

**KLASIFIKASI UJARAN KEBENCIAN DI *TWITTER*
MENGUNAKAN ALGORITMA *RADIAL BASIS FUNCTION*
NETWORK (RBFN)**

SKRIPSI

OLEH:

DANI GUNAWAN SITOMPUL

188160010



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2025

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 16/3/26

Access From (repository.uma.ac.id)16/3/26

**KLASIFIKASI UJARAN KEBENCIAN DI TWITTER
MENGUNAKAN ALGORITMA *RADIAL BASIS FUNCTION*
NETWORK (RBFN)**

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana di Fakultas Teknik
Universitas Medan Area



Oleh :

DANI GUNAWAN SITOMPUL

188160010

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN**

2025

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

HALAMAN PENGESAHAN

Judul Skripsi : *Klasifikasi Ujaran Kebencian di Twitter Menggunakan Algoritma Radial Basis Function Network (RBFN)*
Nama : Dani Gunawan Sitompul
NPM : 188160010
Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing

Dr. Arnes Sembiring, ST, M.Kom
Dosen Pembimbing



Tanggal Lulus: 30 Juni 2025

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi – sanksi lainnya dengan peraturan yang berlaku, apabila dikemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 30 Juni 2025



Dani Gunawan Sitompul
188160010

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Dani Gunawan Sitompul
NPM : 188160010
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul : **Klasifikasi Ujaran Kebencian di Twitter Menggunakan Algoritma Radial Basis Function Network (RBFN)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan
Pada Tanggal : 30 Juni 2025
Yang menyatakan



(Dani Gunawan Sitompul)

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan ujaran kebencian di media sosial *Twitter* menggunakan algoritma *Radial Basis Function Network* (RBFN). Latar belakang penelitian ini didasarkan pada maraknya penyebaran ujaran kebencian di media sosial yang dapat memicu konflik sosial dan mengganggu keharmonisan masyarakat. Data penelitian diperoleh secara langsung dari *Twitter* melalui API V2 dengan kata kunci “ujaran kebencian”, berjumlah 600 *tweet* berbahasa Indonesia. Tahapan penelitian meliputi *text preprocessing* yang terdiri dari *case folding*, *tokenization*, *stopwords removal*, *stemming*, serta pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Model RBFN digunakan sebagai metode klasifikasi untuk membedakan antara *tweet* yang mengandung ujaran kebencian (*hate speech*) dan yang tidak (*non-hate speech*). Evaluasi dilakukan menggunakan *confusion matrix* dengan metrik akurasi, presisi, dan *recall* pada tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Hasil pengujian menunjukkan bahwa model memperoleh akurasi tertinggi sebesar 87,7% pada pembagian data 70% untuk pelatihan dan 30% untuk pengujian. Hasil ini menunjukkan bahwa algoritma RBFN mampu memberikan performa yang baik dalam mengklasifikasikan ujaran kebencian di *Twitter*. Penelitian ini diharapkan dapat menjadi acuan dalam pengembangan sistem deteksi ujaran kebencian secara otomatis untuk menciptakan lingkungan media sosial yang lebih sehat dan aman.

Kata Kunci: Ujaran Kebencian, *Twitter*, *Text Preprocessing*, *Radial Basis Function Network* (RBFN), TF-IDF, Klasifikasi.

ABSTRACT

This research aims to classify hate speech on the social media platform Twitter using the Radial Basis Function Network (RBFN) algorithm. The background of this study stems from the widespread dissemination of hate speech on social media, which can trigger social conflicts and disrupt public harmony. The dataset was collected directly from Twitter through API V2 using the keyword “ujaran kebencian,” consisting of 600 Indonesian-language tweets. The research stages include text preprocessing, which involves case folding, tokenization, stopwords removal, stemming, and word weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The RBFN model was applied to classify tweets into hate speech and non-hate speech categories. Model evaluation was conducted using a confusion matrix with accuracy, precision, and recall metrics across three data-splitting scenarios: 70:30, 80:20, and 90:10. The results show that the model achieved the highest accuracy of 87.7% under the 70:30 data-splitting configuration. These findings indicate that the RBFN algorithm performs effectively in classifying hate speech on Twitter. This study is expected to serve as a reference for developing automated hate speech detection systems to promote a safer and more positive online environment.

Keywords: Hate Speech, Twitter, Text Preprocessing, Radial Basis Function Network (RBFN), TF-IDF, Classification.

RIWAYAT HIDUP

Penulis dilahirkan di Medan pada tanggal 2 Januari 2000 dari Bapak Rudolf Mangasa Sitompul dan mendiang ibu Junita Magdalena Mare Mare. Penulis merupakan anak ke dua dari 4 bersaudara dan memiliki satu kakak perempuan dan dua adik perempuan. Penulis pertama kali mengenyam pendidikan di bangku SD Swasta Santo Paulus Martubung Medan pada tahun 2006 dan lulus pada tahun 2012. kemudian penulis melanjutkan pendidikan ke jenjang SMP pada tahun 2012 di SMP Swasta Santo Paulus Martubung Medan dan lulus pada tahun 2015. Pada tahun yang sama, Penulis melanjutkan ke jenjang selanjutnya yaitu di SMA Santo Paulus Martubung Medan dan lulus pada tahun 2018. Pada bulan Juli tahun 2018, Penulis melanjutkan pendidikan di bangku kuliah dan terdaftar sebagai mahasiswa Fakultas Teknik dengan mengambil Program Studi Teknik Informatika di Universitas Medan Area.



