

**KLASIFIKASI TWEET PELECEHAN *ONLINE* PADA MEDIA
SOSIAL X MENGGUNAKAN *GATED RECURRENT UNIT***

SKRIPSI

INDAH MARIANA NAINGGOLAN

191402024



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

**KLASIFIKASI TWEET PELECEHAN *ONLINE* PADA MEDIA SOSIAL X
MENGUNAKAN *GATED RECURRENT UNIT***

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Teknologi Informasi

INDAH MARIANA NAINGGOLAN

191402024



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

PERSETUJUAN

Judul

: KLASIFIKASI TWEET PELECEHAN ONLINE PADA
MEDIA SOSIAL X MENGGUNAKAN GATED
RECURRENT UNIT

Kategori

: Skripsi

Nama Mahasiswa

: Indah Mariana Nainggolan

Nomor Induk Mahasiswa

: 191402024

Program Studi

: Sarjana (S-1) Teknologi Informasi

Fakultas

: Ilmu Komputer dan Teknologi Informasi

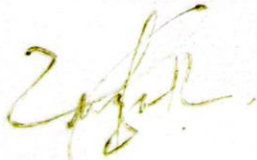
Universitas Sumatera Utara

Medan, 4 Juli 2024

Komisi Pembimbing:

Pembimbing 2,

Pembimbing 1,



Umayra Ramadhani Putri Nasution, S.TI., M.Kom.

Sarah Purnamawati, S.T., M.Sc.

NIP. 199104112021022001

NIP. 198302262010122003

Diketahui/disetujui oleh

Program Studi S-1 Teknologi Informasi

Ketua,



Dedy Ansandi, S.T., M.Kom.

NIP. 197908312009121002

PERNYATAAN**KLASIFIKASI PELECEHAN *ONLINE* PADA MEDIA SOSIAL X
MENGUNAKAN *GATED RECURRENT UNIT*****SKRIPSI**

Saya mengakui bahwa skripsi ini merupakan hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 27 April 2024



INDAH MARIANA NAINGGOLAN

191402024

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur dan terima kasih kepada Tuhan Yesus Kristus, karena atas berkat dan kasih-Nya penulis dapat menuntaskan penyusunan skripsi yang berjudul “Klasifikasi *Tweet* Pelecehan *Online* Pada Media Sosial X Menggunakan *Gated Recurrent Unit*” diaman hal ini diperlukan sebagai persyaratan untuk dapat meraih gelar Sarjana Komputer dari Program Studi S1 Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi.

Penulis dapat menyelesaikan penyusunan skripsi disertai dengan dukungan, doa dan semangat dari seluruh pihak yang terlibat selama perjalanan ini. Oleh karena itu, penulis mengucapkan terima kasih yang setulus-tulusnya kepada :

1. Keluarga Penulis, Ayah Marukil Nainggolan dan Ibu Ramlia Panjaitan, serta saudara penulis Van Asel Nainggolan yang selalu memberikan doa, semangat, dukungan hingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Dr.Maya Silvi Lidya,M.Sc selaku Dekan Fasilkom-TI USU.
3. Bapak Dedy Arisandi ST., M.Kom., selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Bapak Ivan Jaya S.Si., M.Kom., selaku Sekretaris Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
4. Ibu Sarah Purnamawati S.T.,M.Sc., selaku Dosen Pembimbing 1 dan Ibu Umayya Putri Nasution S.TI.,M.Kom selaku Dosen Pembimbing 2 yang sudah memberikan bimbingan dengan memberikan saran dan masukan kepada penulis hingga poses penyelesaian skripsi ini.
5. Ibu Ulfi Andayani S.Kom.,M.Kom dan Ibu Fanindia Purnamasari S.TI.,M.IT selaku dosen pembimbing penulis yang telah banyak memberikan saran kepada penulis.
6. Dosen Pengajar di Program Studi Teknologi Informasi Universitas Sumatera Utara.
7. Staff dan pegawai di Program Studi Teknologi Informasi Universitas Sumatera Utara.

8. Teman seperjuangan penulis sejak maba Margaretha Gok Asi Naibaho dan Putri Laura Tobing yang telah memberikan banyak hal positif untuk membangun motivasi kepada penulis.
9. Sahabat Penulis yang selalu memberi dukungan dan doa dalam penyelesaian skripsi ini, Aurora Ivana Simanjuntak.
10. Semua pihak yang tidak dapat dituliskan satu per satu yang selalu memberikan bantuan, perhatian dan doa kepada penulis dalam penyelesaian skripsi ini.
Penulis mengucapkan terima kasih.

Medan, 27 April 2024



INDAH MARIANA NAINGGOLAN

191402024

ABSTRAK

Saat ini, perkembangan teknologi komunikasi dan informasi sangat berkembang pesat dalam berbagai bidang kehidupan manusia, terutama dalam penggunaan media sosial. Salah satu media sosial paling banyak digunakan adalah X (twitter) yang dimana sering digunakan untuk berbagai informasi secara *real-time*, termasuk berita terbaru dan memberikan ruang untuk berdiskusi terhadap topik tertentu. Akan tetapi, penggunaan X ini tidak hanya memberikan dampak positif tetapi juga dampak negatif dimana banyak pihak-pihak atau pengguna yang menyalahgunakan fungsi dari X itu sendiri sehingga meningkatnya kejahatan yang muncul, salah satunya adalah pelecehan *online* (*online harassment*). Mengidentifikasi pelecehan *online* sangat penting dilakukan guna melindungi pengguna dan mencegah semakin meluasnya praktik pelecehan dan kekerasan secara *online* di Indonesia. Tetapi, proses ini tentunya akan membutuhkan waktu yang cukup lama jika dilakukan secara manual. Oleh sebab itu, dibutuhkan sebuah pendekatan yang dapat mengolah data *tweet* pelecehan menjadi suatu yang dapat melakukan proses klasifikasi pelecehan secara efektif. Penelitian ini bertujuan untuk mengklasifikasi *tweet* pelecehan *online* pada media sosial X menggunakan *gated recurrent unit* dan *word embedding fastText*. Data yang digunakan pada penelitian ini berjumlah 7006 data dengan melakukan *crawling* menggunakan tool *tweet harvest* pada aplikasi X. Berdasarkan evaluasi menggunakan *confusion matrix* menghasilkan akurasi sebesar 0.91. Maka dapat diambil kesimpulan bahwa algoritma yang digunakan pada penelitian ini sudah cukup baik dalam mengklasifikasi pelecehan *online* pada media sosial X.

Kata kunci: *Pelecehan Online, Gated Recurrent Unit, FastText, Confusion Matrix*

CLASSIFICATION OF ONLINE HARASSMENT TWEET ON SOCIAL MEDIA X USING GATED RECURRENT UNIT

ABSTRACT

The development of communication and information technology is growing rapidly in various fields of human life, especially in the use of social media. One of the most widely used social media is X (twitter) which is often used for real-time information, including the latest news and provides a space for discussion on certain topics. However, the use of X not only has a positive impact but also a negative impact where many parties or users abuse the function of X itself resulting in an increase in crimes that arise, one of which is online harassment. Identifying online harassment is very important to protect users and prevent the widespread practice of online harassment and violence in Indonesia. However, this process will certainly take a long time if done manually. Therefore, an approach is needed that can process harassment tweet data into something that can effectively classify harassment. This research aims to classify online harassment tweets on social media X using gated recurrent units and word embedding fastText. The data used in this research amounted to 7006 data by crawling using the tweet harvest tool in the X application. Based on the evaluation using confusion matrix, the accuracy is 0.91. So it can be concluded that the algorithm used in this study is good enough in classifying online harassment on social media X.

Keywords: Online Harrassment, Gated Recurrent Unit, FastText, Confusion Matrix

DAFTAR ISI

PERSETUJUAN	Error! Bookmark not defined.
PERNYATAAN	i
UCAPAN TERIMA KASIH.....	iii
ABSTRAK.....	v
ABSTRACT.....	vi
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian.....	4
1.4. Batasan Masalah.....	4
1.5. Manfaat Penelitian.....	4
1.6. Metodologi Penelitian	4
BAB 2 LANDASAN TEORI.....	7
2.1. Pelecehan <i>Online</i>	7
2.2. Jenis-jenis Pelecehan <i>Online</i>	8
2.3. X (Twitter).....	11
2.4. Recurrent Neural Network (RNN)	11
2.5. Long Short Term Memory (LSTM)	12
2.6. Gated Recurrent Unit (GRU)	14
2.7. <i>Word Embedding</i>	15
2.8. Penelitian Terdahulu	16
2.9. Perbedaan Penelitian	18
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	20
3.1. Dataset	20
3.2. Arsitektur Umum.....	21
3.2.1 Input Data.....	22
3.2.2 <i>Pre-processing</i>	25
3.2.3 Word Embedding	29
3.2.4 Model Building	30
3.3. Perancangan Aplikasi Sistem	33

3.3.1	Gambaran Tampilan Halaman Depan.....	33
3.3.2	Gambaran Tampilan Training Data	34
3.3.3	Gambaran Tampilan Testing Data	35
3.4	Motode Evaluasi.....	37
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM.....		39
4.1.	Implementasi Sistem	39
4.1.1.	Spesifikasi <i>Hardware</i> dan <i>Software</i>	39
4.1.2.	Penerapan Tampilan Antarmuka.....	39
4.2.	Implementasi Model.....	43
4.2.1.	Model <i>Training</i>	43
4.2.2.	Hasil Pengujian Sistem	50
4.3.	Metode Evaluasi	53
BAB 5 KESIMPULAN DAN SARAN		62
5.1.	Kesimpulan.....	62
5.2.	Saran.....	62
DAFTAR PUSTAKA		63

DAFTAR TABEL

Tabel 2.1	Kata Kunci Pelecehan <i>Online</i>	10
Tabel 2.2	Penelitian Terdahulu	18
Tabel 3.1	Data hasil <i>Crawling</i> Pada X(Twitter)	21
Tabel 3.2	Dataset <i>Tweet</i> Pelecehan <i>Online</i>	24
Tabel 3.3	Pembagian Dataset	25
Tabel 3.4	Contoh Penerapan Proses <i>Cleaning Data</i>	26
Tabel 3.5	Contoh Penerapan Proses <i>Case Folding</i>	27
Tabel 3.6	Contoh Penerapan Proses <i>Punctual Removal</i>	27
Tabel 3.7	Contoh Penerapan Proses <i>Normalization</i>	28
Tabel 3.8	Contoh Penerapan <i>Stopword Removal</i>	29
Tabel 3.9	Contoh Penerapan <i>Stemming</i>	29
Tabel 3.10	Contoh Penerapan <i>Tokenization</i>	30
Tabel 3.11	Implementasi Evaluasi <i>Confussion Matrix</i>	38
Tabel 4.1	Hasil <i>Training</i>	47
Tabel 4.2	Hasil Pengujian Sistem	50
Tabel 4.3	Keterangan <i>Confussion Matrix</i> Pada Kelas <i>Racist</i>	54
Tabel 4.4	Keterangan <i>Confussion Matrix</i> Pada Kelas <i>Embarrass</i>	55
Tabel 4.5	Keterangan <i>Confussion Matrix</i> Pada Kelas <i>Physical Threatened</i>	56
Tabel 4.6	Keterangan <i>Confussion Matrix</i> Pada Kelas <i>Sexually Harassed</i>	57
Tabel 4.7	Keterangan <i>Confussion Matrix</i> Pada Kelas Netral	58
Tabel 4.8	Evaluasi	60

DAFTAR GAMBAR

Gambar 2.1	Contoh Tweet Rasis	8
Gambar 2.2.	Contoh Tweet Olokan dan Ejekan	9
Gambar 2.3	Contoh Tweet Ancaman Fisik.....	9
Gambar 2.4	Contoh Tweet Pelecehan Seksual	10
Gambar 2.5	Arsitektur Umum Recurrent Neural Network (Amidi, 2018).....	12
Gambar 2.6	Arsitektur Umum Long Short Term Memory (Wildan et al, 2018).....	12
Gambar 2.7	Arsitektur Umum Recurrent Neural Network (Arfianti, 2021)	15
Gambar 3.1.	Arsitektur Umum	23
Gambar 3.2	Embedding Matrix	31
Gambar 3.3	Gambaran Halaman Depan Aplikasi Web	35
Gambar 3.4	Gambaran Hampilan Training Data	36
Gambar 3.5	Gambaran Hampilan Testing Data.....	37
Gambar 3.6	Gambaran Hampilan Uji Coba Realtime	38
Gambar 4.1	Tampilan Halaman Depan/Beranda	41
Gambar 4.2	Tampilan Halaman Training Data.....	42
Gambar 4.3	Tampilan Hasil Training Data.....	42
Gambar 4.4	Tampilan Halaman Testing Data	43
Gambar 4.5	Tampilan Hasil Testing Data	43
Gambar 4.6	Tampilan Uji Coba Realtime	44
Gambar 4.7	Tampilan Grafik Akurasi dan Grafik Loss Kelas Racist	45
Gambar 4.8	Tampilan Grafik Akurasi dan Grafik Loss Kelas Embarrass.....	45
Gambar 4.9	Tampilan Grafik akurasi dan Grafik Loss Kelas Physical Threatened	46
Gambar 4.10	Tampilan Grafik Akurasi dan Grafik Loss Kelas Sexually Harassed..	46
Gambar 4.11	Tampilan Grafik Akurasi dan Grafik Loss Kelas Netral	47
Gambar 4.12	Confussion Matrix Racist.....	54
Gambar 4.13	Confussion Matrix Embarrass.....	55
Gambar 4.14	Confussion Matrix Physically Threatened	56
Gambar 4.15	Confussion Matrix Sexually Harassed	57
Gambar 4.16	Confussion Matrix Netral.....	58

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Dewasa ini kemajuan teknologi komunikasi dan informasi sangat berkembang pesat di berbagai aspek kehidupan manusia, terutama pada penggunaan media sosial yang menjadi gaya baru bagi masyarakat modern yang dapat dilakukan dengan mudah tanpa batas ruang dan waktu. Menurut hasil riset *We Are Social Hootsuite* mengungkapkan bahwa pengguna aktif media sosial di Indonesia meningkat 12,35% atau sebesar 191 juta jiwa pada Januari 2022. Salah satu media sosial yang biasa digunakan adalah X. Pada bulan Juli 2020, didapati bahwa jumlah pengguna X di Indonesia terbanyak di posisi kedelapan di atas Thailand, Filipina dan Meksiko dimana tercatat sebanyak 11,2 juta pengguna aktif (Statista, 2020).

X sering digunakan sebagai alat untuk berbagi informasi secara real-time, termasuk berita terbaru, pembaruan dari organisasi, selebriti, dan politisi, serta diskusi publik tentang topik tertentu. Fitur-fitur yang dimiliki oleh X seperti *tweet*, *reply*, *quote tweet*, *like* dan sebagainya memberi kebebasan setiap pengguna untuk mengungkapkan apapun yang mereka mau. Namun hal ini memberikan dampak negatif sehingga semakin banyak kejahatan yang ditemui di dunia maya atau media sosial, salah satunya adalah Pelecehan Online (*Online Harrassment*).

Pelecehan Online (*Online Harrasment*) merupakan tindakan yang dilakukan secara daring atau *online* dengan tujuan merendahkan, mengintimidasi, mengganggu seseorang atau kelompok tertentu dan dapat terjadi diberbagai *platform* media sosial. Mengalami pelecehan online pada umumnya mengarah pada konsekuensi serius bagi kesehatan mental, trauma psikologis dan emosional pada korban, serta dapat berdampak pada kehidupan pribadi dan profesional mereka contohnya seperti rendahnya harga diri, kecemasan dan kesepian (Celuch et al., 2023). Beberapa faktor terjadinya pelecehan *online* antara lain yaitu pandangan politik, ras atau etnis, penampilan fisik, gender, agama dan orientasi seksual (Duggan, 2017).

Menurut laporan dari survey yang dilakukan oleh Pew Research Center 2017 berikut jenis-jenis pelecehan *online* yang sering terjadi dan diterima yaitu kata-kata yang rasis (*racist*) sebesar 27%, olokan atau ejekan (*embarrass*) sebesar 20%, kicauan yang mengancam fisik (*physically threatened*) sebesar 10%, selanjutnya kicauan pelecehan yang selalu berulang (*harassed for a sustained period*) sebesar 7%, lalu kata-kata yang mengandung pelecehan seksual (*sexually harassed*) sebesar 7%, dan penguntit (*stalking*) sebesar 6%. Data tersebut berdasarkan salah satu kasus pelecehan *online* yang umumnya terjadi di media sosial, dan juga komunikasi pada game *online*, situs web bagian kolom komentar, dan email.

Polling Indonesia yang bekerja sama dengan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) telah melakukan riset pada 5.900 sampel dan hasil dari riset tersebut mengatakan ada sekitar 49 persen netizen pernah menjadi sasaran pelecehan atau kekerasan secara *online* di media sosial (Pratomo, 2019). Berdasarkan data KOMNAS Perempuan pada tahun 2022 merilis laporan tentang kekerasan atau pelecehan berbasis *online* yang dimana kasus tertinggi adalah siber atau *online* sebanyak 869 kasus. KOMNAS Perempuan menyatakan bahwa terdapat 383 kasus siber di ranah publik yang dilakukan oleh “teman sosial media”. Hingga per maret 2023 angka kasus siber yang dilaporkan dari lembaga layanan secara keseluruhan mengalami peningkatan sebanyak 112 kasus (Komnas perempuan, 2023).

Berdasarkan angka kasus yang termasuk tinggi pada kasus pelecehan dan kekerasan yang dilakukan secara *online*/siber ini sudah seharusnya menjadi hal yang serius yang harus diperhatikan semua pihak. Akan tetapi sampai saat ini, pemantauan terhadap kasus ini masih dilakukan secara manual oleh lembaga dan pihak yang berwenang sehingga penanganan kasus pelecehan *online* di Indonesia masih tergolong lambat. Oleh karena itu, permasalahan ini membutuhkan suatu pendekatan yang dapat mengotomatiskan proses identifikasi pelecehan *online* yang terjadi pada media sosial X secara efektif berdasarkan klasifikasinya.

