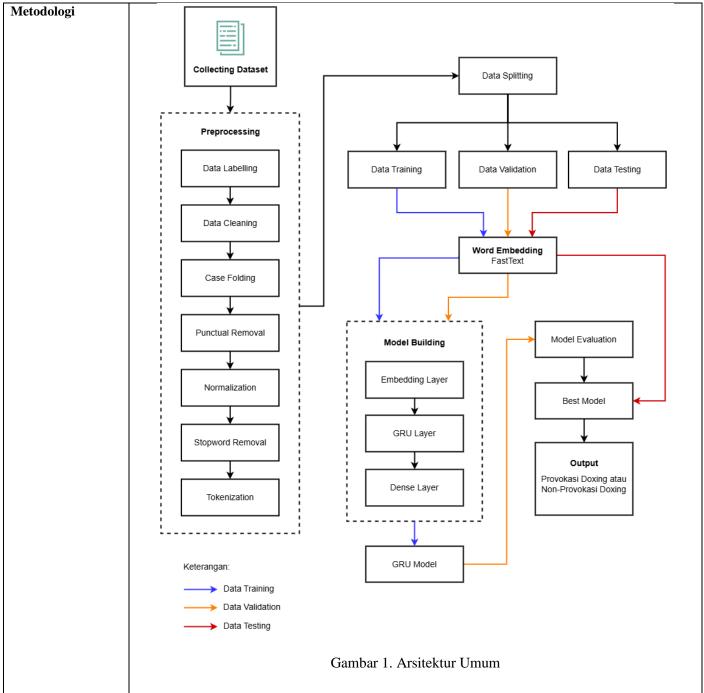
Provokasi doxing, yaitu upaya mendorong atau mengajak orang lain untuk mengungkap informasi pribadi seseorang secara daring tanpa izin, semakin sering muncul di platform media sosial. Karakteristik media sosial yang memungkinkan informasi menyebar dengan cepat memperburuk situasi, di mana provokasi-provokasi tersebut berpotensi memicu tindakan berbahaya seperti intimidasi, ujaran kebencian, hingga teror terhadap individu tertentu. Meskipun secara hukum provokasi doxing belum sepenuhnya dikategorikan sebagai tindakan ilegal, dampaknya tetap signifikan karena dapat mendorong banyak orang untuk terlibat dalam pengungkapan informasi pribadi orang lain. Permasalahan ini semakin kompleks karena sifat provokasi doxing yang sering disamarkan sebagai "aksi solidaritas" atau "kampanye sosial" untuk memberikan "sanksi sosial" sehingga sulit teridentifikasi secara langsung. Oleh karena itu, diperlukan sistem deteksi yang mampu mengidentifikasi provokasi doxing di media sosial, khususnya dalam konteks bahasa Indonesia. Sistem ini diharapkan dapat membantu meningkatkan kesadaran pengguna, meminimalkan risiko penyalahgunaan dan penyebaran informasi pribadi, dan berperan sebagai langkah preventif dalam menjaga keamanan serta privasi digital.



## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id



Gambar 1 menunjukkan arsitektur umum penelitian yang terdiri dari beberapa tahapan utama sebagai berikut.

### 1. Collecting Dataset

Data yang digunakan dalam penelitian ini dikumpulkan dari *platform* media sosial X (sebelumnya dikenal sebagai Twitter) dengan metode crawling. Fokus pengumpulan data adalah cuitan berbahasa Indonesia yang mengandung potensi provokasi *doxing*, contohnya menggunakan frasa seperti "spill aja NIK-nya biar mampus", "sebarkan



## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

alamat rumah", atau seruan lainnya yang mendorong pengungkapan informasi pribadi tanpa izin. Data yang terkumpul disimpan dalam format CSV dan digunakan sebagai dataset untuk membangun model identifikasi provokasi *doxing*.

#### 2. Preprocessing

Pada tahap preprocessing, teks mentah dari cuitan yang dikumpulkan diubah menjadi format yang lebih bersih dan siap untuk diproses. Tahapan-tahapan preprocessing meliputi:

- a. Data Labeling, mmberikan label pada data yang relevan sesuai kategori, yaitu Provokasi *Doxing* dan Non-Provokasi *Doxing*.
- b. Data Cleaning, menghapus data duplikat dan entri yang tidak relevan, seperti cuitan kosong atau yang tidak memuat informasi penting.
- c. Case Folding, mengubah semua huruf dalam cuitan menjadi huruf kecil agar seragam dan menghindari perbedaan karena kapitalisasi.
- d. Punctual Removal: Menghapus tanda baca yang tidak relevan (misalnya titik, koma, tanda tanya) yang tidak memiliki pengaruh pada konteks makna teks.
- e. Normalization, mengubah kata-kata yang ditulis dengan variasi yang berbeda ke dalam bentuk yang standar, misalnya "nomor hp" menjadi "nomor telepon".
- f. Stopword Removal, menghilangkan kata-kata umum seperti "dan", "di", "yang" yang tidak memiliki kontribusi signifikan terhadap pengklasifikasian.
- g. Tokenization, memecah teks cuitan menjadi kata-kata individu (token), yang nantinya akan digunakan sebagai input ke model.

### 3. Data Splitting

Pada tahap ini, data yang telah dikumpulkan dibagi menjadi tiga bagian: data latih (training), data validasi (validation), dan data uji (testing). Data latih digunakan untuk membangun dan melatih model agar mampu mempelajari pola dari data. Data validasi berfungsi untuk menguji performa model selama proses pelatihan dan membantu menghindari overfitting dengan menyesuaikan parameter model. Terakhir, data uji digunakan untuk mengevaluasi kinerja model secara objektif setelah pelatihan selesai.