KATA PENGANTAR

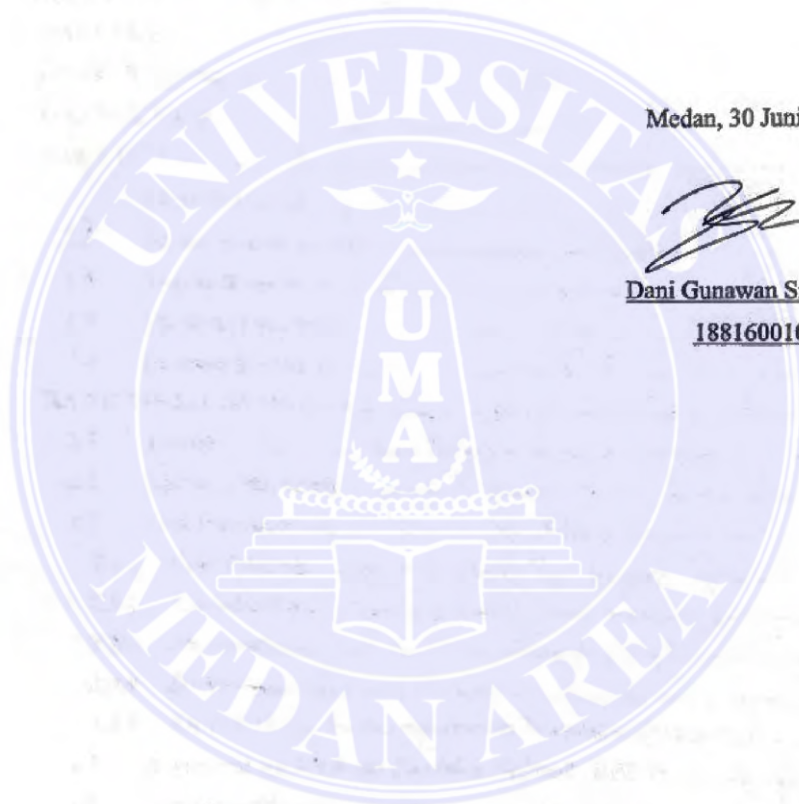
Puji syukur penulis panjatkan kepada Tuhan Yang Maha Kuasa atas segala karuniaNya sehingga skripsi ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini dengan judul “**Klasifikasi Ujaran Kebencian di *Twitter* Menggunakan Algoritma *Radial Basis Function Network (RBFN)***”. Skripsi ini merupakan salah satu syarat untuk menyelesaikan pendidikan untuk mencapai gelar sarjana di Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area. Pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc. selaku Rektor Universitas Medan Area.
2. Bapak Dr. Eng. Supriatno, ST., M.T selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom selaku Kepala Program Studi Teknik Informatika Universitas Medan Area.
4. Bapak Dr. Arnes Sembiring, ST, M.Kom selaku Dosen pembimbing yang telah membantu dan membimbing sehingga penulis dapat menyelesaikan skripsi ini.
5. Kedua Orang tua penulis tercinta yang tiada henti memberikan semua dukungan dan doa terbaik untuk penulis.
6. Seluruh Dosen dan Staf Program Studi Teknik Informatika Universitas Medan Area.
7. Seluruh teman-teman yang sudah memberikan dukungannya selama penulisan skripsi ini, khususnya teman-teman Teknik Informatika angkatan 2018

Penulis menyadari bahwa penelitian ini masih memiliki kekurangan oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan demi kesempurnaan penelitian ini. Penulis berharap tugas penelitian ini dapat bermanfaat baik kalangan pendidikan maupun masyarakat. Akhir kata penulis ucapkan terima kasih.

Medan, 30 Juni 2025


Dani Gunawan Sitompul
188160010



DAFTAR ISI

ABSTRAK	iv
ABSTRACT	v
RIWAYAT HIDUP	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA	6
2.1 <i>Twitter</i>	6
2.2 Ujaran Kebencian	6
2.3 <i>Text Preprocessing</i>	7
2.3.1 <i>Case Folding</i>	7
2.3.2 <i>Tokenization</i>	7
2.3.3 <i>Stopwords</i>	7
2.3.4 <i>Stemming</i>	8
2.3.5 <i>TF-IDF (Term Frequency-Inverse Document Frequency)</i>	8
2.4 <i>Algoritma Radial Basis Function Network (RBFN)</i>	9
2.5 <i>Confusion Matrix</i>	12
2.6 Penelitian Terdahulu	14

BAB III METODE PENELITIAN	17
3.1 Pengumpulan Data	18
3.2 <i>Text Preprocessing</i>	19
3.2.1 <i>Case Folding</i>	20
3.2.2 <i>Tokenization</i>	20
3.2.3 <i>Stopwords</i>	21
3.2.4 <i>Stemming</i>	21
3.2.5 <i>Term-Frequency Inverse Document Frequency (TF-IDF)</i>	22
3.3 Perhitungan Manual <i>Radial Basis Function Neural Network</i>	23
3.4 <i>Confusion Matrix</i>	27
BAB IV HASIL DAN PEMBAHASAN.....	29
4.1 Hasil	29
4.2 Pembahasan.....	37
BAB V KESIMPULAN DAN SARAN	41
5.1 Kesimpulan	41
5.2 Saran.....	41
DAFTAR PUSTAKA.....	42