Penelitian mengenai pengklasifikasian *tweet* pelecehan *online* pada media sosial twitter sebelumnya sudah pernah dilakukan dengan metode *Naïve Bayes* dan hasil yang didapatkan yaitu tingkat akurasi sebesar 85% dimana menggunakan 1500 dataset dengan 1200 data latih dan 300 data uji. Pada penelitian ini dikatakan belum memiliki simulasi yang sempurna, sehingga dibutuhkannya penyempurnaan sesuai dengan

kebutuhan (Sandes, 2018). Penelitian selanjutnya dilakukan oleh Kurnia et al., (2022) berjudul *Klasifikasi Tweet Cyberbullying Dengan Menggunakan Algoritma Random Forest*. Pada penelitian ini melakukan proses latih data dengan dua model yang berbeda berdasarkan perbedaan kedalaman pohon, daun, bobot kelas dan fitur maksimum. Hasil akurasi antara 2 model berbeda tersebut sebesar 93,3% dan 93,2%.

Dalam penelitian ini, penulis mengajukan penelitian menggunakan metode *Gated Recurrent Unit*. *Gated Recurrent Unit* merupakan salah satu dari dua varian *Recurrent Neural Network* dimana memiliki arsitektur yang saat ini umum digunakan dan variasi lainnya yaitu *Long Short Term Memory*. RNN juga dikenal menjadi salah satu algoritma yang cukup baik dalam mengatasi permasalahan *Natural Language Processing* (Tarwani et al., 2017). Pada penelitian terdahulu yang dilakukan oleh Biswas et al., (2015) berjudul *Sentimen Analisis Terhadap Ulasan Film* dengan menggunakan 50.000 dataset ulasan film IMDB, ditemukan bahwa *Gated Recurrent Unit* menghasilkan tingkat akurasi tertinggi yaitu sebesar 97%, disusul oleh arsitektur *Long Short Term Memory* dengan tingkat akurasi 95%. Penelitian selanjutnya terkait Identifikasi Judul Clickbait Pada Berita Bahasa Indonesia (Hutabarat, 2021) menggunakan GRU dihasilkan nilai akurasi yang cukup baik yaitu sebesar 0.88 dengan loss sebesar 0.3.

Berdasarkan latar belakang diatas, penulis mengajukan sebuah pendekatan untuk memperoleh metode yang paling baik dalam pengklasifikasian. Oleh karena itu, penulis mengajukan penelitian dengan judul **“KLASIFIKASI TWEET PELECEHAN ONLINE PADA MEDIA SOSIAL X MENGGUNAKAN GATED RECURRENT UNIT”**.

1.2. Rumusan Masalah

Pelecehan *online* pada media sosial memiliki dampak yang sangat merugikan bagi pengguna dan melakukan identifikasi pelecehan *online* sangat penting untuk dilakukan guna melindungi pengguna *online* dan mencegah semakin meluasnya praktik-praktik pelecehan dan kekerasan secara *online* di Indonesia. Tetapi, proses ini tentunya akan memakan waktu yang cukup lama jika melakukannya secara manual. Oleh sebab itu, dibutuhkan suatu sistem yang dapat mengotomatiskan proses identifikasi pelecehan *online* yang terjadi pada media sosial secara efektif.

1.3. Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengklasifikasi tweet pelecehan *online* pada media sosial twitter menggunakan metode *Gated Recurrent Unit* (GRU).

1.4. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Data yang digunakan berupa *text* di dalam *file* dokumen berekstensi *.csv*
2. Data merupakan *tweet* berbahasa indonesia hasil *crawling* dari media sosial X dan dilabeli secara manual
3. Penelitian dilakukan dengan memperhatikan kata, tidak terhadap angka, emoji dan gambar
4. Output yang dihasilkan merupakan sistem berbasis web yang dapat mengklasifikasi data tweet kedalam kelas Rasis, Ejekan & Olokan, Ancaman Fisik, Pelecehan Seksual

1.5. Manfaat Penelitian

Berikut manfaat penelitian yang dilakukan :

1. Mempelajari dan memahami bagaimana kinerja dari algoritma *Gated Recurrent Unit* dalam pengklasifikasian *tweet* pelecehan *online* pada media sosial twitter dan mengukur seberapa besar tingkat akurasi yang dihasilkan
2. Sebagai sistem pendukung dalam mendeteksi pelecehan *online* secara efektif, sehingga dapat mengurangi potensi terjadinya pelecehan *online* di media sosial, terkhusus twitter.

1.6. Metodologi Penelitian

Berikut tahapan yang akan dilakukan dalam penelitian ini :

1. Studi Literatur

Studi Literatur merupakan tahapan dimana penulis melakukan pengumpulan data dari berbagai sumber referensi. Referensi berasal dari buku, skripsi, jurnal, artikel dan sumber informasi lainnya melalui media *online* (internet) yang berhubungan dengan klasifikasi, pelecehan *online* (*online harassment*) pada media sosial twitter, algoritma *Gated Recurrent Unit*, serta *text processing*.

2. Analisis Permasalahan

Pada tahap ini, dilakukan analisis permasalahan berdasarkan data-data yang digunakan dimana memiliki tujuan untuk memahami konsep dari metode *Gated Recurrent Unit* (GRU) beserta faktornya sehingga dapat menyelesaikan permasalahan tentang pelecehan *online* pada media sosial twitter.

3. Perancangan

Tahap selanjutnya yaitu perancangan, dimana penulis melakukan perancangan pada arsitektur umum, melakukan pengumpulan data serta melakukan perancangan tampilan antarmuka pada *system*.

4. Implementasi

Tahap implementasi merupakan langkah selanjutnya setelah dilakukannya tahap analisis. Pada tahap ini menggunakan Bahasa pemrograman python untuk mengimplementasikan metode *Gated Recurrent Unit* (GRU).

5. Pengujian

Setelah tahap implementasi, selanjutnya akan dilakukan pengujian terhadap performa berdasarkan hasil dari implementasi metode *Gated Recurrent Unit* dalam mengklasifikasi pelecehan *online* sesuai dengan tujuan penelitian.

6. Penyusunan Laporan

Pada tahap ini, dilakukan penyusunan laporan dengan melengkapi dokumentasi laporan penelitian yang menjadi hasil akhir berupa skripsi.

1.7. Sistematika Penulisan

Dalam penelitian ini, sistematika penulisan dibagi menjadi 5 (lima) bagian utama, yaitu sebagai berikut.

Bab 1: Pendahuluan

Dalam bab pendahuluan, berisikan penjelasan mengenai latar belakang dilakukannya penelitian yang berjudul “Klasifikasi Pelecehan *Online* Pada Media Sosial X Menggunakan *Gated Recurrent Unit*”, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab 2: Landasan Teori

Pada bab landasan teori, berisikan penjelasan mengenai dasar-dasar konsep dan teori yang berkaitan dengan penelitian ini untuk dipahami dan dikaji dalam proses

penyelesaian masalah. Bab ini menjelaskan mengenai pelecehan *online* dan penjelasan terkait metode yang digunakan yaitu *Gated Recurrent Unit*. Selain itu, juga diuraikan penelitian-penelitian terkait yang sebelumnya sudah dilakukan dan akan dijadikan sebagai referensi dalam penelitian ini.

Bab 3: Analisis dan Perancangan Sistem

Pada bab analisis dan perancangan sistem, berikan penjelasan mengenai analisis permasalahan penelitian dan perancangan metode *Gated Recurrent Unit* dan sistem yang akan di rancang.

Bab 4: Implementasi dan Pengujian Sistem

Pada bab implementasi dan pengujian sistem berisikan penjelasan mengenai implementasi atau penerapan sistem yang sudah dirancang sebelumnya pada bab 3 serta memaparkan hasil dari pengujian sistem.

Bab 5: Kesimpulan dan Saran

Pada terakhir yaitu berisikan kesimpulan dari seluruh pengerjaan penelitian yang sudah diuraikan serta saran yang diajukan oleh penulis untuk pengembangan penelitian selanjutnya.

BAB 2

LANDASAN TEORI

2.1. Pelecehan *Online*

Pelecehan *Online* (*Online Harassment*) dikenal sebagai “*cyber harassment*” merupakan tindakan yang dilakukan secara daring atau *online* dengan tujuan merendahkan, mengintimidasi, mengganggu seseorang atau kelompok tertentu dan sering terjadi diberbagai platform media sosial. Tindakan ini dilakukan secara berulang kali untuk menargetkan orang atau kelompok tertentu dengan langsung menghubungi mereka, atau secara tidak langsung menyebarkan informasi pribadi mereka, sehingga dapat menyebabkan trauma psikologis dan emosional pada korban, serta berdampak pada kehidupan pribadi dan profesional mereka.

Pada survey yang sudah dilakukan oleh *Pew Research Center Survey of U.S. Citizens* (2017) menjelaskan beberapa hal yang mencakup pelecehan *online*:

1. Mengirimkan *e-mail* kepada korban secara berlebihan yang tidak diminta atau diinginkan sehingga dapat mengancam penerima secara tidak langsung.
2. Melakukan sabotase elektronik dengan mengirim virus melalui *email*.
3. Mengirimkan komentar *online* yang meremehkan suatu hal.
4. Pesan negative yang dikirim langsung kepada korban.
5. Mencoba menyamar sebagai korban dan mengirimkan pesan yang provokatif, kontroversial, atau menggoda yang menyebabkan orang lain bereaksi buruk terhadap korban.
6. Mengirimkan informasi demografis korban dan mengirim gambar yang berorientasi seksual atau situs pornografi.
7. Mencari dan mengumpulkan berbagai informasi korban yang mungkin telah diposkan oleh situs berita dengan maksud untuk mencari informasi pribadi dan kemudian menggunakan informasi ini untuk melecehkan, mengancam, dan mengintimidasi korban baik secara *online* maupun di dunia nyata

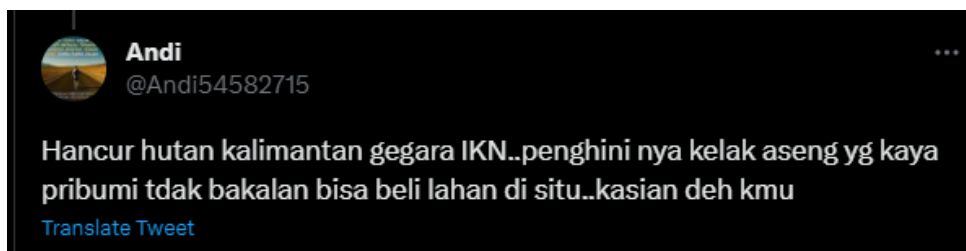
2.2. Jenis-jenis Pelecehan *Online*

Menurut survei yang dilakukan oleh *Pew Research Center* 2017, berikut jenis-jenis pelecehan *online* yang terjadi di media sosial :

1. Rasis (*Racist*)

Rasisme adalah suatu sistem kepercayaan atau pandangan yang didasarkan pada keyakinan bahwa satu ras atau kelompok ras memiliki keunggulan atau nilai yang lebih tinggi dibandingkan dengan ras atau kelompok lainnya. Rasisme melibatkan deskriminasi, penindasan, atau perlakuan tidak adil terhadap individu atau kelompok berdasarkan ras atau latar belakang etnis mereka.

Tindakan rasisme seringkali dilakukan di lingkungan sekitar baik itu di “dunia nyata” ataupun “dunia maya” seperti media sosial. Contoh tindakan yang secara disengaja atau tidak disengaja digunakan oleh masyarakat untuk melakukan ejekan/umpatan terhadap sesamanya, seperti: “Dasar cina lo, pendatang”, “Dasar jawa lo, orang pindahan”, “Orang timur item-item/keling ya” dan lain sebagainya (Prayoga, 2020). Ada pula yang melakukan tindakan tersebut secara diam-diam atau sembunyi dengan cara menggunakan “bahasa simbol” tertentu, seperti kata “Aseng dan Pribumi” yang merupakan sebutan sarkas sekaligus ekspresi dari sentimen anti Cina dan pribumi (Sumanto Al Qurtuby, 2017).



Gambar 2.1 Contoh Tweet rasis

2. Olokan atau Ejekan (*Embarrass*)

Olokan atau Ejekan merujuk pada tindakan atau ucapan yang bertujuan untuk mempermalukan atau merendahkan seseorang. Tindakan ini sering terjadi tidak hanya dilakukan secara langsung tetapi melalui sosial media atau komunitas *online*. Contoh tindakan yang dilakukan berupa ejekan, penghinaan, komentar merendahkan dan penyebaran meme atau gambar yang menghina seseorang.

Dikutip dari postingan tweet seseorang di media sosial twitter “Pantes ga ada otak, otaknya tertinggal di tong sampah”. (Twitter @Areajulid diunggah pada 02

Oktober 2022). Dari kalimat tersebut terdapat kata “gaa da otak” yang dapat diartikan sebagai seseorang yang tidak dapat berpikir, dan lain sebagainya.

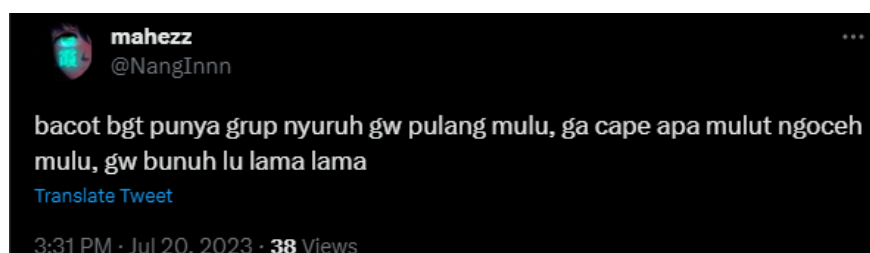


Gambar 2.2. Contoh Tweet olokan dan ejekan

3. Ancaman Fisik (*Physically Threatened*)

Ancaman Fisik yang dilakukan pada media sosial merujuk pada tindakan mengancam atau mengintimidasi seseorang secara fisik yang dapat berupa ucapan, pesan atau konten yang menyiratkan niat atau kemungkinan melakukan kekerasan fisik terhadap seseorang. Tindakan ini juga memiliki dampak yang fatal pada kesejahteraan dan kenyamanan setiap individu yang menjadi korban yang menciptakan rasa takut, kecemasan dan merusak suasana hati bahkan dapat meningkatkan resiko terjadinya tindak kekerasan dalam kehidupan nyata.

Contoh ancaman fisik yang terjadi pada media sosial seperti pada kalimat “*mending usir saja dari jogja*” dan “*siap-siap nyawamu terancam di jogja*”. Contoh tersebut biasanya merupakan tanggapan atas peristiwa yang sering juga dijumpai pada dunia nyata. “usir, tendang, labrak” merupakan kata-kata yang digunakan untuk mengancam seseorang secara fisik (Apriliawati, 2015).



Gambar 2.3 Contoh Tweet ancaman fisik

4. Pelecehan Seksual (*Sexually Harrased*)

Pelecehan Seksual merupakan suatu tindakan yang menggunakan komentar, pesan atau konten yang bersifat seksual kepada seseorang tanpa izin atau yang tidak dikehendaki dan berakibat pada dampak psikologis yang signifikan pada korban, termasuk kecemasan, depresi, trauma dan penurunan harga diri.



Gambar 2.4 Contoh Tweet pelecehan seksual

Mengutip dari (Sendes, 2018) menguraikan kata kunci yang biasa digunakan yang didapatkan melalui berbagai referensi dan media seperti berita terkait kategori pelecehan *online*. Berikut klasifikasi kata-kata pelecehan online yang dapat dilihat pada Tabel 2.1

Tabel 2.1 Kata Kunci Pelecehan *Online*

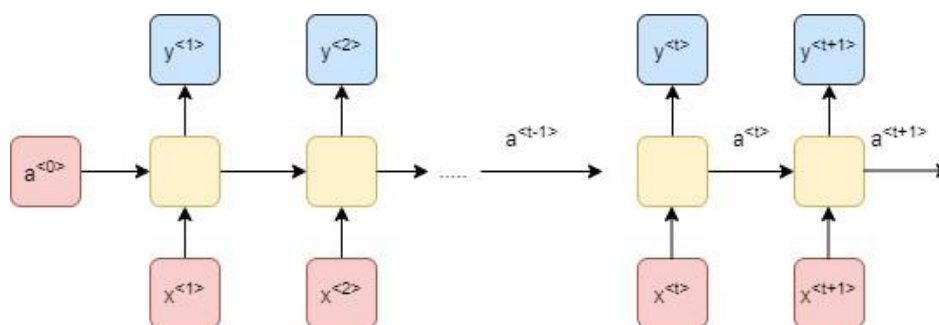
Kelas	Kata Kunci Pelecehan <i>Online</i>
Kata pelecehan kategori Rasis	<ul style="list-style-type: none"> - Aseng - Pribumi - Cina - Kafir - Kadrin
Kata pelecehan kategori Olokan atau Ejekan (<i>Embarrass</i>)	<ul style="list-style-type: none"> - Goblok - Tolol - Bego - Babi - Dasar ndeso - Otak - Anjing - Taik - Penjilat
Kata pelecehan kategori Ancaman Fisik (<i>Physically Threatened</i>)	<ul style="list-style-type: none"> - Labrak - Jambak - Bunuh - Gampar - Bacok - Tonjok - Tendang - Cekek - Santet - Hajar
Kata pelecehan kategori Pelecehan Seksual (<i>Sexually Harrased</i>)	<ul style="list-style-type: none"> - Pelacur - Montok - Pantat - Homo - Lesbi - Bugil - Pelakor

2.3.X (Twitter)

X(Twitter) merupakan *platform* media sosial yang populer yang memungkinkan pengguna berbagi pesan yang disebut dengan “*tweet*” dengan panjang maksimal 280 karakter. Twitter diluncurkan pada tahun 2006 dan telah menjadi salah satu platform paling banyak digunakan di dunia untuk berkomunikasi secara *real-time*, berbagi informasi, dan terlibat dalam percakapan global. Pengguna twitter dapat mengikuti akun-akun lain untuk melihat tweet yang mereka bagikan di aliran informasi mereka. Mereka dapat membalas *tweet*, menyukai, dan membagikannya kembali (*retweet*) kepada pengikut mereka. Dengan fitur-fitur ini, Twitter memfasilitasi diskusi, interaksi sosial, dan berbagi informasi secara cepat dan mudah (about.twitter.com).