### 4. Word Embedding

Setelah tahap preprocessing, data teks yang sudah bersih diubah menjadi representasi numerik menggunakan *library* FastText. FastText mengonversi setiap kata ke dalam vektor yang menangkap informasi semantik dan sintaksis dari kata-kata tersebut. Representasi vektor ini sangat penting karena model pembelajaran mesin hanya dapat



## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

bekerja dengan data numerik, dan embedding FastText memungkinkan model untuk memahami hubungan antar kata.

#### 5. Model Building

Pada tahap ini, dilakukan pembangunan model dengan menggunakan algoritma Gated Recurrent Unit (GRU). GRU adalah salah satu jenis Recurrent Neural Network (RNN) yang digunakan untuk memproses data sekuensial seperti teks, karena kemampuannya dalam menyimpan konteks urutan data secara efisien. Arsitektur model terdiri dari beberapa lapisan sebagai berikut:

- a. Embedding Layer, merupakan layer pertama yang menerima input berupa vektor hasil embedding dari FastText.
- b. GRU Layer, merupakan lapisan inti yang bertugas memproses urutan kata dalam cuitan, dengan mempertimbangkan konteks dari kata-kata sebelumnya.
   GRU efektif dalam menangkap hubungan jangka panjang dalam data teks.
- c. Dense Layer, lapisan Dense digunakan untuk mengubah hasil GRU menjadi representasi yang lebih sederhana.

### 6. Model Evaluation

Setelah model terbentuk, kinerjanya dievaluasi menggunakan data validasi dan data uji. Proses evaluasi mengukur metrik seperti akurasi, precision, recall, dan f1-score untuk memastikan model berfungsi dengan baik.

### 7. Best Model

Setelah melalui tahap pelatihan dan validasi, model yang telah terlatih berhasil dibentuk. Model ini mampu mengenali pola-pola yang berkaitan dengan provokasi doxing. Selanjutnya, model ini siap untuk diuji menggunakan data uji guna mengukur kinerjanya dan memastikan performa yang optimal.

#### 8. Output

Model yang telah dilatih akan mengklasifikasikan cuitan ke dalam dua kelas:

- a. Provokasi *Doxing*, di mana cuitan berisi provokasi atau dorongan untuk mengungkap informasi pribadi seseorang tanpa izin, misalnya, "Sebar alamatnya biar kena batunya."
- b. Non-Provokasi *Doxing*, di mana cuitan tidak mengandung provokasi *doxing*, baik berupa diskusi umum maupun konten lain yang tidak terkait.



## UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

#### PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155 Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

-		•		
к	ei	e	re	nsi

- Akiti, C., Squicciarini, A., Rajtmajer, S., 2020. A semantics-based approach to disclosure classification in user-generated online content. Find. Assoc. Comput. Linguist. Find. ACL EMNLP 2020 3490–3499. https://doi.org/10.18653/v1/2020.findings-emnlp.312
- Aliapoulios, M., Take, K., Ramakrishna, P., Borkan, D., Goldberg, B., Sorensen, J., Turner, A., Greenstadt, R., Lauinger, T., McCoy, D., 2021. A Large-Scale Characterization of Online Incitements to Harassment Across Platforms. Proc. ACM SIGCOMM Internet Meas. Conf. IMC 621–638. https://doi.org/10.1145/3487552.3487852
- Aliyah, A., Azzahra, N., Isma Putri, A., Aini Rakhmawati, N., Teknologi Sepuluh Nopember, I., Teknik Kimia, J., Sukolilo, K., Koresprodensi Penulis, S., 2024. Analisis Sentimen Twitter Terhadap Tren Penyebaran Informasi Pelaku Kejahatan Menggunakan Algoritma Naives Bayes 85–97.
- Amrutha, B.R., Bindu, K.R., 2019. Detecting hate speech in tweets using different deep neural network architectures. 2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019 923–926. https://doi.org/10.1109/ICCS45141.2019.9065763
- Karimi, Y., Squicciarini, A., Wilson, S., 2022. Automated Detection of Doxing on Twitter. Proc. ACM Human-Computer Interact. 6, 1–24. https://doi.org/10.1145/3555167
- Marpaung, A., Rismala, R., Nurrahmi, H., 2021. Hate Speech Detection in Indonesian Twitter Texts using Bidirectional Gated Recurrent Unit. KST 2021 2021 13th Int. Conf. Knowl. Smart Technol. 186–190. https://doi.org/10.1109/KST51265.2021.9415760
- Nuhil Mehdy, A.K.M., Mehrpouyan, H., 2021. A Multi-Input Multi-Output Transformerbased Hybrid Neural Network for Multiclass Privacy Disclosure Detection.
- Nuhil Mehdy, A.K.M., Mehrpouyan, H., 2020. A user-centric and sentiment aware privacy-disclosure detection framework based on multi-input neural network. CEUR Workshop Proc. 2573, 21–26.
- Patihullah, J., Winarko, E., 2019. Hate Speech Detection for Indonesia Tweets Using Word Embedding And Gated Recurrent Unit. IJCCS (Indonesian J. Comput. Cybern. Syst. 13, 43. https://doi.org/10.22146/ijccs.40125
- SAFEnet, 2022. Represi Digital di Indonesia Masih Terus Berlanjut [WWW Document]. URL https://safenet.or.id/id/2022/03/represi-digital-di-indonesia-masih-terus-berlanjut-sepanjang-2021/

Medan, 14 Oktober 2024 Mahasiswa yang mengajukan,

Talitha Syafiyah NIM. 211402018