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	17
Gambar 3. 2 Developer Portal Twitter	18
Gambar 3. 3 Crawling Data	19
Gambar 3. 4 Alur Text Preprocessing.....	20
Gambar 3. 5 Alur Algoritma RBFN.....	23
Gambar 4. 1 Developer Portal X.....	29
Gambar 4. 2 Crawling Data	30
Gambar 4. 3 Case Folding.....	30
Gambar 4. 4 Tokenization.....	31
Gambar 4. 5 Stopwords.....	31
Gambar 4. 6 Stemming	32
Gambar 4. 7 TF-IDF	33
Gambar 4. 8 Pembagian Data & Algoritma RBFNN.....	34
Gambar 4. 9 Confusion Matrix Pengujian 1	35
Gambar 4. 10 Confusion Matrix Pengujian 2	36
Gambar 4. 11 Confusion Matrix Pengujian 3	37

DAFTAR TABEL

Tabel 2. 1 Cunfusion Matrix	13
Tabel 3. 1 Tahapan Case Folding.....	20
Tabel 3. 2 Tahapan Tokenization.....	20
Tabel 3. 3 Tahapan Stopwords.....	21
Tabel 3. 4 Tahapan Stemming	22
Tabel 3. 5 Tahapan TF-IDF	22
Tabel 3. 6 Hasil Klasifikasi Menggunakan Algoritma RBFN	27
Tabel 3. 7 Hasil Confusion Matrix.....	27
Tabel 4. 1 Hasil Pengujian	34



BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial sebagai ruang publik berperan sebagai media untuk berdiskusi, bertukar pikiran, serta berkomunikasi secara bebas dan demokratis. Saat ini peran tersebut mulai terganggu dengan adanya intervensi kekuatan ekonomi politik oleh kelompok tertentu, sehingga netralitas masyarakat dalam bermedia terusik. Media sosial yang selayaknya berguna untuk membangun kebebasan berpendapat dan solidaritas demokrasi, justru digunakan sebagai alat penyebaran ujaran kebencian untuk menggiring opini publik, bahkan menyerang lawan yang memiliki perbedaan pandangan dan ideologi. Persebaran ujaran kebencian tersebut dilakukan bukan sekedar oleh unsur ketidaksengajaan, bersifat iseng, atau hiburan semata, tetapi dapat diasumsikan memiliki agenda *setting* tertentu (Kusumasari & Arifianto, 2020). Ujaran kebencian pada media sosial telah bermetamorfosis menjadi bagian dari sebuah panggung ekonomi politik yang dengan sengaja dirancang oleh pihak tertentu untuk mendapatkan keuntungan dibalik wacana dan produksinya. Ujaran kebencian di media sosial telah menjadi ancaman bagi demokrasi kebebasan pers, sebagai ruang publik di Indonesia, selain meresahkan masyarakat pengguna informasi. Masyarakat menjadi kesulitan dalam membedakan antara teks informasi yang bersifat kritik di ruang publik dan teks ujaran kebencian, selain mengalami kesulitan dalam menentukan informasi mana yang bersifat akurat dan informasi mana yang merupakan berita palsu/hoax. Akibatnya masyarakat pengguna ruang publik di media sosial kerap mempercayai informasi yang mengandung ujaran kebencian tersebut dengan mengunggah ulang dan meneruskan pesan tersebut melalui media sosial terkait, sehingga berimplikasi pada penyebaran ujaran kebencian dengan cepat dan menjangkau audiens yang luas (Kusumasari & Arifianto, 2020). Media sosial merupakan sarana untuk berinteraksi orang-orang satu sama lain dengan cara menciptakan, berbagi, serta bertukar informasi dan gagasan melalui kata-kata, gambar, dan video dalam sebuah jaringan dan komunitas virtual (Saputra, 2019). Menurut Surat Edaran Mabes Polri

No:SE/6/X/2015, tanggal 8 Oktober 2015, ujaran kebencian di definisikan sebagai “tindak pidana yang berbentuk, penghinaan, pencemaran nama baik, penistaan, perbuatan yang tidak menyenangkan, memprovokasi, menghasut, penyebaran berita bohong, dimana semua tindakan di atas memiliki tujuan atau bisa berdampak pada tindak diskriminasi, kekerasan, penghilangan nyawa, dan atau konflik sosial” (Dharmawan dkk., 2023). Ujaran kebencian adalah ucapan dan atau tulisan yang dibuat seseorang di muka umum untuk tujuan menyebar atau menyulut kebencian sebuah kelompok terhadap kelompok lain yang berbeda baik karena ras, agama keyakinan, gender, etnisitas, kecacatan, dan orientasi seksual (Hakiem dkk., 2019). Permasalahannya, keleluasaan berdiskusi di media sosial ini menyiratkan beberapa dampak negatif. Salah satu yang dipotret ialah hadir dan meningkatnya intensitas ujaran kebencian. Semakin hari semakin banyak kasus-kasus ujaran kebencian di dunia maya sehingga menyebabkan perpecahan di negeri ini. Hal tersebut terjadi dikarenakan kurangnya menjaga perkataan keji, mencaci maki lidah yang kotor, kemudian bencana kata-kata kutukan baik kepada binatang benda atau manusia, penghinaan dan kata-kata ejekan, mengumpat, adu domba dll (Safitri, 2020). Ujaran kebencian tersebut dapat menyebabkan terjadinya perpecahan antargolongan dan memecah belah persatuan bangsa Indonesia sehingga harus dapat ditangani dengan cepat sebelum berkembang menjadi sangat besar dan luas. Oleh karena itu klasifikasi ujaran kebencian diperlukan untuk mengurangi tersebarnya *tweet - tweet* yang berisikan ujaran kebencian tersebut. Klasifikasi ujaran kebencian yang dilakukan secara manual akan memakan waktu dan juga tenaga yang banyak, maka dari itu klasifikasi ujaran kebencian secara otomatis diperlukan untuk mempermudah proses klasifikasi tersebut sehingga kedepannya dapat digunakan oleh pihak yang membutuhkan seperti misalnya Kepolisian Negara Republik Indonesia (Hakiem dkk., 2019).