2.4. Recurrent Neural Network (RNN)

Recurrent Neural Network merupakan jenis arsitektur jaringan saraf yang dirancang khusus dengan tujuan untuk melakukan proses data secara berurutan atau sekuensial. RNN juga dikenal sebagai jaringan saraf berulang, karena nilai neuron yang sebelumnya terdapat di *hidden layer* akan digunakan kembali sebagai data input. Selanjutnya penggunaan neuron pada *hidden layer* tersebut akan disimpan kembali ke dalam *context layer*, dan nilai tersebut akan terus di *update* sampai kondisi RNN terpenuhi (). *Recurrent Neural Network* merupakan metode yang kompleks dan dinamis disebabkan hasil tidak hanya dipengaruhi oleh *input*, tetapi juga dari *output* sebelumnya (). Berikut arsitektur umum RNN :



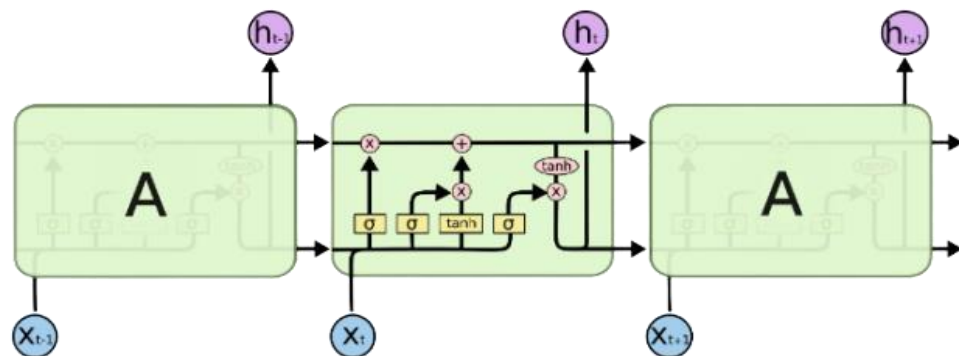
Gambar 2.5 Arsitektur Umum *Recurrent Neural Network* (Amidi, 2018)

Pada gambar 2.5, setiap waktu (t), didapatkan $x^{<t>}$ yang dimana sebagai input dengan fungsi aktivitas dan disimbolkan dengan $a^{<t>}$ serta memiliki output yang disimbolkan dengan $y^{<t>}$. Dalam RNN, sering ditemui terjadinya permasalahan seperti *vanishing* dan *exploding*, dikarenakan RNN memiliki banyaknya jumlah lapisan sehingga

membuat nilai gradient semakin kecil saat proses *training* yang dimana nilainya mendekati nol atau menghilang. Sedangkan *exploding gradient* dapat terjadi ketika nilai gradiennya memiliki jumlah yang besar. Dengan demikian, untuk menyelesaikan permasalahan tersebut dapat menggunakan algoritma *Long Short Term Memory* dan *Gated Recurrent Unit*.

2.5. Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) merupakan salah satu bentuk modifikasi dari *Recurrent Neural Network* (RNN) yang memiliki kemampuan untuk memprediksi secara akurat terhadap suatu variabel (Owen, et al, 2022). LSTM juga dapat mengatasi masalah *vanishing gradient* yang sering terjadi pada RNN. LSTM menggunakan unit memori khusus disebut *cell state* dan *gates unit* yang mampu mempelajari dan mengingat informasi dalam jangka panjang dengan mengontrol aliran informasi didalamnya. Berikut arsitektur umum LSTM :



Gambar 2.6 Arsitektur Umum *Long Short Term Memory* (Wildan et al, 2018)

Pada gambar 3 diatas menunjukkan alur kerja dari *memory cells* ketika LSTM bekerja pada setiap *neurons*. Ada tiga jenis *gate*, diantaranya *forget gate*, *input gate*, dan *output gate*. *Forget gate* berfungsi untuk menghapus informasi yang tidak dibutuhkan lagi pada *cell*. Pada *forget gate* menggunakan fungsi sigmoid yang menghasilkan nilai antara 0 sampai 1, sehingga menunjukkan seberapa pentingnya informasi yang akan di hapus. Berikut rumus yang digunakan:

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

Keterangan:

f_t = *forget gate*

σ = fungsi aktivasi sigmoid

W_f = *weight*

h_{t-1} = *hidden state* sebelumnya

x_t = *input*

b_f = *bias*

Input Gate berfungsi untuk mengendalikan informasi yang akan dikelola dan ditambahkan ke *cell state*. Pada *input gate* menggunakan fungsi sigmoid untuk mengontrol seberapa banyak informasi akan ditambahkan. Fungsi tangen hiperbolik(tanh) yang menghasilkan vector kandidat nilai yang ditambahkan. Berikut rumus yang digunakan :

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

$$\check{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \check{C}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

Keterangan :

I_t = *input gate*

\check{C}_t = kandidat *cell state*

C_t = *cell state*

$ht-1$ = *hidden state* sebelum

\tanh = fungsi aktivasi tanh

W_i, W_c = *weight* pada *input gate* dan *cell state*

b_i, b_c = *bias input gate* dan *cell state*

Output Gate berfungsi untuk mengendalikan informasi yang berada dalam *cell state* yang akan dikeluarkan kedalam lapisan berikutnya atau sebagai output dari LSTM. Berikut rumus yang digunakan :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

Keterangan :

ot = *output gate*

ht = *hidden state*

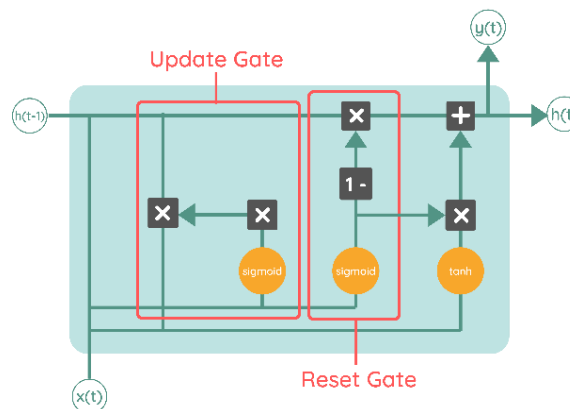
W_o = *weight output gate*

Bo = nilai bias

2.6. Gated Recurrent Unit (GRU)

Gated Recurrent Unit merupakan jenis arsitektur turunan dari RNN yang serupa dengan LSTM tetapi GRU menyederhanakan struktur LSTM dengan menggabungkan beberapa *gate* menjadi satu *control gate*. GRU diperkenalkan pertama kali oleh Chung et al (2014). GRU memiliki tujuan utama yaitu untuk membuat setiap *recurrent unit* dapat menangkap *dependencies* dengan cara dapat beradaptasi terhadap skala waktu yang beragam (Rafika, 2020).

Di dalam GRU, memiliki 2 *gate* sebagai komponen pengatur alur informasi, antara lain *reset gate* dan *update gate*. *Reset Gate* merupakan *gate* yang memberi keputusan tentang bagaimana untuk menggabungkan input baru dengan informasi masa lalu, sedangkan *update gate* merupakan *gate* yang akan memberi keputusan seberapa banyak informasi masa lalu yang akan tetap disimpan. Pada *reset* dan *update gate* menggunakan fungsi sigmoid untuk menghasilkan vector yang memiliki nilai antara 0 dan 1. Berikut arsitektur *reset gate* dan *update gate* :



Gambar 2.7 Arsitektur Umum *Gated Recurrent Unit* (Arfianti, 2021)

Langkah awal yang dilakukan dalam membangun model GRU dengan menghitung *update gate* menggunakan persamaan berikut ini :

$$Z_t = \sigma(W_{xz} X_t + W_{hz} H_{t-1} + b_z) \quad (8)$$

Langkah selanjutnya menghitung *reset gate*, berikut persamaan *reset gate* :

$$R_t = \sigma(W_{xr} X_t + W_{hr} H_{t-1} + b_r) \quad (9)$$

Keterangan :

R_t = *reset gate*

Z_t = *update gate*

σ = aktivasi fungsi sigmoid

H_{t-1} = output pada blok GRU sebelumnya (dalam timestamp $t - 1$)

X_t = input pada timestamp sekarang

W_{xr} , W_{hr} , W_{xz} , W_{xz} = parameter weight

b_r , b_z = parameter bias

Kemudian menghitung *candidate hidden state* yang akan menggunakan *reset gate* untuk menyimpan informasi yang relevan dari masa lalu dan diakhiri dengan menyimpan informasi yang akan diteruskan ke blok selanjutnya yang disebut dengan *hidden state* dan memerlukan *update date*. Berikut persamaan *candidate hidden state* dan *hidden state* :

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \cdot H_{t-1}) W_{hh} + b_h) \quad (10)$$

$$H_t = Z_t \cdot H_{t-1} + (1 - Z_t) \cdot \tilde{H}_t \quad (11)$$

Keterangan:

\tilde{H}_t = *candidate hidden state*

H_t = *hidden state*

X_t = input pada timestamp sekarang

R_t = *reset gate*

W_{xh} , W_{hh} = parameter weight

b_h = parameter bias

Z_t = *update gate*

2.7. Word Embedding

Word Embedding merupakan suatu metode yang digunakan untuk melakukan konversi pada serangkaian kata yang berupa karakter *alphanumeric* ke dalam bentuk vector atau array yang berisi angka atau bilangan didalam ruang multidimensional (Silitonga, 2021). Representasi ini memungkinkan mesin untuk memahami dan memanipulasi makna kata-kata berdasarkan hubungan dan konteksnya. Representasi ini dapat dilakukan dan diperoleh melalui teknik-teknik seperti Word2Vec, Glove, dan FastText. Dalam penelitian ini teknik yang digunakan dalam tahap *word embedding* yaitu FastText, yang dimana pertama kali diperkenalkan oleh facebook. FastText sendiri mampu menangkap informasi yang lebih kaya dan umum dalam kata-kata yang memiliki morfologi yang kompleks atau yang jarang terlihat dalam data pelatihan.

2.8 Penelitian Terdahulu

Penelitian mengenai pengklasifikasian tweet pelecehan *online* pada media sosial twitter sebelumnya sudah pernah dilakukan oleh Sandes (2018) dengan metode *Naïve Bayes*. Data yang digunakan merupakan 1500 data tweet yang berasal dari twitter berbahasa indonesia yang dibagi menjadi 1200 data latih dan 300 data uji. Hasil dari penelitian ini mendapatkan nilai akurasi sebesar 85%.

Penelitian selanjutnya mengenai klasifikasi tweet *cyberbullying* dengan menggunakan metode *random forest* oleh Kurnia, et al (2022), dimana pada penelitian ini melakukan proses latih data dengan dua model yang berbeda berdasarkan perbedaan kedalaman pohon, daun, bobot kelas dan fitur maksimum. Pengujian ini juga melakukan perbandingan *confusion matrix* dari kedua model tersebut dan mengukur nilai *true positive*-nya. Nilai akurasi menunjukkan hasil sebesar 93,3% dengan kedalaman maksimum 350 menggunakan 100 pohon, bobot kelas tidak diberikan, jumlah daun minimum sebanyak 1, dan fitur maksimum bernilai akar 2. Sedangkan dengan kedalaman maksimum 200, menggunakan 250 pohon, bobot kelasnya seimbang, banyak daun minimum 3, dan fitur maksimal bernilai log2 menghasilkan akurasi sebesar 93,2%.

Triyana, et al (2022) melakukan penelitian tentang Deteksi *Cyberbullying* Pada *Tweet* Berbahasa Inggris Dengan Metode *Support Vector Machine*. Penelitian ini menggunakan *tweet* berbahasa inggris karena bahasa ini masih menjadi prioritas utama ASEAN, dan Amerika Serikat menjadi negara dengan pengguna Twitter terbanyak nomor 1 didunia. Pada pengujian ini menggunakan SVM yang diikuti dengan teknik *preprocessing* serta pembobotan kata dan mendapatkan hasil akurasi sebesar 90,59%, *recall* 92,74%, *precision* 95,42% dan *f-measure* 94,06%.

Pada penelitian yang dilakukan oleh Biswan et al (2015) yang berjudul sentimen analisis terhadap ulasan film dengan menggunakan 50.000 dataset ulasan film IMDB, ditemukan bahwa Gated Recurrent Unit menghasilkan tingkat akurasi tertinggi yaitu sebesar 97%, disusul oleh arsitektur Long Short Term Memory dengan tingkat akurasi 95%.

Hutabarat (2021) melakukan penelitian tentang identifikasi judul *clickbait* pada berita Bahasa Indonesia menggunakan *Gated Recurrent Unit* dihasilkan nilai akurasi yang cukup baik yaitu sebesar 0.88 dengan loss sebesar 0.3. Pada penelitian ini

menggunakan kumpulan data berjumlah 8.613 judul berita berasal dari 12 media *online* local bahasa indonesia yang sudah diberikan label *clickbait* dan *non-clickbait*.

Tabel 2.2 Penelitian Terdahulu

No.	Peneliti	Judul	Metode	Keterangan
1.	Sandes (2018)	Klasifikasi Pelecehan <i>Online</i> Pada Media Sosial Twitter Menggunakan Metode <i>Naïve Bayes</i>	<i>Naïve Bayes</i>	Pada penelitian ini melakukan keseluruhan tahapan pengujian dan memperoleh hasil akurasi tertinggi dengan text processing tanpa stemming menghasilkan nilai akurasi sebesar 85%.
2.	Kurnia. <i>et al</i> (2022)	Klasifikasi Tweet <i>Cyberbullying</i> dengan Menggunakan Algoritma Random Forest	<i>Random Forest</i>	Pada penelitian ini melakukan proses latih data dengan dua model yang berbeda berdasarkan perbedaan kedalaman pohon, daun, bobot kelas dan fitur maksimum. Hasil akurasi antara 2 model berbeda tersebut sebesar 93,3% dan 93,2%.
3.	Triyana <i>et al</i> (2022)	Deteksi <i>Cyberbullying</i> Pada Tweet Berbahasa Inggris Dengan Metode <i>Support Vector Machine</i>	<i>Support Vector Machine</i> (SVM)	Pada penelitian ini menggunakan dataset yang tersedia di Github dan Kaggle sebanyak 24.783 data tweet dan melakukan pengujian dengan metode SVM . Hasil akurasi yang didapatkan sebesar

Tabel 2.2 Penelitian Terdahulu (Lanjutan)

No.	Peneliti	Judul	Metode	Keterangan
				90,59% dengan nilai <i>recall</i> 92,74% , <i>precision</i> 95,42% dan <i>f-measure</i> 94,06%. Penelitian ini dibagi menjadi 2 kelas klasifikasi yaitu Positif dan Negatif.
4.	Biswan et al (2015)	Sentiment Analysis with <i>Gated Recurrent Unit</i> .	<i>Gated Recurrent Unit</i> (GRU)	Penelitian ini melakukan sentiment analisi terhadap ulasan film IMBD menggunakan metode GRU dengan dataset sebanyak 50.000 data dan menghasilkan akurasi sebesar 97% .
5.	Hutabarat (2021)	Identifikasi Judul <i>Clickbait</i> Pada Berita Bahasa Indonesia Menggunakan <i>Gated Recurrent Unit</i>	<i>Gated Recurrent Unit</i> (GRU)	Pada penelitian ini menggunakan 8.613 judul berita yang diambil dari portal berita sebanyak 12 media <i>online</i> local bahasa indonesia, dan menunjukkan hasil akurasi sebesar 0.88 dengan loss sebesar 0.3.

2.9. Perbedaan Penelitian

Ada beberapa hal yang membedakan penelitian ini dengan semua penelitian yang sudah dijelaskan sebelumnya. Hal pembeda pertama adalah metode yang digunakan didalam penelitian ini, ditemukan bahwa belum ada penelitian yang melakukan klasifikasi

pelecehan online pada media sosial X menggunakan *Gated Recurrent Unit* (GRU) dan melakukan pengambilan data menggunakan *tweet harvest*. Selain itu, penulis juga menggunakan *Word Embedding fastText* sebagai pembobotan kata, dimana penelitian terdahulu lebih banyak menggunakan TF-IDF dan Word2Vec. Penulis juga melakukan tujuh tahapan dalam *preprocessing* yaitu *cleaning*, *case folding*, *punctual removal*, *normalization*, *stopword removal*, *stemming* dan *tokenization*.

BAB 3

ANALISIS PERANCANGAN

3.1 Dataset

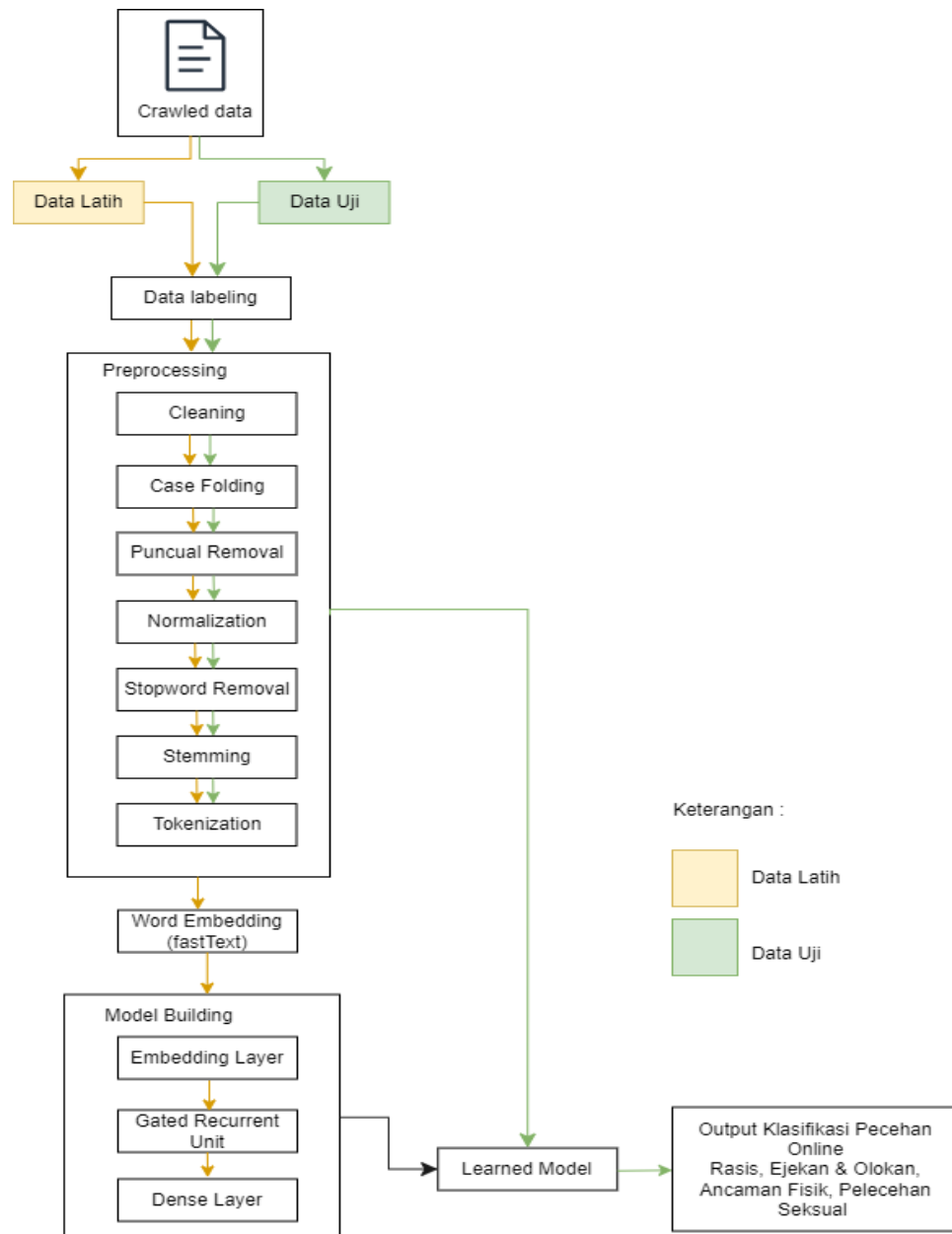
Pada penelitian ini penulis melakukan pengumpulan dataset yang berisikan *tweet* bahasa indonesia, dimana inputnya berupa teks dengan panjang maksimal 280 karakter yang di dapatkan melalui aplikasi X(Twitter) dengan melakukan *crawling data* menggunakan tool *Tweet Harvest*. Data dari hasil crawling berjumlah 7006 *tweet* yang disimpan ke dalam file berformat *.csv*. Data tersebut akan dibagi menjadi data training dan data testing dengan pembagian yaitu 80% dari keseluruhan dataset sebanyak 5604 data dijadikan sebagai data training dan 20% dari keseluruhan dataset sebanyak 1402 data dijadikan sebagai data testing. Berikut hasil *crawling data* dari aplikasi X yang dapat dilihat pada Tabel 3.1