Banyak peneliti yang melakukan pendeteksian ujaran kebencian dalam beberapa tahun terakhir untuk mengatasi masalah klasifikasi ujaran dengan menggunakan berbagai algoritma seperti (Susanti dkk., 2022) dengan judul

“Klasifikasi Data *Tweet* Ujaran Kebencian Di Media Sosial Menggunakan *Naïve Bayes Classifier*” menggunakan data yang bersumber dari *kaggle.com* dengan label positif dan negatif, hasilnya mendapatkan nilai akurasi 69%. Penelitian (Septiawan & Chairani, 2023) dengan judul “Perbandingan Akurasi Metode Deteksi Ujaran Kebencian dalam Postingan Twitter Menggunakan Metode SVM dan Decision Trees yang Dioptimalkan dengan Adaboost”, data yang digunakan berjumlah 22878 dengan 3 label yaitu positif, negatif, dan netral. Mendapatkan hasil akurasi 90,03% serta *Decision Tree* sebesar 89,53%.

Penelitian ini akan menggunakan data primer dengan menggunakan teknik API *Twitter*, adapun pembersihan data meliputi *case folding*, *tokenization*, *stopwords*, *stemming*, dan TF-IDF. Algoritma yang digunakan adalah *Radial Basis Function Network* (RBFN), yang merupakan pengembangan dari *feedforward neural network*. Jaringan RBF terdiri dari tiga lapisan, yaitu lapisan masukan (*input layer*), lapisan tersembunyi (*hidden layer*) dan lapisan keluaran (*output layer*) (Ambarwati, 2023). Setiap node pada *hidden layer* mengaktifkan fungsi aktivasi secara non-linier pada vektor masukan dengan fungsi radial basis. Pada *hidden layer* mentransformasikan *input* secara non-linier dan kemudian pada *output layer* menerima sinyal *output* dari *hidden layer* dan diproses secara linier (Ambarwati, 2023). RBFN memiliki struktur jaringan yang mudah namun kemampuan konvergensi yang sangat cepat serta kemampuan pendekatan non-linier yang sangat tinggi (Chen dkk., 2019). Alasan peneliti menggunakan algoritma RBFN ini karena merupakan arsitektur yang populer digunakan dalam klasifikasi. RBFN bersifat *feedforward*, tipe jaringan (*multilayer*) serta keunggulan fungsi aktivasi yang digunakan. Berdasarkan uraian tersebut, maka penelitian ini bertujuan menerapkan algoritma RBFN dalam mengklasifikasi ujaran kebencian di *Twitter*. Klasifikasi adalah proses untuk menemukan model atau fungsi yang dapat menggambarkan dan membedakan kelas data atau konsep, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas yang belum diketahui dari suatu objek pengamatan (Sihombing & Yuliati, 2021). Metode klasifikasi yang umum digunakan pada disiplin ilmu statistika adalah Analisis Diskriminan dan Regresi

Logistik. Namun, era data yang semakin populer menunjukkan pertumbuhan pesat dari volume data yang luar biasa banyak, sehingga menghasilkan set data besar (*big data*). Oleh karena itu, kita membutuhkan alat analisis yang kuat dan serbaguna untuk mengungkap informasi berharga dari set data besar dan mengubahnya menjadi pengetahuan yang terorganisir. (Sihombing & Yuliati, 2021).

1.2 Perumusan Masalah

Berdasarkan latar belakang yang sudah jelaskan diatas, rumusan masalah pada penelitian ini adalah bagaimana mengklasifikasikan ujaran kebencian di *Twitter* dengan menggunakan algoritma RBFN?

1.3 Tujuan Penelitian

Adapun tujuan pada penelitian ini yang dapat dilakukan yaitu :

1. Mengimplementasikan algoritma RBFN pada klasifikasi ujaran kebencian di *Twitter*.
2. Menganalisis hasil klasifikasi ujaran kebencian berdasarkan nilai akurasi, presisi, dan *recall* dengan menggunakan algoritma RBFN.

1.4 Manfaat Penelitian

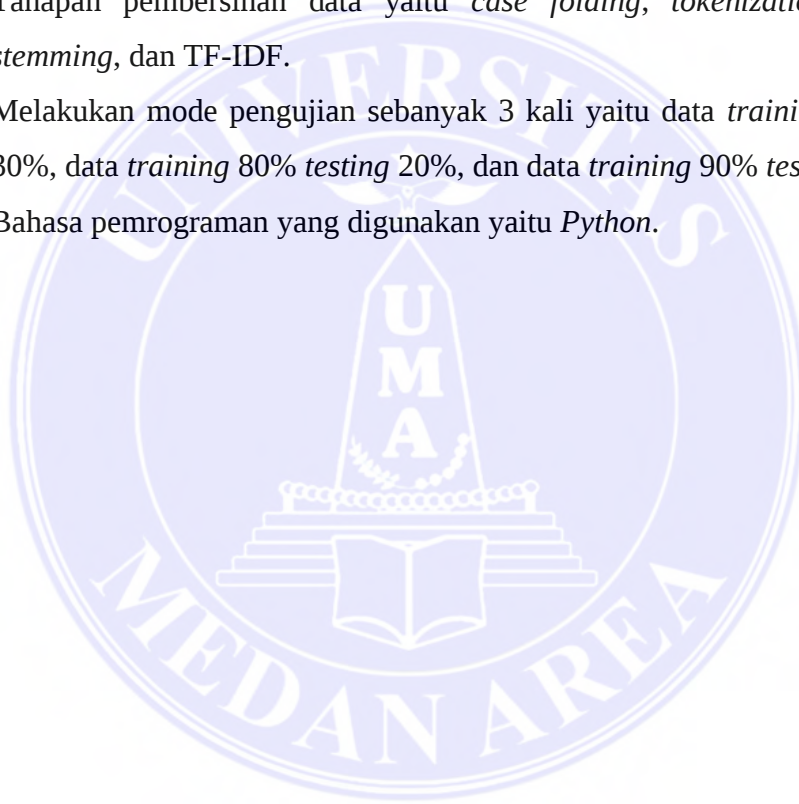
Adapun manfaat yang diperoleh dari penelitian ini yaitu :

1. Menganalisis dan mengklasifikasikan ujaran kebencian memungkinkan pemahaman yang lebih baik tentang jenis-jenis diskriminasi dan prasangka yang ada dalam masyarakat untuk pengembangan program pendidikan yang bertujuan memerangi intoleransi, mengedukasi masyarakat tentang keragaman, dan mempromosikan nilai-nilai inklusi.
2. Membantu menciptakan lingkungan *online* yang lebih aman dan ramah bagi semua pengguna sosial media.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut.

1. Data yang digunakan adalah data primer yang bersumber dari *Twitter*.
2. Teknik pengambilan data yang digunakan yaitu menggunakan API *Twitter*.
3. Dataset berjumlah 600 *record* berbahasa Indonesia dengan kata kunci “ujaran kebencian”, dimana 1 *record* adalah 1 *tweet*.
4. Menggunakan algoritma Radial Basis Function Network.
5. Tahapan pembersihan data yaitu *case folding*, *tokenization*, *stopwords*, *stemming*, dan TF-IDF.
6. Melakukan mode pengujian sebanyak 3 kali yaitu data *training* 70% *testing* 30%, data *training* 80% *testing* 20%, dan data *training* 90% *testing* 10%.
7. Bahasa pemrograman yang digunakan yaitu *Python*.



BAB II TINJAUAN PUSTAKA

2.1 *Twitter*

Twitter merupakan layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter akan tetapi pada tanggal 07 November 2017 bertambah hingga 280 karakter yang dikenal dengan sebutan kicauan atau *tweet* (Susanti dkk., 2022). Seorang pengguna dapat mengikuti pengguna lain dan pengguna yang diikuti tidak perlu mengikuti kembali. Setiap pengguna dapat mengikuti pengguna lain untuk menerima *tweet* dari pengguna lain dengan menjadi *follower* pengguna yang diikuti (Saskia, 2021).

Twitter sangat memudahkan penggunanya untuk saling berinteraksi dan berteman dengan pengguna yang lain. *Twitter* juga memiliki fitur *Top Trending* yaitu fitur yang memudahkan penggunanya dalam melihat *tweet* apa yang sedang populer dan sering dikicaukan oleh pengguna *Twitter*. *Twitter* juga memiliki konten *Verified Account* dimana konten untuk mendapatkan Lencana Akun Verifikasi di *Twitter* dan tidak sembarang orang bisa mendapatkan lencana tersebut di akunnya. Hanya kalangan selebritis, politikus, maupun orang-orang yang memiliki pengaruh saja. Awal kemunculannya, *Twitter* membuka layanan *Verified Account* bagi orang biasa, namun saat ini sudah tidak ada lagi (Bara dkk., 2022).

2.2 Ujaran Kebencian

Pengertian tuturan kebencian yaitu tuturan yang muncul di tengah masyarakat dengan memasuki budaya baru yang belum sepenuhnya terdapat kesadaran dan pemahaman dengan adanya kelebihan maupun kelemahan (Claudia & Wijayanto, 2020). Ujaran kebencian (*hate speech*) merupakan tindakan komunikasi yang dilakukan oleh individu atau kelompok sebagai bentuk provokasi, hasutan, hinaan, penistaan, pencemaran nama baik, serta penyebaran berita bohong dalam aspek seperti

ras, warna, kulit, gender, etnis, cacat fisik, orientasi seksual, kewarganegaraan, agama, dan lain-lain (Permatasari & Subyantoro, 2020).

2.3 Text Preprocessing

Text preprocessing merupakan tahapan awal sebelum mengolah data lebih lanjut dan kemudian digunakan dalam proses pembuatan sistem. Tahapan dalam *text preprocessing* ada berbagai macam tahap dan tahapan preprocessing yang akan digunakan nantinya akan menyesuaikan dengan proses yang akan dilalui. *Text preprocessing* dilakukan untuk mengubah data menjadi lebih mudah dan sesuai untuk diproses oleh sistem, dan juga untuk membuang data-data yang tidak diperlukan oleh sistem tersebut. *Text preprocessing* sangat penting dalam melakukan analisis mengenai teks, terutama untuk media sosial yang sebagian besar berisi kata-kata atau kalimat yang tidak formal dan tidak terstruktur serta memiliki *noise* yang besar (Hakiem dkk., 2019). Tahapan pada *text preprocessing* adalah sebagai berikut:

2.3.1 Case Folding

Case Folding, bertujuan untuk mengubah semua karakter huruf pada teks menjadi huruf kecil. Selain itu, proses ini juga menghilangkan karakter non-huruf seperti tanda baca, angka, dan *Uniform Resource Locator* (URL) (Sandryan dkk., 2021).