Tabel 3.1 Data hasil *crawling* pada X(Twitter)

No	<i>Tweet</i>
1	@siregar_elang @DokterTifa Mau Susah kerjaan... Mau negara Banyak hutang... Mau pejabat banyak yg Koruptor... Mau Bandara & Pelabuhan Mangkrak... Mau sesama anak bangsa tengkar mulu... Mau Minyak goreng Langka... Mau Tol & BBM naik.. KAGAK PEDULI.. Yang penting Presidennya punya muka ndeso gitu..? Dasar..
2	@basuki_btp Tuan @basuki_btp. Anda membuat bbrpa ekor bangsat negri ini sakit hati. Anda adalah cina, aseng, asing, penista agama dan dianggap minoritas di Indonesia. Mengapa ingin membuat Pribumi" sejahtera..? Kasihan yg ingin memperbudak saudaranya sndiri, tuan."
3	@ardillasantix Aku maunya bugil langsung berduaiaan sm kamu
4	@OUR5UNSHINE @JAEBUM5UN EH ELU JANGAN SOK IYE DEH, NTAR GW GAMPAR MULUT LU LO KALO BERANI AMA GUE AWAS LU.

3.2 Arsitektur Umum

Pada penelitian akan dilakukan dalam beberapa tahapan proses, yaitu tahapan pengumpulan data yang diperoleh dari aplikasi X(Twitter) dengan proses *crawling* menggunakan *Tweet Harvest* dan disimpan kedalam file berformat .csv. Selanjutnya, data akan melewati tahapan *preprocessing*. Adapun tahapan *preprocessing* terdiri dari *cleaning*, *case folding*, *punctual removal*, *normalization*, *stopword removal*, *stemming* dan *tokenization*. Setelah itu, hasil data yang sudah melewati *tokenization* akan memasuki tahapan *word embedding* dimana data tersebut akan diubah kedalam bentuk vektor. Kemudian, *matrix* yang berisikan vektor dijadikan sebagai input atau masukan terhadap model untuk melalui proses *training*. Hasil dari model akan dilakukan pengujian kemampuan dengan menggunakan data *testing* untuk mengklasifikasi *tweet* pelecehan *online*. Gambar arsitektur umum pada penelitian ini dijabarkan dalam Gambar 3.1 sebagai berikut.



Gambar 3.1. Arsitektur Umum

3.2.1 Input Data

Didalam input data terdapat kolom *text* berisikan *tweet* yang dikirimkan oleh pengguna pada aplikasi X(Twitter) dalam bahasa indonesia serta kolom *class* yang berisikan pelabelan *tweet* pelecehan online. Ada 5 *class* yang akan dikategorikan yaitu *Racist* (Rasis), *Embarrass* (Ejekan & Olokan), *Physically Threatened* (Ancaman Fisik), *Sexually Harassed* (Pelecehan Seksual) dan *Netral* dimana setiap pernyataan yang mengandung pelecehan *online* akan dilabeli 1 dan setiap pernyataan yang tidak mengandung pelecehan seperti pernyataan informasi, pendapat ataupun berita akan dilabeli 0. Pada

penentuan kelas tersebut peneliti dibantu oleh seorang ahli bahasa bernama Afni Sanelsyah Manullang S.S untuk melabeli dan memvalidasi data secara manual berdasarkan kata kunci yang sudah ditentukan pada Bab 2 sebelumnya. Berikut tabel yang berisikan data *tweet* atau pernyataan hasil *crawling* dari media sosial X yang sudah dilabeli dapat dilihat pada tabel 3.2

Tabel 3.2 Dataset Tweet Pelecehan *Online*

No	Text	Class				
		<i>Racist</i>	<i>Embarass</i>	<i>Pysically_ treatened</i>	<i>Sexually_ harrassed</i>	<i>Netral</i>
1	@yantisilver SaveRohingya FUCK KAFIR TERORIS BUDHA, LIAT AJA GW BUNUH LU	1	0	1	0	0
2	cina merupakan negara terbesar dengan penduduk terbanyak	0	0	0	0	1
3	Kita habis-habisan aja, bang. Usir kafir & cina dari nusantara. merekalah racun akidah, tukang sogok, kristenisasi, pelacur, miras. Kafir anjing!	1	1	0	0	0
4	@SHMSBS7 @bjork najis taik goblok setanlo dasar warga setan nyebar hoak mlu pgen gw santet lama lama	0	1	1	0	0
...
5600	Yuk sayang kirimin PAP BUGIL KAMU dong ?? https://t.co/G6a8lM2lun	0	0	0	1	0

Tabel 3.2 Dataset Tweet Pelecehan *Online* (Lanjutan)

No	Text	Class				
		<i>Racist</i>	<i>Embarass</i>	<i>Pysically_tr eatened</i>	<i>Sexually_h arrassed</i>	<i>Netral</i>
5601	Dick riding dyandra tuh gaguna banget anjeng mereka tuh bangsat semuaaa. Dasar penjilat	0	1	0	0	0
5602	Pen gue cekek si putri nuraini ini. Bangsat. Gue gabisa liat orang begitu dh dikasih tau malah ngotot. Anjinggg maju lu sini mo gue cekek.	0	1	1	0	0
5603	@hhaijuli hah gimana? gw bacok lu	0	0	1	0	0
5604	Capek lho jadi saya Di sekolah swasta dibilang pribumi Di sekolah negeri dibilang cina aseng Mau kalian gimana sih? Ttd, Anak ampyang	0	0	0	0	1

Terdapat pembagaaian pada dataset yaitu data *training* dan data *testing* dengan perbandingan 80:20 dari hasil *crawling* data di aplikasi X. Pada penelitian ini menggunakan data *validation* yang akan berguna untuk memastikan bahwa data yang sedang diproses sesuai dengan kebutuhan serta membersihkan dan mempersiapkan data

agar siap digunakan dalam tahap analisis atau pemodelan selanjutnya. Berikut pembagian dataset yang dapat dilihat pada Tabel 3.3.

Tabel 3.3 Pembagian dataset

Dataset Training		Dataset Testing
Training	Validation	1042
5002	602	

3.2.2 Pre-processing

Pada tahapan *preprocessing* akan dilakukan beberapa tahapan yang bertujuan untuk menghasilkan data yang baik dan lebih terstruktur sehingga mudah untuk dipelajari oleh mesin dan dapat meningkatkan akurasi yang baik pula. Berikut merupakan tahapan *preprocessing* pada penelitian ini :

3.2.2.1. Cleaning Data

Tahapan awal yang dilakukan pada *preprocessing* yaitu *cleaning data* yang bertujuan untuk membersihkan data agar meminimalisir *noise* pada proses identifikasi sehingga meningkatkan keakuratan dalam dataset. Adapun tugas yang dilakukan pada proses *cleaning data* yakni menghapus emoji, RT (*Retweet*), *username* (@), URL serta menghilangkan entitas karakter HTML yang tidak diperlukan. Tahapan *cleaning data* pada *tweet* dapat dilihat pada Tabel 3.4 berikut ini .

Tabel 3.4 Contoh penerapan proses *cleaning* data

<i>Tweet sebelum proses cleaning data</i>	<i>Tweet sesudah proses cleacing data</i>
@yunicitra AJG baru aja diomongin, ternyata ibu sama anak sama aja ?? Andaikan di dunia ini gaada yg namanya dosa menyakiti hati orang, udah gw blokir dari tahun lalu ajgg liat aja kau ya babi goblok tolol anjing ga ada otak.!! □ □ ѿ	AJG baru aja diomongin, ternyata ibu sama anak sama aja ?? Andaikan di dunia ini gaada yg namanya dosa menyakiti hati orang, udah gw blokir dari tahun lalu ajgg liat aja kau ya babi goblok tolol anjing ga ada otak.!!

3.2.2.2. Case Folding

Tahapan selanjutnya yaitu *case folding*. Pada tahapan ini dilakukan proses mengubah semua huruf dalam teks menjadi huruf kecil. Tujuan dari *case folding* yaitu untuk mengurasi variasi dan kompleksitas dalam data teks karena tidak semua *tweet* memiliki kekonsistenan dalam penggunaan huruf. Oleh karena itu, diperlukannya perataan semua kata kedalam bentuk standart (huruf kecil). Tahap *case folding* dapat dilihat pada Tabel 3.5 berikut ini.

Tabel 3.5 Contoh penerapan proses *case folding*

<i>Tweet sebelum proses case folding</i>	<i>Tweet sesudah proses case folding</i>
AJG baru aja diomongin, ternyata ibu sama anak sama aja ?? Andaikan di dunia ini gaada yg namanya dosa menyakiti hati orang, udah gw blokir dari tahun lalu ajgg liat aja kau ya babi goblok tolol anjing ga ada otak.!!	ajg baru aja diomongin, ternyata ibu sama anak sama aja ?? andaikan di dunia ini gaada yg namanya dosa menyakiti hati orang, udah gw blokir dari tahun lalu ajgg liat aja kau ya babi goblok tolol anjing ga ada otak.!!

3.2.2.3. Punctual Removal

Pada tahapan *punctual removal* akan dilakukan proses menghapus tanda baca seperti titik, koma, tanda tanya dan tanda seru dari teks. Proses ini bertujuan untuk menghilangkan karakter yang tidak relevan dengan analisis teks atau pemodelan, sehingga memudahkan dalam pemrosesan dan menganalisis teks. Tahap *punctual removal* dapat dilihat pada Tabel 3.6 berikut ini.

Tabel 3.6 Contoh penerapan proses *punctual removal*

<i>Tweet sebelum proses punctual removal</i>	<i>Tweet sesudah proses punctual removal</i>
ajg baru aja diomongin, ternyata ibu sama anak sama aja ?? andaikan di dunia ini gaada yg namanya dosa menyakiti hati orang, udah gw blokir dari tahun lalu ajgg liat aja kau ya babi goblok tolol anjing ga ada otak.!!	ajg baru aja diomongin ternyata ibu sama anak sama aja andaikan di dunia ini gaada yg namanya dosa menyakiti hati orang udah gw blokir dari tahun lalu ajgg liat aja kau ya babi goblok tolol anjing ga ada otak

3.2.2.4. Normalization

Normalization dalam tahapan *preprocessing* merupakan proses mengubah atau memperbaiki bentuk kata seperti *typo* atau singkatan kedalam bentuk standart seperti “ajg” menjadi “anjing”, “aja” menjadi “saja”, “gaada” menjadi “tidak ada” yang tidak sesuai kedalam bahasa indonesia. Hal ini bertujuan untuk menjadikan kata lebih terstruktur dan spesifik karena semakin spesifik suatu kata maka semakin mudah untuk dapat diproses. Pada tahap *normalization* ini penulis menggunakan kamus yang sudah disediakan sebelumnya. Tetapi kamus yang digunakan juga memiliki kekurangan yaitu masih terdapat kata-kata yang belum bisa di normalisasikan karena menggunakan *typo* atau singkatan yang berlebihan atau baru sehingga tidak terdapat pada kamus. Tahap *normalization* dapat dilihat pada Tabel 3.7 berikut ini.

Tabel 3.7 Contoh penerapan proses *normalization*

<i>Tweet sebelum proses normalization</i>	<i>Tweet sesudah proses normalization</i>
ajg baru aja diomongin ternyata ibu sama anak sama aja andaikan di dunia ini gaada yg namanya dosa menyakiti hati orang udah gw blokir dari tahun lalu ajgg liat aja kau ya babi goblok tolol anjing ga ada otak	anjing baru saja diomongin ternyata ibu sama anak sama saja andaikan di dunia ini tidak ada yang namanya dosa menyakiti hati orang udah aku blokir dari tahun lalu anjing lihat saja kamu ya babi goblok tolol anjing tidak ada otak

Tidak Baku	Baku
ajg	anjing
aja	saja
ga	tidak
gw	aku
kau	kamu
liat	lihat

3.2.2.5. Stopword Removal

Tahapan *stopword removal* merupakan proses dimana kata-kata yang umum dan memiliki sedikit kontribusi terhadap makna teks akan dihapus dari teks. Tahapan ini bertujuan untuk membuat kalimat menjadi lebih sederhana, mengurangi dimensi dataset

teks dan mempercepat pemrosesan teks tanpa mengubah makna dari kalimatnya. Kata yang tidak terlalu memiliki makna dalam sebuah kalimat yaitu kata penghubung dimana hanya menjelaskan konteks dan hanya dipahami oleh manusia saja. Kata-kata tersebut seperti “dan,” “atau,” “di,” “dari,” “ke,” dan sebagainya. Adapun *library* yang digunakan adalah NLTK (*Netural Language Tool Kit*). Dengan *stopwords*, fokus akan diarahkan pada kata-kata yang memiliki makna penting dalam konteks analisis. Tahap *stopword removal* dapat dilihat pada Tabel 3.8 berikut ini.

Tabel 3.8 Contoh penerapan proses *stopword removal*

<i>Tweet sebelum proses <i>stopword removal</i></i>	<i>Tweet sesudah proses <i>stopword removal</i></i>
anjing baru saja diomongin ternyata ibu sama anak sama saja andaikan di dunia ini tidak ada yang namanya dosa menyakiti hati orang udah aku blokir dari tahun lalu anjing lihat saja kamu ya babi goblok tolol anjing tidak ada otak	anjing diomongin ternyata ibu anak dunia tidak namanya dosa menyakiti hati orang blokir anjing lihat babi goblok tolol anjing tidak otak

3.2.2.6. Stemming

Pada tahapan ini akan dilakukan proses pemotongan akhiran kata (suffix) atau awalan kata (prefix) untuk mendapatkan kata dasar atau bentuk kata yang lebih umum. Tujuan *stemming* dalam mengurangi imbuhan yaitu guna mengoptimalkan proses dan kinerja dari algoritma yang digunakan. Adapun contoh imbuhan seperti dalam kata “memakan” akan dipotong menjadi bentuk dasarnya yang adalah “makan”, dalam kata “berlari” menjadi “lari” dan sebagainya. Proses *stemming* ini penulis menggunakan *library* Sastrawi. Tahap *stemming* dapat dilihat pada Tabel 3.9 berikut ini.

Tabel 3.9 Contoh penerapan proses *stemming*

<i>Tweet sebelum proses <i>stemming</i></i>	<i>Tweet sesudah proses <i>stemming</i></i>
anjing diomongin ternyata ibu anak dunia tidak namanya dosa menyakiti hati orang blokir anjing lihat babi goblok tolol anjing tidak otak	anjing omong nyata ibu anak dunia tidak nama dosa sakit hati orang blokir anjing lihat babi goblok tolol anjing tidak otak

3.2.2.7. Tokenization

Pada tahapan tokenisasi akan dilakukan pembagian teks menjadi unit-unit yang lebih kecil yang disebut sebagai token. Token dapat berupa kata, frasa atau karakter terpisah. Pada penelitian ini dilakukan pembagian berdasarkan kata yang dimana kata tersebut akan menjadi *input* pada proses selanjutnya. Tujuannya dari tokenisasi untuk mempermudah pemrosesan dan analisis teks dengan mengidentifikasi unit-unit terpisah yang relevan. Tahap *tokenization* dapat dilihat pada Tabel 3.10 berikut ini.