2.3.2 Tokenization

Tokenization yaitu memisahkan teks menjadi unit-unit yang lebih kecil yang disebut token. Token bisa berupa kata, frasa, atau karakter tergantung pada kebutuhan. Misalnya, kalimat "Saya sedang belajar pemrosesan teks" dapat ditokenisasi menjadi ["Saya", "sedang", "belajar", "pemrosesan", "teks"] (Septiawan & Chairani, 2023).

2.3.3 Stopwords

Pada tahap ini dilakukan pemilihan kata penting dari hasil sebelumnya yaitu *tokenizing*. Proses ini menggunakan algoritma *stoplist* (membuang kata yang tidak penting) dan *wordlist* (menyimpan kata penting) dengan menggunakan kamus list

Indonesia, contoh: di, aku, jika, yang, dari, kamu, mau, dia, mereka, dll (Roy & Ojha, 2020).

2.3.4 Stemming

Pada tahapan ini berfungsi untuk memperkecil kata-kata menjadi akar kata atau mengubah kata-kata ke bentuk dasar. Contohnya, kata-kata "berlari", "berlari", dan "berlari" dapat distem menjadi "lar" (Septiawan & Chairani, 2023).

2.3.5 TF-IDF (*Term Frequency-Inverse Document Frequency*)

Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu teknik yang digunakan dalam pengolahan teks dan pemodelan bahasa alami. Tujuan utama dari metode TF-IDF adalah untuk mengevaluasi seberapa penting suatu kata (*term*) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar (Septiawan & Chairani, 2023).

Rumus metode *Term Frequency-Inverse Document Frequency* (TF-IDF):

$$idf = \log \frac{N}{df} \quad (2.1)$$

$$w(k, d) = tf(k, d) * idf \quad (2.2)$$

Keterangan :

Idf = *Inverse document frequency*

$W(k,d)$ = Bobot kata (*term*) yang tersedia dalam dokumen

$tf(k,d)$ = jumlah frekuensi munculnya kata didalam dokumen

N = jumlah seluruh dokumen yang tersedia didalam *database*

df = jumlah dokumen yang mengandung *term*

2.4 Algoritma *Radial Basis Function Network* (RBFN)

Radial Basis Function Network atau disingkat RBFN merupakan arsitektur dari ANN atau *Artificial Neural Network* yang bersifat *feed forward* dan dapat melakukan proses klasifikasi dengan waktu singkat. RBFN melakukan proses pembelajaran yang sangat cepat karena neuron disetel secara lokal. Jaringan syaraf tiruan RBFN merupakan salah satu teknik pengembangan jaringan syaraf tiruan yang pelatihannya memakai teknik pembelajaran terawasi. Jenis jaringan pada RBFN hampir mirip dengan jaringan syaraf tiruan *Multilayer Perceptron*. Waktu pelatihan (*training*) pada jaringan RBFN sangat cepat dan memiliki kemampuan generalisasi yang baik (Wahyuningrum, 2020).

Keunggulan dari RBFN ini merupakan mempunyai perulangan komputasi yang lebih cepat dibandingkan dengan teknik lain untuk jaringan syaraf tiruan. RBFN memakai fungsi linier antara input dan lapisan tersembunyi. Jumlah lapisan tersembunyi di RBFN diperbarui secara berurutan hingga kesalahan target tercapai. Untuk setiap neuron pada lapisan tersembunyi yang ditambahkan, jarak antara vektor *input* dan vektor *output* dihitung menerapkan norma *Euclidean*. Kemudian, neuron diaktifkan menggunakan fungsi aktivasi *Sigmoid* sebagai jenis umum dari radial basis fungsi (RBF) (Madhiarasan, 2020).

RBFN merupakan jaringan saraf tiga lapis dengan banyak neuron di lapisan tersembunyi. Pada RBFN, fungsi aktivasi basis radial nonlinier ditugaskan ke lapisan tersembunyi, sedangkan fungsi linier ditugaskan ke lapisan keluaran. Menyerupai dengan MLPNN, RBFN dapat dilatih menggunakan algoritma *backpropagation* yang diawasi (Kermani dkk., 2020). Rumus *backpropagation* adalah sebagai berikut.