Tabel 3.10 Contoh penerapan proses *tokenization*

<i>Tweet sebelum proses tokenization</i>	<i>Tweet sesudah proses tokenization</i>
anjing diomongin ternyata ibu anak dunia tidak namanya dosa menyakiti hati orang blokir anjing lihat babi goblok tolol anjing tidak otak	"anjing","omong","nyata","ibu","anak","dunia","tidak","nama","dosa","sakiti","hati","orang","blokir","anjing","lihat","babi","goblok","tolol","anjing","tidak","otak"

3.2.3 Word Embedding

Word Embedding adalah tahapan yang dilakukan untuk mengubah token-token yang telah terbentuk menjadi sebuah *vector*. Tahapan ini memungkinkan model pembelajaran mesin untuk memahami dan menganalisis hubungan antara kata-kata dan menyematkan (embed) ke dalam ruang *vector* yang menggambarkan makna mereka. Pada penelitian ini tahapan *word embedding* dilakukan menggunakan *library fastText*. Berikut merupakan contoh dari *embedding matrix* yang adalah vektor dari kata "sialan" pada gambar 3.2

0.0385	0.0052	0.0329	0.1034	-0.0343	-0.0889
0.0206	0.0257	-0.025	-0.1365	0.0008	-0.0109
-0.0203	-0.0251	0.0277	0.0097	0.0408	-0.0012
0.0236	-0.0348	0.0167	-0.0162	-0.0032	0.0623
-0.0061	-0.0056	-0.0335	-0.0314	0.1194	-0.0317
0.0558	-0.0101	0.0311	-0.0179	0.0805	-0.0388
0.0149	0.0108	0.0616	-0.0102	0.0322	0.0099
-0.0335	0.0302	0.0039	0.0316	-0.0588	0.0135
0.0131	-0.0381	-0.0189	0.0187	-0.0184	0.024
0.0028	0.0054	0.0693	0.0247	-0.0293	0.031
0.0031	0.0219	-0.0303	0.0173	0.0303	0.0119
-0.0228	0.0008	0.027	-0.0263	0.0614	0.0132
0.0217	0.0023	0.0034	0.0206	-0.0191	-0.0124
-0.0274	0.0019	0.0192	0.0104	0.0027	-0.0071
-0.0354	0.0106	-0.0512	-0.0893	0.0191	0.0003
-0.0055	-0.0075	0.0327	0.0097	-0.0774	-0.13940001
-0.0016	0.0247	-0.0078	0.0133	-0.0065	0.0306
0.0053	-0.0108	0.0494	0.0101	0.0131	-0.0173
0.0144	0.0042	-0.0843	-0.044	0.0032	0.023
-0.0052	-0.0216	-0.0696	-0.0173	0.0233	0.0223
-0.0578	-0.0198	-0.0316	-0.0263	-0.0057	-0.017
0.002	-0.0229	-0.0276	-0.0236	0.0095	-0.0121
0.0062	-0.0059	0.1086	-0.0195	-0.0115	0.0058
0.0457	0.0137	-0.004	-0.1928	0.0152	0.0181
-0.0214	-0.0043	0.0177	-0.0166	-0.0298	0.0181
0.0134	0.0175	-0.0406	0.0113	0.171	0.0069
-0.0192	0.0234	-0.0229	-0.0459	0.25799999	0.1288
-0.0347	-0.0137	-0.0172	-0.0205	0.0151	0.0097
-0.0253	0.011	0.0189	-0.0005	0.0289	0.0533
-0.024	0.0341	-0.0081	0.0048	-0.0057	0.0024
0.0605	0.0011	0.0231	-0.0491	-0.0074	0.0434
0.1216	-0.0213	-0.0174	-0.0147	0.0017	-0.027
-0.0355	0.0096	-0.0397	-0.0168	0.0179	-0.0171
-0.0212	0.0882	0.0235	0.0176	0.0293	0.0232
0.0334	0.0038	0.0069	0.0019	-0.0267	-0.0021
0.013	0.0302	0.0055	0.0166	0.0328	0.0075
-0.0248	0.0004	-0.0364	-0.108	-0.0088	0.0104
0.0244	0.0035	-0.0397	-0.0024	-0.037	0.0121

Gambar 3.2 Embedding Matrix

3.2.4 Model Building

Setelah melakukan *word embedding*, langkah selanjutnya adalah *model building*. Pada tahapan ini bertujuan untuk membentuk model dengan memanfaatkan algoritma *Gated Recurrent Unit*. Model akan dilengkapi dengan beberapa layer yang menerima inputan sehingga terjadinya *underfitting* dan *overfitting*. Berikut layer yang digunakan dalam penelitian ini :

1. *Embedding Layer*

Embedding Layer merupakan sebuah lapisan membentuk matriks input dari hasil sekuens hasil tokenisasi dan sudah diberikan *padding* agar memiliki panjang yang sama dan akan menerima parameter *weight* yang berisikan *matrix embedding* hasil dari proses *word embedding*. Sesuai dengan dimensi yang dihasilkan, output yang dibentuk merupakan vektor berdimensi 300. Berikut potongan kode dari *embedding layer*.

```
input_layer = Input(shape=(max_len,))
embedding_layer= Embedding(nb_words,embed_dim,input_length=max_len,
weights=[embedding_matrix],trainable=False)(input_layer)
```

Pada *embedding layer* akan menerima lapisan input dari model. Lapisan ini menentukan bentuk data input yang akan diterima oleh model, '*max_len*' merupakan panjang maksimum dari setiap sampel data yang akan dimasukkan ke dalam model.

Embedding layer menggunakan lapisan *embedding* yang bertanggung jawab untuk melakukan konversi representasi kata-kata dalam bentuk teks menjadi vektor, antara lain *nb_words* merupakan jumlah kata unik pada dataset, *embed_dim* merupakan dimensi dari vektor *embedding*, *input_lenght* menunjukkan panjang maksimum dari setiap sampel data yang dimasukkan kedalam model, *self.embedding_matrix* merupakan sekumpulan angka yang sudah dimiliki untuk kata-kata yang sudah ada, *trainable_False* merupakan perintah untuk tidak mengubah angka-angka yang sudah ada selama melatih model.

2. *Gated Recurrent Unit Layer*

Setelah melakukan tahapan *Embedding Layer*, maka output dari *Embedding layer* akan menjadi inputan bagi lapisan GRU dan akan dilatih menggunakan model *Gated Recurrent Unit* yang terdiri dari 32 unit neuron. Berikut potongan kode dari *Gated Recurrent Unit Layer*.

```
gru_layer = GRU(32)(embedding_layer)
```

Pada kode diatas, membuat *GRU Layer* pada model yang membantu model memahami pola urutan dalam data, seperti urutan kata dalam teks. Parameter unit neuron 32 pada *GRU layer* ini bisa dianggap sebagai kemampuan atau kompleksitas *GRU layer*, semakin banyak unit maka semakin rumit modelnya. Dalam hal ini juga

menggunakan *batch size* sebesar 64 untuk dapat mengoptimalkan efisiensi model. Berikut cara kerja GRU *layer*:

- a. Langkah awal yang dilakukan pada *layer* ini dimulai dengan menerima input yang berasal dari *layer* sebelumnya yaitu *embedding layer* yang berupa vektor 300 dimensi.
- b. Selanjutnya, inputan tersebut akan masuk kedalam dua *gate* pada *layer* GRU, yaitu *reset gate* dan *update gate*. *Reset gate* akan memutuskan seberapa banyak informasi yang harus dilupakan atau dipertahankan, sedangkan *update gate* akan mengatur seberapa banyak informasi masa lalu yang harus tetap disimpan. Lalu, akan menghasilkan output berupa nilai probabilitas terhadap masing-masing kelas yaitu positif (1) dan negatif (0) sehingga hal ini akan memberikan estimasi seberapa jauh model dapat menganalisis input atau text berdasarkan masing-masing kelas.
- c. Langkah selanjutnya adalah menentukan kandidat *hidden state*. Kandidat *hidden state* akan menggabungkan informasi dari langkah sebelumnya yang sudah diperbaharui berdasarkan nilai *update gate* dengan input baru menggunakan fungsi aktivasi. *Output* yang dihasilkan dari kandidat *hidden state* akan menjadi dasar perhitungan *hidden state*.
- d. Tahap berikutnya adalah pada *hidden state*, *output* hasil akhir atau informasi akhir dari proses sebelumnya akan diteruskan ke *time step* berikutnya sebagai *hidden state*.

3. Dense Layer

Dense Layer merupakan lapisan *fully connected network* yang dimana setiap unit neuron dalam lapisan tersebut terhubung dengan setiap unit di lapisan sebelumnya dan setelahnya. *Dense Layer* akan menghasilkan output antara 0 sampai 1 yang merupakan hasil yang diperoleh dari pemodelan. Pada *dense layer* ini menggunakan aktivasi ReLu (*Rectified Linier Unit*) karena pada penelitian ini adalah *multiclass* hasil klasifikasi yang diperoleh dari pemodelan. Selain itu, relu juga memiliki karakteristik yang sederhana dan efektif dalam mengatasi masalah *vanishing gradient*. Berikut adalah potongan kode dari *dense layer*.

```
second_dense = Dense(units='32',activation='relu')(gru_layer)
```

Pada kode diatas, *dense layer* terhubung sepenuhnya dengan setiap neuron di layer sebelumnya yaitu *gru layer* dimana *dense layer* akan mengambil output dari GRU *layer* sebelumnya sebagai inputnya. Pada *dense layer* menggunakan parameter unit neuron sebanyak 32 dan setiap neuron berkontribusi dalam pemrosesan data serta mempelajari pola yang ada. Dense layer memiliki fungsi utama untuk mempelajari pemetaan fitur yang lebih abstrak dari data inputan, maka dengan melakukan kombinasi linier dan menerapkan ke fungsi aktivasi non-linier, dense layer mampu mengekstraksi pola-pola yang kompleks dari data.

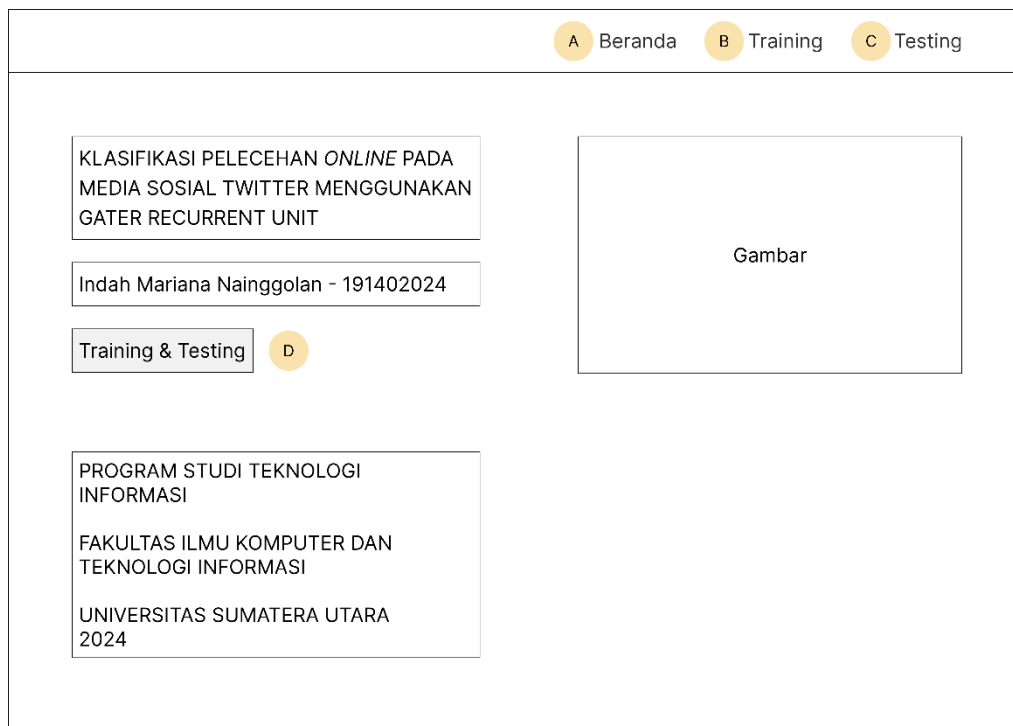
Pada penelitian ini menggunakan 50 epoch untuk menghasilkan output yang baik. Dalam pemodelan menggunakan fungsi optimasi Adam (Adaptive Moment Estimation) dengan tujuan untuk menemukan nilai parameter yang menghasilkan loss atau kerugian yang minimum. Permasalahan yang ada di dalam Machine Learning dapat diselesaikan dengan mengoptimasi Adam dengan menggunakan dataset dalam jumlah yang lebih banyak.

3.3 Perancangan Aplikasi Sistem

Pada tahapan ini merupakan penjelasan tentang perancangan antarmuka sistem aplikasi klasifikasi pelecehan *online* pada media sosial X(Twitter) menggunakan *Gated Recurrent Unit* yang bertujuan untuk memberikan gambaran umum seperti tata letak menu dan alur sistem yang dapat memudahkan pengguna ketika melakukan pengoperasian sistem.

3.3.1 Gambaran Tampilan Halaman Depan

Halaman depan atau bisa disebut sebagai *hero page* yang akan muncul pertama kali saat sistem dijalankan. Pada halaman ini akan menampilkan informasi tentang judul penelitian, nama dan nim penulis, serta ada beberapa *button* yang memiliki fungsi masing-masing. Berikut gambaran tampilan halaman depan web aplikasi yang dapat dilihat pada gambar 3.3.



Gambar 3.3 Gambaran halaman depan aplikasi web

Keterangan:

1. Label A yaitu menu beranda yang akan menampilkan halaman awal aplikasi web
2. Label B yaitu menu *training* dimana ketika pengguna klik menu tersebut akan beralih ke halaman *training* data
3. Label C yaitu menu *testing* dimana ketika pengguna klik menu tersebut akan beralih ke halaman *testing* data
4. Label D yaitu *button training* dan *testing* dimana ketika pengguna klik menu tersebut akan beralih ke proses *training* data

3.3.2 Gambaran Tampilan Training Data

Tampilan training data merupakan halaman dimana pengguna dapat melakukan proses training pada dataset ke dalam model. Adapun caranya yaitu meng-*input* terlebih dahulu file dataset berformat .csv kemudian klik *button* mulai training. Setelah selesai akan ditampilkan hasil *preprocessing* dan *training* data dan dilengkapi dengan nilai akurasi serta *loss* grafik *history* keakuratan data *training* dan data validasi. Berikut gambaran tampilan training data pada Gambar 3.4.

Gambar 3.4 Gambaran tampilan *training* data

Keterangan:

1. Label A yaitu digunakan untuk input epoch yang akan menentukan berapa kali model akan melakukan pembelajaran saat melewati dataset *training*
2. Label B yaitu untuk memasukkan file dataset berformat .csv untuk dilakukan proses *training*
3. Label C yaitu tombol untuk mengunggah data *training*
4. Label D yaitu menampilkan tabel data hasil *preprocessing*
5. Label E yaitu menampilkan grafik akurasi masing masing kelas
6. Label F yaitu menampilkan grafik *loss* masing masing kelas

3.3.3 Gambaran Tampilan Testing Data

Tampilan training data merupakan halaman dimana pengguna dapat melakukan proses testing pada dataset. Adapun caranya yaitu menginput terlebih dahulu file dataset berformat .csv kemudian klik tombol mulai testing. Setelah selesai akan ditampilkan tabel prediksi data hasil *testing*, dilengkapi dengan *classification report* yang berisikan nilai dari *Precision*, *Recall*, *F1-Score* dan *Accurasy* dan menampilkan *classification matrix*. Berikut gambaran tampilan testing data pada Gambar 3.5.

[Beranda](#) [Training](#) **Testing**

TESTING
 Upload dataset untuk proses testing

Input file

Pilih file
No file chosen

Mulai testing

Hasil preprocessing

C

D
 Classification Report

E
 Confussion Matrix

Gambar 3.5 Gambaran tampilan *testing* data

Keterangan:

1. Label A yaitu untuk memasukkan file dataset berformat .csv untuk dilakukan proses *training*
2. Label B yaitu tombol untuk mengunggah data *testing*
3. Label C yaitu menampilkan tabel prediksi data hasil dari proses *testing*
4. Label D yaitu menampilkan *classification report* masing masing kelas
5. Label E yaitu menampilkan *classification matrix* masing masing kelas

Gambar 3.6 Gambaran tampilan uji coba *realtime*

Keterangan:

1. Label A yaitu untuk memasukkan kalimat atau text *tweet*
2. Label B yaitu tombol untuk mengunggah data *testing*
3. Label C yaitu menampilkan tabel prediksi data hasil dari proses *testing*

3.4 Metode Evaluasi

Metode evaluasi digunakan untuk mengetahui seberapa akurat dan berhasil sistem dalam melakukan klasifikasi pelecehan *online* dan metode evaluasi yang digunakan pada penelitian ini adalah *confussion matrix* dengan cara menghitung jumlah data benar dan jumlah data yang salah. Berikut pengukuran evaluasi yang dilakukan berdasarkan *confussion matrix* pada tabel dengan perhitungan *Accuracy*, *Precision*, *Recall* dan *F1-Score*. Berikut penerapan *confussion matrix* yang dapat dilihat pada tabel 3.11.

Tabel 3.11 Implementasi Evaluasi *Confussion Matrix*

Classification		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP (True Positive)	FP (False Positive)
	Negative	TN (True Negative)	FN (False Negative)

Keterangan:

1. TP (*True Positive*) : Banyaknya nilai actual yang memiliki nilai positif dan data prediksi memiliki nilai positif.
2. TN (*True Negative*) : Banyaknya nilai actual yang memiliki nilai positif dan data prediksi memiliki nilai negatif.
3. FP (*False Postive*) : Banyaknya nilai actual yang memiliki nilai negatif dan data prediksi memiliki nilai positif.
4. FN (*False Negative*) : Banyaknya nilai actual yang memiliki nilai negatif dan data prediksi memiliki nilai ngeatif.

BAB 4

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Implementasi Sistem

Pada perancangan sistem web klasifikasi pelecehan *online* pada media sosial X menggunakan *Gated Recurrent Unit* digunakan beberapa perangkat keras (*hardware*) dan perangkat lunak (*software*) sebagai sarana pendukung, antara lain :

4.1.1. Spesifikasi Hardware dan Software

Perangkat keras yang digunakan dalam melakukan perancangan sistem pada penelitian ini memiliki spesifikasi sebagai berikut:

1. Laptop ASUS X441UB
2. Processor 7th Gen Intel CORE i3-7020U
3. Memory RAM 8 GB
4. SSD 258GB

Perangkat lunak yang digunakan dalam melakukan perancangan sistem pada penelitian ini memiliki spesifikasi sebagai berikut:

1. Sistem Operasi: Windows 10 Home 64 bit Operation System
2. Bahasa pemrograman Python 3.8 dengan *library* GRU, numpy, nltk, pandas, matplotlib, library sastrawi, keras dan tensorflow
3. Microsoft *Visual Studio Code*

4.1.2. Penerapan Tampilan Antarmuka

Berikut perancangan sistem yang sudah dibentuk pada bab 3, maka penerapan tampilan antarmuka sistem, yaitu sebagai berikut:

1. Tampilan Halaman Depan/Beranda

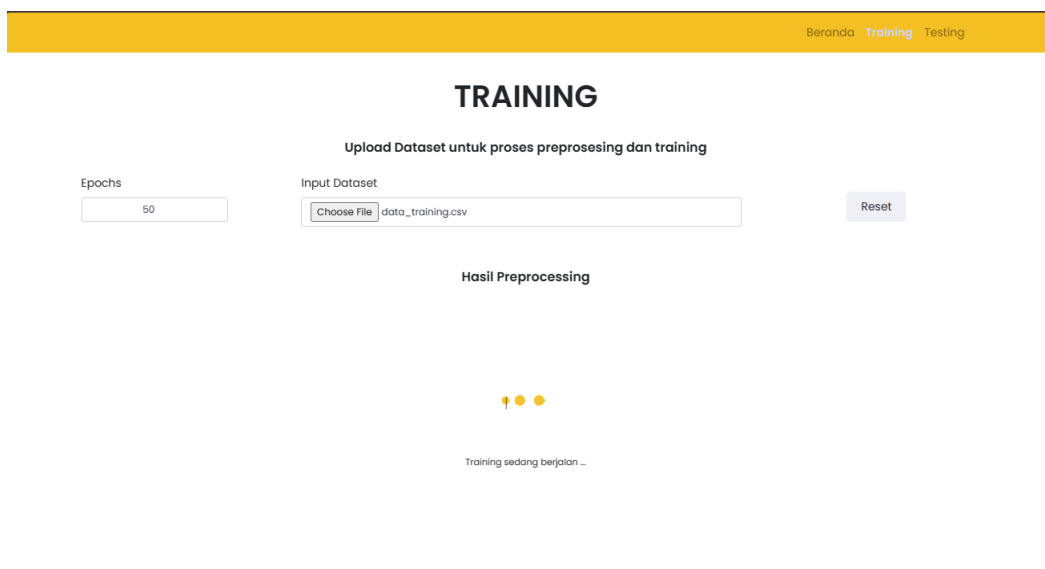
Pada tampilan halaman depan akan menampilkan *hero page* ketika sistem sudah mulai dijalankan oleh pengguna. Terdapat beberapa fitur didalamnya antara lain menu navbar yang berisikan menu *training* dan *testing* yang bisa digunakan oleh pengguna saat melakukan kedua proses tersebut. Kemudian, pada halaman depan juga akan menampilkan identitas penulis seperti judul, nama dan nim. Tampilan halaman depan/beranda dapat dilihat pada gambar 4.1.



Gambar 4.1 Tampilan halaman depan/beranda

2. Tampilan Halaman Training Data

Pada tampilan halaman training data akan ditampilkan ketika pengguna menjalankan fitur training data. Pada halaman ini, pengguna dapat mengunggah file dataset yang berformat .csv serta menekan tombol mulai *training* data. Selain itu, pengguna juga dapat menentukan epoch untuk menghasilkan hasil yang lebih baik. Pada sistem ini menggunakan epoch 50. Tampilan halaman training data dapat dilihat pada gambar 4.2.



Gambar 4.2 Tampilan halaman training data

Setelah selesai mengunggah file dataset maka akan menampilkan hasil training data dalam bentuk tabel. Adapun hasil dari proses tersebut adalah tabel hasil *preprocessing*, grafik nilai akurasi untuk masing-masing kelas, dan grafik nilai loss untuk masing-masing kelas. Terdapat fitur tambahan pada halaman *training* data yaitu adanya fitur *search* dan *pagination* agar memudahkan pengguna ketika mencari kalimat atau kata kunci tertentu.

The screenshot displays a table titled 'Hasil Preprocessing'. It includes a 'Show' dropdown set to '10' and a 'Search' input field. The table has columns for 'Text', 'Preprocessing', 'Racist', 'Embarrass', 'Physically_threatened', 'Sexually_harassed', and 'Netral'. The data rows show various text samples and their corresponding sentiment labels (0 or 1).

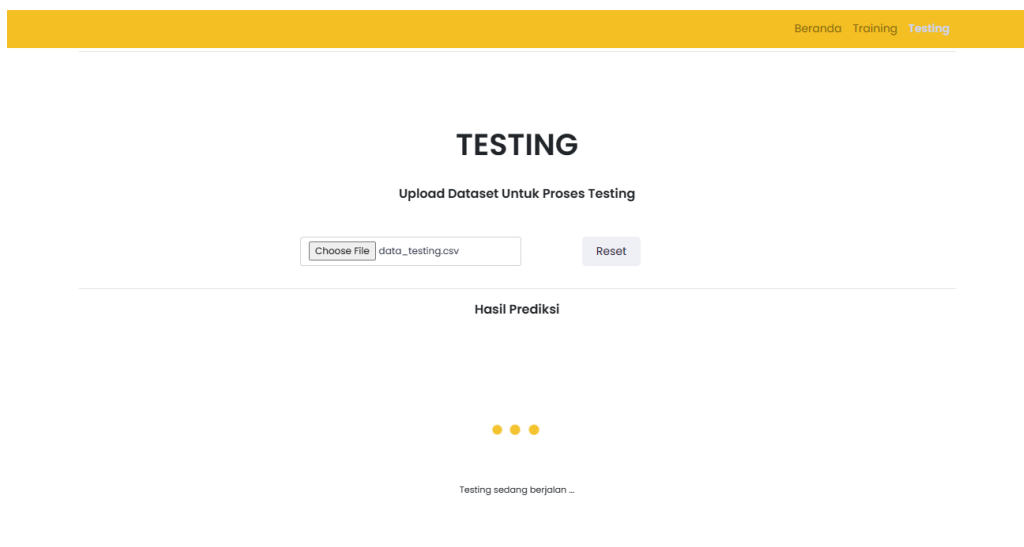
Text	Preprocessing	Racist	Embarrass	Physically_threatened	Sexually_harassed	Netral
@AbahDeon Dasar pelacur ?? orang cuma bercadar aja dibacotin begini, dia ga merugikan orang lain juga. Begini bentuk manusia hati tapi iblis	dasar lacur orang cadar dibacotin tidak rugi orang bentuk manusia hati iblis	0	0	0	1	0
@aniesbaswedan @NasDem @PKSejahtera @DPP_PKB @DPP_PartaiUmmat @desa_bersatu @cakimiNOW Apa yg salah ??? Anies gak bokep koqqq...wkwwkw Ganjar tuhnh ngebokep sejatiii...wkwwkw Kalo luuu sakit hati dgn habib serang habib dong...amin gak ada hubungannya dgn bokepan...wkwwkw Dasar pelacur otaknya najiss...wkwwk cuiihhh?????????????	salah anies tidak bokep koqqq ganjar tonton bokep sejatiii sakit hati habib serang habib amin tidak hubung bokepan dasar lacur otak najiss cuiihhh	0	1	0	1	0
@geloraco Otakmu yang palsu, dasar pelacur	otak palsu dasar lacur	0	1	0	1	0
@Ina F&W @Mn&Idn	dasar lacur munikari	0	0	0	1	0

Gambar 4.3 Tampilan hasil training data

3. Tampilan Halaman Testing Data

Pada tampilan halaman testing data akan ditampilkan ketika pengguna menjalankan fitur *testing* data. Pada halaman ini, pengguna dapat mengunggah

file dataset yang berformat .csv serta menekan tombol mulai *testing* data. Tampilan halaman ini dapat dilihat pada gambar 4.4.



Gambar 4.4 Tampilan halaman testing data

Setelah selesai proses testing, akan menghasilkan tabel prediksi, *classification report* dan *confussion matrix* dari setiap masing-masing kelas. Tampilan hasil prediksi dapat dilihat pada gambar 4.5.

Beranda
Training
Testing

Choose File

data_testing.csv

Reset

Hasil Prediksi

Show

10

entries

Search:

Text	Preprocessing	Prediction				Keterangan
		Racist	Embarrass	Physically_threatened	Sexually_harassed	
@txtdarisisange lu bisa gak sih gak usah nyampah, dasar pelacur	tidak tidak sampah dasar lacur	0	0	0	1	pelecehan
Freeport takkan berdiam saja setelah negosiasi kalah..	freeport tidak diam negosiasi kalah	0	0	0	1	pelecehan
Mungkin maksudnya aseng asing yang bernama MA LAI KAT	maksud asing asing nama lai kat	0	0	0	0	bukan pelecehan
RT @julius_hans: @Majid_Zyn @kevinadrian666 GW BUNUH LU BERDUA KALO KETEMU AWAS AJA...	rt bunuh dua ketemu awas	0	0	1	0	pelecehan

Gambar 4.5 Tampilan hasil testing data

Selain itu terdapat fitur uji coba realtime, dimana fitur ini dapat digunakan dengan cara memasukkan kalimat yang ingin di testing dan akan menghasilkan hasil prediksi secara *real time*. Berikut tampilan uji coba realtime dapat dilihat pada gambar 4.6.

Beranda Training Testing

Data Testing - Uji Coba Realtime

Masukkan Text

Reset

Hasil Prediksi Real Time

Input	Racist	Embarrass	Physically_threatened	Sexually_harassed	Keterangan
yo @scndbys dasar lu pelacur hina anjing	Netral	Positif	Netral	Positif	pelecehan

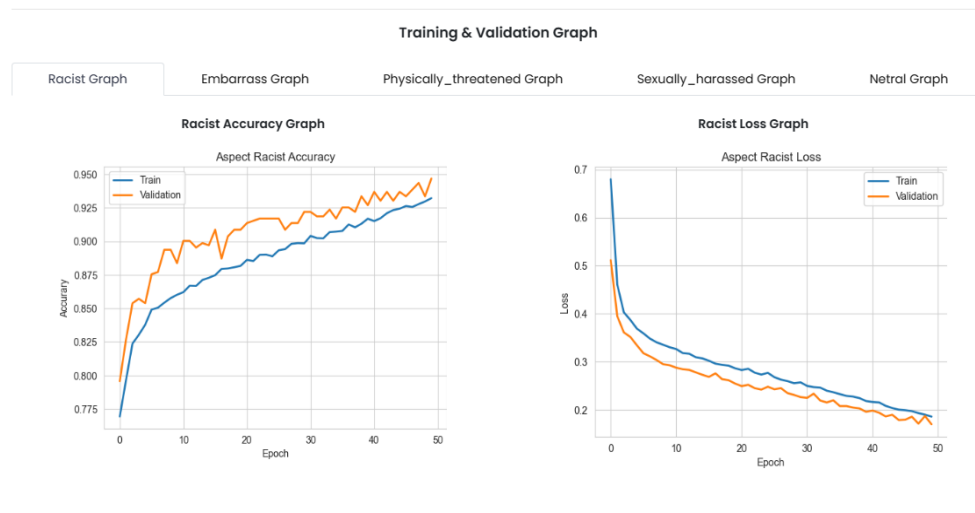
Gambar 4.6 Tampilan uji coba realtime

4.2. Implementasi Model

4.2.1. Model Training

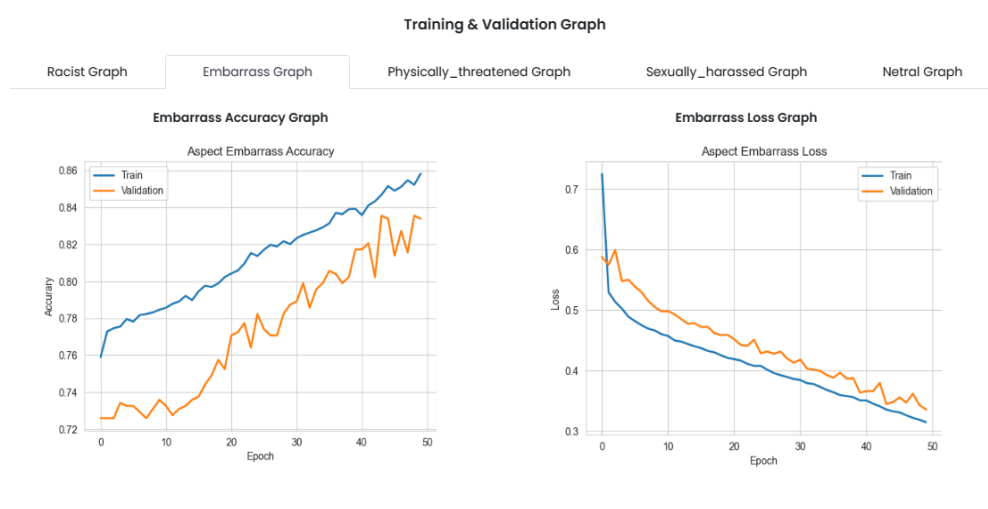
Dalam proses pelatihan model, penulis melakukan beberapa percobaan dengan mengubah *hyperparameter* untuk mendapatkan hasil dan performa model yang optimal. Penulis mengatur ulang *hyperparameter* seperti *batch size*, jumlah unit neuron, dan epoch dalam berbagai percobaan pelatihan. Hal tersebut merupakan proses *hyperparameter tuning*. Model dapat dikatakan baik ketika model mencapai akurasi tertinggi tetapi tidak *overfit*.

Pada penelitian, penulis menemukan model dengan performa terbaik dengan menggunakan *batch size* sebesar 64 dan banyaknya epoch yang digunakan sebanyak 50 serta jumlah unit neuron yang digunakan sebanyak 32 unit. Berdasarkan proses tersebut sehingga didapati akurasi validasi masing-masing kelas diperoleh rata-rata sebesar 91.7% dengan *loss* sebesar 25%. Berikut hasil grafik akurasi dan grafik loss pada masing-masing kelas untuk model paling baik dapat dilihat pada gambar 4.7 sampai 4.11.



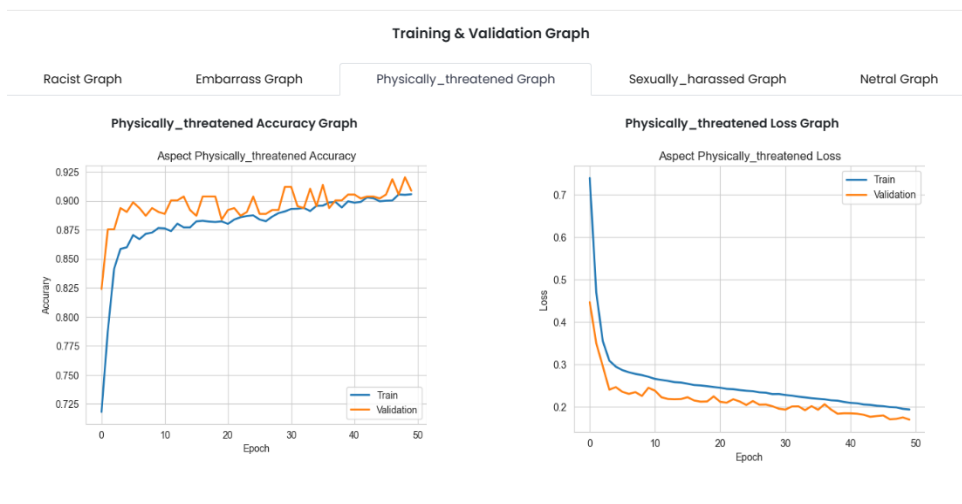
Gambar 4.7 Tampilan grafik akurasi dan grafik loss kelas racist

Pada gambar di atas, grafik akurasi pada kelas rasis data *training* dan data *validation* menunjukkan tingkat akurasi mencapai 0.95 dengan nilai *loss* sebesar 0.19. Hal tersebut menunjukkan kinerja dengan model yang cukup baik.



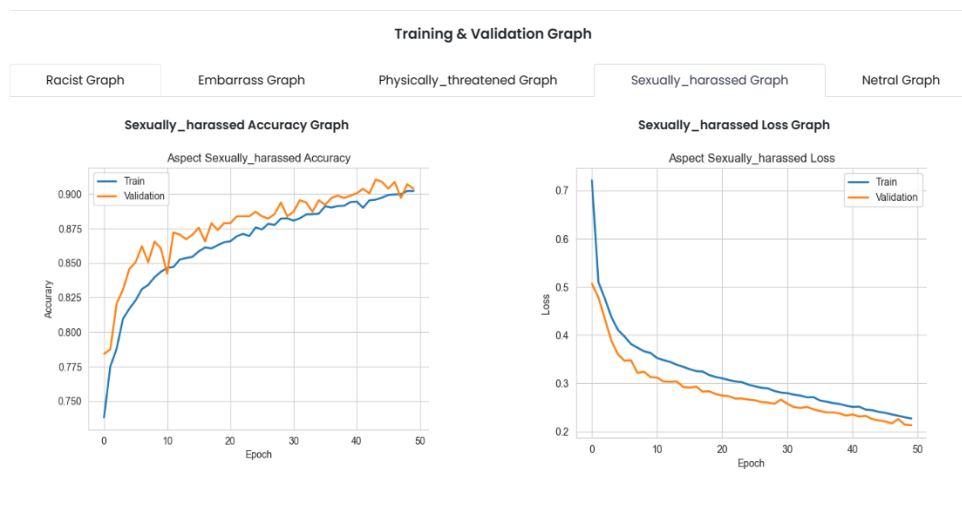
Gambar 4.8 Tampilan grafik akurasi dan grafik loss kelas embarrass

Pada gambar di atas, grafik akurasi pada kelas *embarrass* menunjukkan tingkat akurasi mencapai 0.86 dengan nilai *loss* sebesar 0.31.



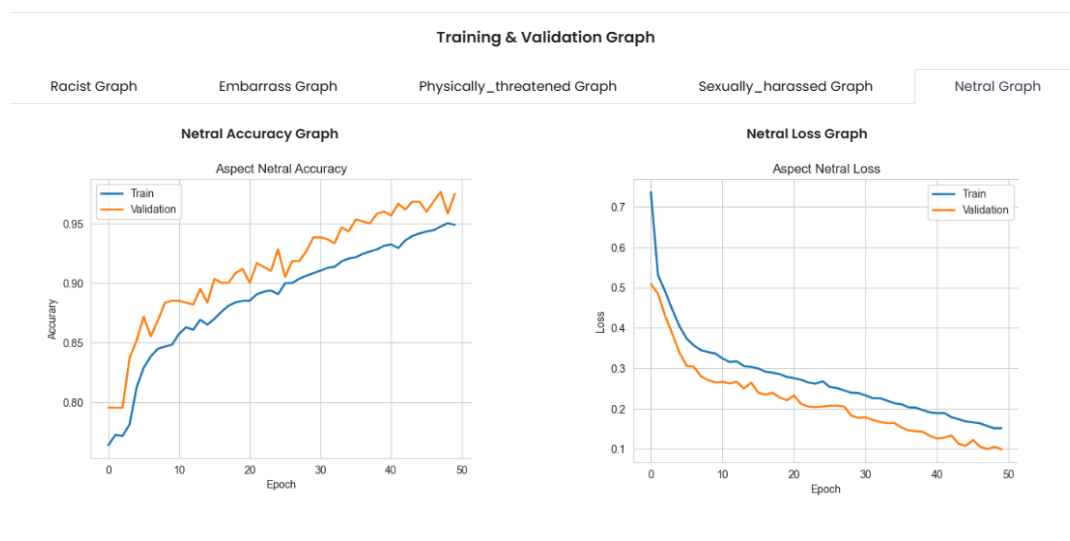
Gambar 4.9 Tampilan grafik akurasi dan grafik loss kelas physical threatened

Pada gambar di atas, grafik akurasi pada kelas *physical threatened* menunjukkan tingkat akurasi mencapai 0.91 dengan nilai loss sebesar 0.19.