1. Inisialisasi bobot dengan nilai acak *range* kecil.
2. Fase Propagasi maju :
 - a.) Hitung keluaran unit tersembunyi (z_j):

Z_j merupakan nilai dari hasil keluaran *hidden layer*.

$$z_{net_j} = v_{j0} + \sum_{i=1}^n x_i v_{ji} \quad (2.3)$$

$$z_i = f(z_{net_j}) = \frac{1}{1 + e^{-z_{net_j}}} \quad (2.4)$$

Dimana :

Z_j = unit tersembunyi ke-j

F = fungsi nonlinear

V_{j0} = bias pada unit tersembunyi –j

$\sum x_i$ = jumlah input ke-i

V_{ji} = bobot ke-j

b.) Hitung keluaran unit y_k :

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \quad (2.5)$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-y_{net_k}}} \quad (2.6)$$

Dimana :

Y_k = unit output ke-k

F = fungsi nonlinear

W_{k0} = unit ke-0 dan k

$\sum z_j$ = jumlah unit tersembunyi ke-j

W_{kj} = bobot awal

3. Fase Propagasi Mundur :

Hitung faktor δ di unit keluaran y_k

$$\delta_k = (t_k - y_k)f'(y_{net_k}) = (t_k - y_k)y_k(1 - y_k), t_k = Target \quad (2.7)$$

$$\Delta w_{kj} = \alpha \delta_k z_j, \quad k = 1, 2, \dots, m; j = 1, 2, \dots, p \quad (2.8)$$

δ_k adalah nilai *error* lapisan keluaran.

y_k adalah nilai keluaran yang diharapkan.

Dimana :

δ_k = informasi tentang kesalahan pada unit y_k yang disebarkan kembali ke unit tersembunyi

t_k = output vektor target ke- k

y_k = unit output ke- k

f = fungsi nonlinear

$\sum z_i$ = jumlah unit tersembunyi ke- j

W_{kj} = bobot awal

4. Hitung penjumlahan kesalahan dari unit tersembunyi (δ):

$$\delta_{net_j} = \sum_{k=1}^m \delta_k w_{kj} \quad (2.9)$$

Faktor kesalahan δ di unit tersembunyi :

$$\delta_j = \delta_{net_j} f'(z_{net_j}) = \delta_{net_j} z_j (1 - z_j) \quad (2.10)$$

Suku perubahan bobot ke unit tersembunyi :

$$\Delta v_{ji} = \alpha \delta_j x_i, \quad j = 1, 2, \dots, p; i = 1, 2, \dots, n \quad (2.11)$$

5. Hitung semua perubahan bobot :

Perubahan bobot unit keluaran :

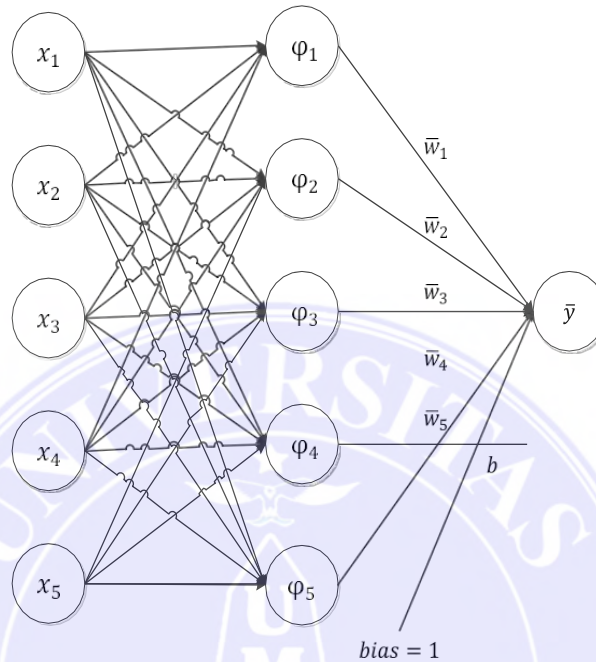
$$w_{kj}(\text{baru}) = w_{kj}(\text{lama}) + \Delta w_{kj}, \quad (k = 1, 2, \dots, m; j = 1, 2, \dots, p) \quad (2.12)$$

Perubahan bobot unit tersembunyi :

$$v_{ji}(\text{baru}) = v_{ji}(\text{lama}) + \Delta v_{ji}, \quad (j = 1, 2, \dots, p; i = 1, 2, \dots, n) \quad (2.13)$$

RBFN mempunyai 3 lapisan, yaitu lapisan masukan, satu lapisan tersembunyi dan lapisan keluaran. RBFN mempunyai bobot antara lapisan tersembunyi dan lapisan keluaran, fungsi aktivasi dapat menghasilkan nilai berupa persamaan

nonlinier pada lapisan tersembunyi, sebaliknya nilai yang dihasilkan oleh metode RBFN pada lapisan keluaran berupa persamaan linier (Prastowo dkk., 2021).



Gambar 2. 1 Arsitektur RBFN

Sumber : (Nikentari dkk., 2018)

2.5 Confusion Matrix

Confusion Matrix adalah suatu proses yang dipakai dalam menganalisa keakuratan dari model klasifikasi yang dibuat untuk mengidentifikasi data dengan kelas yang berbeda. Dengan diukurnya tingkat keakurasian maka dapat diketahui performa dari suatu model klasifikasi tersebut. Hasil akurasi dari sebuah klasifikasi dapat diketahui dari hasil data latih dengan bentuk persentase dari kelompok data latih, yang merupakan hasil dari klasifikasi yang benar (Nugroho, 2019). Sedangkan, hasil uji dari model klasifikasi yang akan dianalisa menggunakan tabel atau sering disebut (*confusion matrix*) dapat memperoleh nilai *accuracy*, *recall*, dan *precision*. Berikut *Confusion matrix* yang dievaluasi dengan tabel yang menerangkan bahwa hasil klasifikasi dari jumlah data *testing* yang benar dan dari jumlah data *testing* yang salah (Anggreany, 2022).