Gambar 4.10 Tampilan grafik akurasi dan grafik loss kelas sexually harassed

Pada gambar diatas, grafik akurasi pada kelas *sexually harassed* menunjukkan tingkat akurasi mencapai 0.90 dengan nilai *loss* sebesar 0.22.



Gambar 4.11 Tampilan grafik akurasi dan grafik loss kelas netral

Pada gambar diatas, grafik akurasi pada kelas netral menunjukkan tingkat akurasi mencapai 0.95 dengan nilai *loss* sebesar 0.15.

Setelah itu, dataset akan dibagi menjadi 2 bagian yaitu data *training* dan data *validation* dengan perbandingan 80% untuk data *training* dan 20% untuk data *validation* yang akan dilatih ke dalam model. Berikut hasil data *training* yang dapat dilihat pada tabel 4.1.

Tabel 4.1 Hasil *Training*

Text	Pre processing	Class				
		Racist	Embarass	Physical_ Threatened	Sexually_ harrassed	Netral
@AbahDeon Dasar pelacur ?? orang cuma bercadar aja dibacotin begini, dia ga merugikan orang lain juga. Begini bentukan manusia hati tapi iblis	dasar lacur orang cadar dibacotin tidak rugi orang bentuk manusia hati iblis	0	0	0	1	0
Berapa banyak pelacur yang ditangkap hanya	berapa bayak lacur	0	0	0	0	1

Tabel 4.1 Hasil *Training* (Lanjutan)

Text	Pre processing	Class				
		Racist	Embarass	Physical_ Threatened	Sexually_ harrassed	Netral
untuk dipermalukan, dilecehkan, dan bahkan diperkosa pihak berwajib? \Bukankah pelecehan ini adalah pornografi dalam arti sebenarnya?	tangkap malu leceh perkosa pihak wajib tidak leceh pornografi arti benar					
@aniesbaswedan @NasDem @PKSejahtera @DPP_PKB @DPP_PartaiUmmat @desa_bersatu @cakimiNOW Apa yg salah ??? Anies gak bokep koqqq...wkwkww Ganjar tuhhh ngebokep sejati..wkwkww Kalo luuu sakit hati dgn habib serang habib dong...amin gak ada hubungannya dgn bokepan..wkwkww Dasar pelacur otaknya najisss...wkkw cuiihhh?????????????	salah anies tidak bokep koqqq ganjar tonton bokep sejatiiii sakit hati habib serang habib amin tidak hubung bokepan dasar lacur otak najisss cuiihhh	0	1	0	1	0

Tabel 4.1 Hasil *Training* (Lanjutan)

Text	Pre processing	Class				
		Racist	Embarass	Physical_ Threatened	Sexually_ harrased	Netral
@geloraco Otakmu yang palsu, dasar pelacur	otak palsu dasar lacur	0	1	0	1	0
@Joe_58w @MprAldo @ganjarpranowo Wkk..wkk dasar pelacur dan mucikari, gue yakin loe pasti ngerjain yg sama begitu juga anak2 loe pasti kalo kgk pengangguran ya tukang jahar atau jilathin panthat junjungan loe! Wkkk....wkk ????????taekbeer x3 kgk neh	dasar lacur mucikari anak anak tidak anggur tukang jahar jilatin pantat junjung taekbeer x tidak nih	0	0	0	1	0
@pretamini @idextratime @barcacentre lah gw demit darimana nya kontrol, lu kalo jadi decul minimal goblok aja jangan goblok banget dasar pelacur	demit kontrol decul minimal goblok jangan goblok banget dasar lacur	0	1	0	1	0
Menurut World Prospects Population, Cina adalah negara	World prospects population	0	0	0	0	1

Tabel 4.1 Hasil *Training* (Lanjutan)

Text	Pre processing	Class				
		Racist	Embarass	Physical_ Threatened	Sexually_ harrassed	Netral
dengan penduduk terbanyak di dunia nomor 1 pada tahun 2022 dengan jumlah penduduk sekitar 1.439.323.776 jiwa.	cina negara penduduk banyak dunia nomor tahun jumlah penduduk jiwa					
@Sigiwarmada Kadrin Yaman ngehina nenek moyang kita..dasar kadrin Yaman, blom berkuasa si Yaman dah berani begini, apalagi si Yaman dah berkuasa.. Habis lah Pribumi., ?????? https://t.co/lrQez4OeBg	kadal gurun yaman hina nenek moyang dasar kadal gurun yaman kuasa yaman berani yaman kuasa habis pribumi	1	0	0	0	0
. @bllssom Hahaha malah diblock, dasar pelacur lu, gw cari sampe ketemu abis tu gw bunuh lu	blok dasar lacur cari ketemu habis bunuh	0	0	1	1	0

4.2.2. Hasil Pengujian Sistem

Pada tahap pengujian sistem, hasil model yang sudah dibangun akan disimpan pada proses *training*, selanjutnya akan diuji menggunakan data *testing* agar mengetahui bagaimana kemampuan sistem dalam mengklasifikasi *tweet* pelecehan *online* menggunakan *Gated Recurrent Unit*. Pada pengujian ini, digunakan data yang belum dikenali oleh model sebanyak 1402 data. Berikut hasil mode *testing* yang dapat dilihat pada tabel 4.2.

Tabel 4.2 Hasil Pengujian Sistem

<i>Tweet</i>	<i>Preprocessing</i>	<i>Class (Manual)</i>	Hasil Prediksi (<i>By system</i>)
udh pintar bhs cina ya km	pintar bahasa cina	Netral	Netral
@yunibadim Bangga jadi ga perawan ? Atau bangga jadi lonte berkelas semi pelacur ? ??	bangga tidak perawan bangga lonte kelas semi lacur	Sexually_harrassed	Racist
@bs_vito Anjing dasar..!!! Gw paham karena loe bukan pribumi onta yaman..!!!	anjing dasar paham bukan pribumi unta yaman	Racist Embarrass	Racist Embarrass
@ningenshikkakx Awat aja si dosky, gw bunuh lu besok	awas dosky bunuh besok	Physical_threatened	Physical_threatened
@_dityaaaa @_t0pher Emang goblok anak satu ini, anjing anjing	goblok anak anjing anjing	Embarrass	Embarrass

Tabel 4.2 Hasil Pengujian Sistem (Lanjutan)

<i>Tweet</i>	<i>Preprocessing</i>	<i>Class (Manual)</i>	<i>Hasil Prediksi (By system)</i>
@_mastito Sodok dari blakang sambil tamparin pantat kamu enak nih kyaknya hehe	sodok blakang tampar pantat enak nih	Sexually_harrassed	Sexually_harrassed
@__calendulaa @indriiiee_ @hachikitchen_ @infotwitwor_ itu bahasa halus kah? kok lucu dah namanya wk kalo pelacur, jalang, lonte kan kedengeran nya kasar banget, purel malah kek nickname wkwkwk	bahasa halus lucu nama lacur jalang lonte dengar kasar banget purel nickname	Netral	Netral
-	-	-	-
@_susuc0klat_ @Xutegr1__ @_susuc0klat_ goblok goblok goblok Makanya jgn nama doang susu coklat tapi otak gak dikasih gizi Beli susu sana	goblok goblok goblok jangan nama susu coklat otak tidak kasih gizi beli susu	Embarrass	Sexually_harrassed

Tabel 4.2 Hasil Pengujian Sistem (Lanjutan)

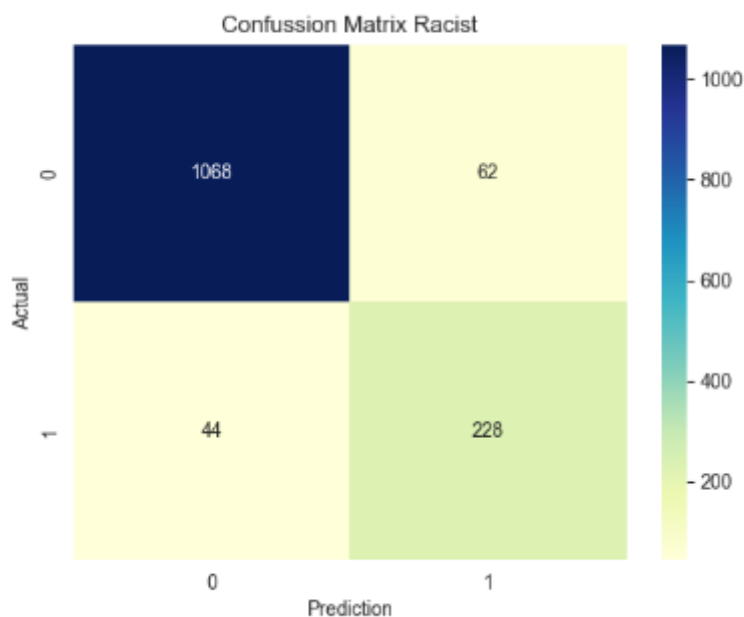
<i>Tweet</i>	<i>Preprocessing</i>	<i>Class (Manual)</i>	<i>Hasil Prediksi (By system)</i>
@fvckdjkk GW EMOSI SMA LU GW BUNUH LU BESOK AWAS AJA LU	emosi bunuh besok awas	Physical_threatened	Physical_threatened
@a_rogansi Dasar kadrun buzzer penjilat kotoran anus boosedan goblog, gak pernah makan sekolahan. Gak tahu jumlah penduduk pribumi Indonesia. Habis otaknya sudah kecampur kotoran anus yohanies.??	dasar kadal gurun buzzer jilat kotor anus boosedan goblok tidak makan sekolah tidak duduk pribumi indonesia habis otak campur kotor anus yohanies	Racist Embarrass	Racist Embarrass
@aizzahraa aku mau remes pantat kamu boleh ga cantik	remes pantat tidak cantik	Sexually_harrassed	Sexually_harrassed
@pnerahu spill kegiatan yg mengasah otak agar lebih pintar dong suhu	spill giat asah otak pintar suhu	Netral	Netral

Tabel 4.2 Hasil Pengujian Sistem (Lanjutan)

<i>Tweet</i>	<i>Preprocessing</i>	<i>Class (Manual)</i>	Hasil Prediksi (<i>By system</i>)
@arragaki @pietrasta ngetik pake ava Rokuta akan gw labrak lu sampai pipis dicelana	etik pakai ava rokuta labrak pipis celana	Embarrass	Sexually_harrassed
Tidak putih bukan berarti tidak cantik, artis juga tidak harus putih . Kenapa Jodie Di Bully segala? Warna kulit negara kita pasti beda sama orang korea, cina, jepang	tidak putih bukan tidak cantik artis tidak putih jodie bully warna kulit negara beda orang korea cina jepang	Netral	Netral

4.3. Metode Evaluasi

Pada bagian ini akan dilakukan evaluasi model dengan cara menghitung nilai akurasi menggunakan *accuracy matrix*. Oleh karena itu, dibutuhkan jumlah data yang benar dan salah saat diprediksi oleh model. Adapun metode evaluasi yang digunakan pada penelitian ini yaitu *confussion matrix* yang dimana *output* yang akan ditampilkan berupa nilai *precision*, *f1-score*, *recall* dan nilai *accuracy*. Berikut hasil yang akan ditampilkan dalam bentuk *heatmap* yang dapat dilihat pada Gambar 4.12 – 4.16.

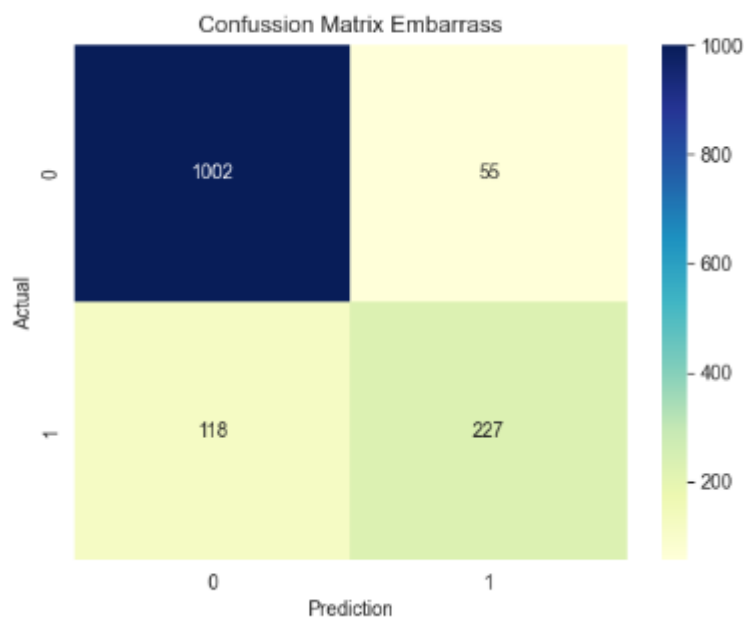


Gambar 4.12 Confussion matrix racist

Pada gambar 4.12 merupakan *confussion matrix* pada kelas rasis. Perhitungan yang dilakukan yaitu menunjukkan jumlah data yang diprediksi benar termasuk kelas rasis sebanyak 228 data dan data yang diprediksi salah atau bukan termasuk kelas rasis sebanyak 44 data. Selanjutnya, jumlah data yang diprediksi benar bukan termasuk kelas rasis sebanyak 1068 dan jumlah data yang diprediksi salah atau termasuk kelas rasis sebanyak 62 data. Berikut informasi mengenai rincian kelas rasis pada *confussion matrix*, dapat dilihat dalam tabel 4.3.

Tabel 4.3 Keterangan confussion matrix pada kelas rasis

No	Kelas Rasis	Jumlah
1	TPT (<i>True Positive Tweet</i>)	228
2	FPT (<i>False Positive Tweet</i>)	62
3	TNT (<i>True Negative Tweet</i>)	1068
4	FNT (<i>False Negative Tweet</i>)	44

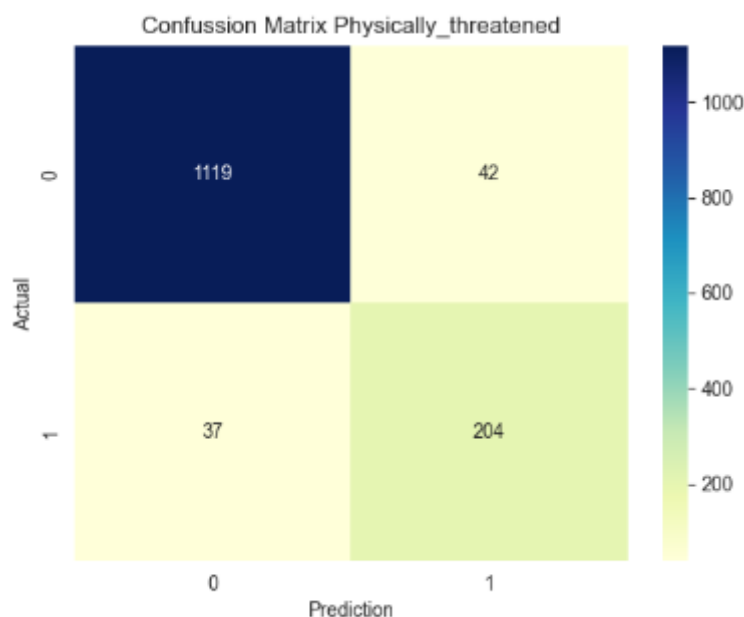


Gambar 4.13 Confussion matrix embarrass

Pada gambar 4.13 merupakan *confussion matrix* pada kelas *embarrass*. Perhitungan yang dilakukan yaitu menunjukkan jumlah data yang diprediksi benar termasuk kelas *embarrass* sebanyak 227 data dan jumlah data diprediksi salah atau bukan termasuk kelas *embarrass* sebanyak 118 data. Selanjutnya, jumlah data yang diprediksi benar bukan termasuk kelas *embarrass* sebanyak 1002 dan jumlah data yang diprediksi salah atau termasuk kelas *embarrass* sebanyak 55 data. Berikut informasi mengenai rincian kelas *embarras* pada *confussion matrix*, dapat dilihat dalam tabel 4.4.

Tabel 4.4 Keterangan confussion matrix pada kelas embarrass

No	Kelas Embarrass	Jumlah
1	TPT (<i>True Positive Tweet</i>)	227
2	FPT (<i>False Positive Tweet</i>)	56
3	TNT (<i>True Negative Tweet</i>)	1002
4	FNT (<i>False Negative Tweet</i>)	118

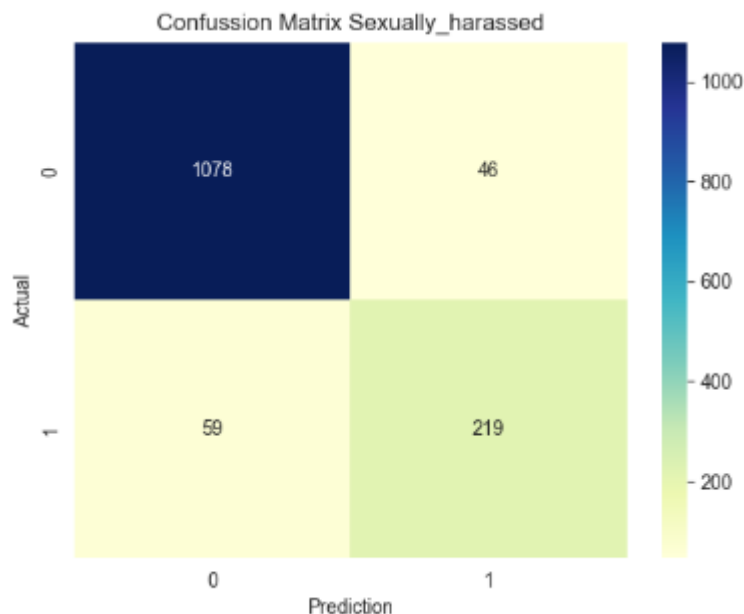


Gambar 4.14 Confussion matrix physically threatened

Pada gambar 4.14 merupakan *confussion matrix* pada kelas *physically threatened*. Perhitungan yang dilakukan yaitu menunjukkan jumlah data yang diprediksi benar termasuk kelas *physically threatened* sebanyak 204 data dan jumlah data diprediksi salah atau bukan termasuk kelas *physically threatened* sebanyak 37 data. Selanjutnya, jumlah data yang diprediksi benar bukan termasuk kelas *physically threatened* sebanyak 1119 serta jumlah data yang diprediksi salah atau termasuk kelas *physically threatened* sebanyak 42 data. Berikut informasi mengenai rincian kelas *physically threatened* pada *confussion matrix*, dapat dilihat dalam tabel 4.5.

Tabel 4.5 Keterangan confusion matrix pada kelas physically threatened

No	Kelas Physically Threatened	Jumlah
1	TPT (<i>True Positive Tweet</i>)	204
2	FPT (<i>False Positive Tweet</i>)	42
3	TNT (<i>True Negative Tweet</i>)	1119
4	FNT (<i>False Negative Tweet</i>)	37

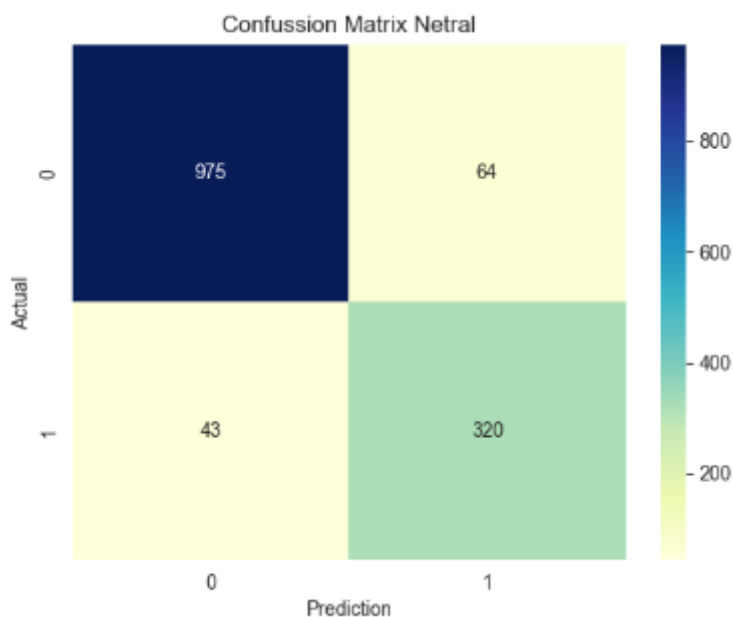


Gambar 4.15 Confussion matrix sexually harassed

Pada gambar 4.15 merupakan *confussion matrix* pada kelas *sexually harassed*. Perhitungan yang dilakukan yaitu menunjukkan jumlah data yang diprediksi benar termasuk kelas *sexually harassed* sebanyak 219 data dan jumlah data diprediksi salah atau bukan termasuk kelas *sexually harassed* sebanyak 59 data. Selanjutnya, jumlah data yang diprediksi benar bukan termasuk kelas *sexually harassed* sebanyak 1078 serta jumlah data yang diprediksi salah atau termasuk kelas *sexually harassed* sebanyak 46 data. Berikut informasi mengenai rincian kelas *sexually harassed* pada *confussion matrix*, dapat dilihat dalam tabel 4.6.

Tabel 4.6 Keterangan confusion matrix pada kelas sexually harassed

No	Kelas Sexually Harassed	Jumlah
1	TPT (<i>True Positive Tweet</i>)	219
2	FPT (<i>False Positive Tweet</i>)	46
3	TNT (<i>True Negative Tweet</i>)	1078
4	FNT (<i>False Negative Tweet</i>)	59



Gambar 4.16 Confussion Matrix Netral

Pada gambar 4.16 merupakan *confussion matrix* pada kelas netral. Perhitungan yang dilakukan yaitu menunjukkan jumlah data yang diprediksi benar termasuk kelas netral sebanyak 320 data dan jumlah data diprediksi salah atau bukan termasuk kelas netral sebanyak 43 data. Selanjutnya, jumlah data yang diprediksi benar bukan termasuk kelas netral sebanyak 975 serta jumlah data yang diprediksi salah atau termasuk kelas netral sebanyak 64 data. Berikut informasi mengenai rincian kelas netral pada *confussion matrix*, dapat dilihat dalam tabel 4.7.

Tabel 4.7 Keterangan *confussion matrix* pada kelas netral

No	Kelas Netral	Jumlah
1	TPT (<i>True Positive Tweet</i>)	320
2	FPT (<i>False Positive Tweet</i>)	64
3	TNT (<i>True Negative Tweet</i>)	975
4	FNT (<i>False Negative Tweet</i>)	43

Berdasarkan tabel informasi mengenai rincian masing-masing kelas pada *confussion matrix*, selanjutnya akan dilakukan evaluasi dengan menghitung *classification report* sebagai berikut:

Kalkulasi *Precision* :

$$Precision \text{ tweet kelas racist} = \frac{TPT}{TPT+FPT} = \frac{228}{228+62} = 0.78$$

$$Precision \text{ tweet kelas embarrass} = \frac{TPT}{TPT+FPT} = \frac{227}{227+56} = 0.80$$

$$Precision \text{ tweet kelas physical threatened} = \frac{TPT}{TPT+FPT} = \frac{204}{204+42} = 0.83$$

$$Precision \text{ tweet kelas sexually harrassed} = \frac{TPT}{TPT+FPT} = \frac{219}{219+46} = 0.83$$

$$Precision \text{ tweet kelas netral} = \frac{TPT}{TPT+FPT} = \frac{320}{320+64} = 0.83$$

Kalkulasi *Recall* :

$$Recall \text{ tweet kelas racist} = \frac{TPT}{TPT+FNT} = \frac{228}{228+44} = 0.84$$

$$Recall \text{ tweet kelas embarrass} = \frac{TPT}{TPT+FNT} = \frac{227}{227+118} = 0.66$$

$$Recall \text{ tweet kelas physical threatened} = \frac{TPT}{TPT+FNT} = \frac{204}{204+37} = 0.85$$

$$Recall \text{ tweet kelas sexually harrassed} = \frac{TPT}{TPT+FNT} = \frac{219}{219+59} = 0.79$$

$$Recall \text{ tweet kelas netral} = \frac{TPT}{TPT+FNT} = \frac{320}{320+43} = 0.88$$

Kalkulasi *F1-Score* :

$$F1\text{-Score tweet kelas racist} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{2 \times 0.84 \times 0.78}{0.84 + 0.78} = 0.81$$

$$F1\text{-Score tweet kelas embarrass} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{2 \times 0.66 \times 0.80}{0.66 + 0.80} = 0.72$$

$$F1\text{-Score tweet kelas physical threatened} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{2 \times 0.85 \times 0.83}{0.85 + 0.83} = 0.84$$

$$F1\text{-Score tweet kelas sexually harrassed} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{2 \times 0.79 \times 0.83}{0.79 + 0.83} = 0.81$$

$$F1\text{-Score tweet kelas netral} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{2 \times 0.88 \times 0.83}{0.88 + 0.83} = 0.86$$

Tabel 4.8 Evaluasi

Jenis klasifikasi	Precision	Recall	F1-Score
Tweet Racist	0.79	0.84	0.81
Tweet Embarrass	0.8	0.66	0.72
Tweet Physically Threatened	0.83	0.85	0.84
Tweet Sexually Harrassed	0.83	0.79	0.81
Tweet Netral	0.83	0.88	0.86
Average	0.82	0.80	0.80

$$Accuracy = \frac{\text{Jumlah Data Yang Benar}}{\text{Jumlah Total Semua Data}}$$

$$Accuracy = \frac{(227+204+219+320)+(1002+1119+1078+975)}{(227+204+219+320)+(1002+1119+1078+975)+(56+42+46+64)+(118+37+59+43)}$$

$$Accuracy = \frac{970+4174}{970+4174+208+257}$$

$$Accuracy = 91.7\%$$

Berdasarkan hasil pengujian dalam mengklasifikasi *tweet* pelecehan *online* pada media sosial X menggunakan *Gated Recurrent Unit* ditemukan bahwa nilai akurasi yang diperoleh sebesar 91.7%. Dari nilai akurasi tersebut dapat disimpulkan bahwa model menggunakan *Gated Recurrent Unit* dapat melakukan klasifikasi *tweet* pelecehan *online* pada media sosial X dengan cukup baik.

Akan tetapi masih terdapat ketidaktepatan model dalam melakukan klasifikasi pada data dikarenakan beberapa faktor. Penulis telah menganalisa beberapa faktor yang memungkinkan model melakukan kesalahan dalam prediksi, sebagai berikut :

1. Terdapat data yang dimana kata-katanya tidak terlalu dominan atau tidak terlalu spesifik sehingga model masih belum mampu dalam memprediksi data.

Contoh :

- Tidak semua kata kafir ditujukan bg agama lain, cb kaji lg tu..”
- @_WHO4MII @DeHoutman_id Klo kafir ga akan mengerti.. Udah gitu aja...

(Pada contoh diatas, sistem memprediksi bahwa tweet tersebut tidak masuk kedalam salah satu kelas karena tidak adanya kata spesifik mendukung kata kunci “kafir” yang dapat mengacu kepada kelas tertentu)

2. Salah satu contoh ketika melakukan testing untuk kelas sexually_harrassed dan ternyata hasil prediksi adalah kelas embarrass, hal ini disebabkan oleh ciri-ciri data yang dimiliki kedua kelas ini pada dataset saling sangkut paut dan tidak ada kata yang mencolok, ditambah lagi jumlah data tidak sebanyak dominan kedua kelas, sehingga dihitung berdasarkan probabilitasnya.

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil dari pengujian yang sudah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut :

1. Metode Gated Recurrent Unit dapat diterapkan untuk klasifikasi tweet pelecehan *online* pada media sosial X dengan memprediksi tweet ke dalam 5 kelas yaitu *racist*, *embarrass*, *physical threatened*, *sexually harassed* dan *neutral* serta memiliki performa yang cukup baik dan efektif sebagai sistem pendukung dalam mengotomatiskan proses klasifikasi pelecehan *online* yang terjadi pada media sosial X
2. Pada keseluruhan tahapan pengujian yang sudah dilakukan dalam mengklasifikasi *tweet* pelecehan *online* pada media sosial X menggunakan Gated Recurrent Unit diperoleh tingkat akurasi sebesar 91.7% dengan *loss* sebesar 2.1% .

5.2. Saran

Penulis memberikan beberapa saran untuk pengembangan sistem dalam penelitian ini yaitu sebagai berikut :

1. Selain melakukan klasifikasi *tweet* pelecehan *online* menggunakan bahasa indonesia, perlu dilakukan penelitian lebih lanjut untuk dapat memproses tweet berbahasa inggris serta dapat dikembangkan dengan pembacaan tanda baca dan *emoticon* sehingga bukan hanya berdasarkan kata saja.
2. Pada penelitian selanjutnya dapat menambahkan lebih banyak dataset yang lebih beragam agar hasil *processing* lebih maksimal dan memiliki keakuratan yang lebih baik saat memprediksi data.

DAFTAR PUSTAKA

- Anna, S., Utama, A. P., Setiawan, B., Widodo, P., Saragih, H. J., & Sukendro, A. (2023). Resolusi Konflik Terhadap Kasus Kekerasan Berbasis Gender Online (KBGO) di Indonesia Pada Masa Pandemi Covid 2019. *Jurnal Kewarganegaraan*, 352-362.
- Arfianti, U. I., Novitasari, D. C., Widodo, N., Hafiyusholeh, M., & Utami, W. D. (2021). Sunspot Number Prediction Using Gated Recurrent Unit (GRU) Algorithm. *Indonesian Journal of Computing and Cybernetics Systems*, 141-152.
- Biswas, S., Chada, E., & Ahmad, F. (2015). Sentiment Analysis with Gated Recurrent Units. *Advances in Computer Science and Information Technology (ACSIT)*.
- Cahyono, A. S. (n.d.). Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat di Indonesia. 140-157.
- Duggan, M. (2017). Online Harassment 2017. *Pew Research Center*, 1-85.
- Halim, J. K., Herwindiati, D. E., & Hendryli, J. (2023). Penerapan Gated Recurrent Unit Untuk Prediksi Zat Pencemar Udara. *Jurnal Ilmu Komputer dan Sistem Informasi*.
- Hutabarat, F. T. (2021). *Identifikasi Judul Clickbait Pada Berita Bahasa Indonesia Menggunakan Gated Recurrent Unit*. Medan.
- Kurnia, S., Prabowo, H. G., Suharmanto, B., Pribadi, M. N., & Chamidah, N. (2022). Klasifikasi Tweet Cyberbullying dengan Menggunakan Algoritma Random Forest. *SENAMIKA*, e-ISSN 2962-6129.
- Li, W., Wu, H., Zhu, N., Jiang, Y., Tan, J., & Guo, Y. (2020). Prediction of Dissolved Oxygen in A Fishery Pond Based on Gated Recurrent Unit (GRU). www.sciencedirect.com.
- Naufal, R. A. (2023). Analisis Sentimen Terhadap Cyberbullying Di Media Sosial Dengan Crowdtangle. *Skripsi*.
- Owen, M., Vincent, Br Ambarita, R., & Indra, E. (2022). Implementasi Metode Long Short Term Memory Untuk Memprediksi Pergerakan Nilai Harga Emas. *Jurnal TEKINKOM*, 96-104.
- PEREMPUAN, K. (2023). *Catatan Tahunan Komnas Perempuan Tahun 2023 Kekerasan terhadap Perempuan di Ranah Publik dan Negara:Minimnya Pelindungan dan Pemulihan*. Jakarta: Komisi Nasional Anti Kekerasan Terhadap Perempuan.

- Pratomo, Y. (2019). *49 Persen Netizen di Indonesia Pernah Mengalami "Bullying" di Medsos*. Tekno Kompas.
- Prayoga, W. (2020). *Perancangan Informasi Edukasi Tentang Rasisme Melalui Media Komik STRIP*. Bandung.
- Qurtuby, S. A. (2017). *Sentimen Anti-Cina di Indonesia*.
<https://www.dw.com/id/sentimen-anti-cina-di-indonesia/a-36974659>.
- Rizki, M., Basuki, S., & Azhar, Y. (2020). Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang. *Repositor*, 331-338.
- Sakban, A. (2020). *Pencegahan Cyberbullying di Indonesia*. Mataram: Research Gate.
- Siagian, O. W. (2019). Resolution of Cyber Harassment Against Public Figures in . *Jurnal Legalitas*, Vol .14 No.2 Page 40.
- Silitonga, D. C. (2022). *Klasifikasi Jenis UU ITE Menggunakan Random Forest Pada Kasus Kejahatan Siber*. Medan.
- Statista. (2020). *Leading countries based on number of Twitter users as of July 2020*.
<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>.
- Tarwani, K., & Edem, S. (2017). Survey on Recurrent Neural Network in Natural Language Processing. *International Journal of Engineering Trends and Technology (IJETT)*, Volume: 48.
- Thige, O. (2023). The Combined Effects of Online Harassment on Social Media Users. *Thesis*.
- Tri Mueri Sandes. (2018). *Klasifikasi Tweet Pelecehan Online Pada Twitter Dengan Menggunakan Naive Bayes*. PEKANBARU: 2018.
- Triyana, Resa; Putra, Oddy Virgantara; Pradhana, Faisal Reza. (2022). Deteksi Cyberbullying Pada Tweet Berbahasa Inggris Dengan Menggunakan Metode Support Vector Machine. *SENDIKO*.
- Wardana, R. P. (2020). *Penerapan Model Gated Recurrent Unit Untuk Peramalan Jumlah Penumpang Kereta Api di PT. KAI (Persero)*. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.
- Wildan Putra Aldi, M., & Aditsania, A. (2018). Analisis dan Implementasi Long Short Term Analisis dan Implementasi Long Short Term.
- Yulianto, A. D. (n.d.). *Implementasi Deep Learning Dengan Menggunakan Pemodelan Gated Recurrent Unit (GRU) Untuk Prediksi Harga Saham di Index Saham Syariah Indonesia*. Semarang: Repositori UNIMUS.



KEPUTUSAN
DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI
NOMOR : 2365/UN5.2.14.D/SK/SPB/2024

DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA

- Membaca : Surat Permohonan Mahasiswa Fasilkom-TI USU tanggal 3 Juli 2024 perihal permohonan ujian skripsi:
Nama : INDAH MARIANA NAINGGOLAN
NIM : 191402024
Program Studi : Sarjana (S-1) Teknologi Informasi
Judul Skripsi : Klasifikasi Tweet Pelecehan Online Pada Media Sosial X Menggunakan Gated Recurrent Unit
- Memperhatikan : Bahwa Mahasiswa tersebut telah memenuhi kewajiban untuk ikut dalam pelaksanaan Meja Hijau Skripsi Mahasiswa pada Program Studi Sarjana (S-1) Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara TA 2023/2024.
- Menimbang : Bahwa permohonan tersebut diatas dapat disetujui dan perlu ditetapkan dengan surat keputusan
- Mengingat : 1. Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.
2. Peraturan Pemerintah Nomor 17 tahun 2010 tentang pengelolaan dan penyelenggara pendidikan.
3. Keputusan Rektor USU Nomor 03/UN5.1.R/SK/SPB/2021 tentang Peraturan Akademik Program Sarjana Universitas Sumatera Utara.
4. Surat Keputusan Rektor USU Nomor 1876/UN5.1.R/SK/SDM/2021 tentang pengangkatan Dekan Fasilkom-TI USU Periode 2021-2026

MEMUTUSKAN

- Menetapkan :
Pertama : Membentuk dan mengangkat Tim Penguji Skripsi mahasiswa sebagai berikut:
Ketua : Ulfi Andayani S.Kom., M.Kom
NIP: 198604192015042004
Sekretaris : Fanindia Purnamasari S.TI,M.IT
NIP: 198908172019032023
Anggota Penguji : Sarah Purnamawati ST., MSc.
NIP: 198302262010122003
Anggota Penguji : Umaya Ramadhani Putri Nasution S.TI, M.Kom.
NIP: 199104112021022001
Moderator : -
Panitera : -
- Kedua : Segala biaya yang diperlukan untuk pelaksanaan kegiatan ini dibebankan pada Dana Penerimaan Bukan Pajak (PNPB) Fasilkom-TI USU Tahun 2024.
- Ketiga : Keputusan ini berlaku sejak tanggal ditetapkan dengan ketentuan bahwa segala sesuatunya akan diperbaiki sebagaimana mestinya apabila dikemudian hari terdapat kekeliruan dalam surat keputusan ini.

- Tembusan :
- 1. Ketua Program Studi Sarjana (S-1) Teknologi Informasi
 - 2. Yang bersangkutan
 - 3. Arsip

Medan, 04 Juli 2024
Ditandatangani secara elektronik oleh:
Dekan



Maya Silvi Lydia
NIP 197401272002122001