Tabel 2. 1 Cunfusion Matrix

Data Aktual	Data Prediksi	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan :

TP (*True Positive*) = Prediksi Benar Bernilai Positif

FN (*False Negative*) = Negatif Terprediksi Positif

FP (*False Positive*) = Positif Terprediksi Negatif

TN (*True Negative*) = Prediksi Benar Bernilai Negatif

Nilai akurasi adalah persentase terhadap jumlah *record* data yang benar diklasifikasikan oleh algoritma sesudah pengujian dibagi dengan jumlah total *record*. *Precision* adalah jumlah komentar positif yang diprediksi sebagai nilai positif pada data aktual. *Recall* adalah jumlah komentar positif aktual yang diprediksi sebagai nilai positif (Afrillia dkk., 2022). Untuk menghitung tingkat akurasi, *precision*, dan *recall* dapat dilihat pada persamaan berikut :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.14)$$

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (2.15)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.16)$$

2.6 Penelitian Terdahulu

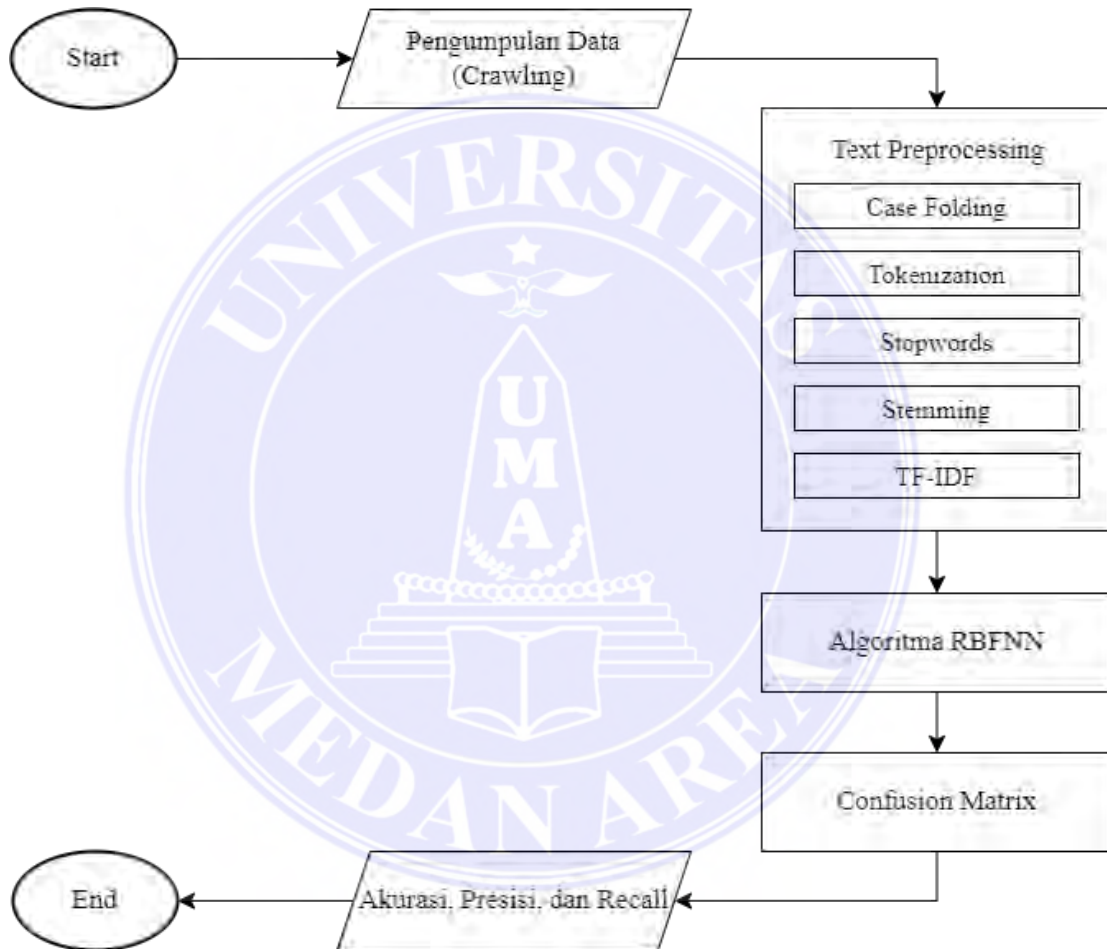
No	Judul	Penulis	Metode	Hasil
1.	Penerapan Radial Basis Function Untuk Klasifikasi Jenis Tanah	(Puspaningrum dkk., 2020)	Radial Basis Function Neural Network	Hasil penelitian yang dilakukan dapat disimpulkan bahwa proses klasifikasi jenis tanah dilakukan dengan menggunakan 36 data. Dimana 21 data berupa data training dan 15 data uji citra yang berisi gambar tanah. Jenis tanah yang digunakan jenis Tanah Entisol, Mollisol, dan Ultisol. Dari uji coba yang telah dilakukan nilai akurasi yang didapatkan yaitu 90,4%.
2.	Penerapan Radial Basis Function Neural Network dalam Pengklasifikasi Daerah Tertinggal di Indonesia	(Wahyuningrum, 2020)	Radial Basis Function Neural Network	Penelitian ini menggunakan data sebanyak 491 kabupaten/kota di Indonesia serta memiliki 27 variabel. Hasil dari penelitian ini yaitu RBFNN memberikan performa yang sangat baik yaitu nilai akurasi sebesar 93,48 %, sensitivitas 81,10 %, spesifikasi 97,43 %, F-Measure 85,36 %.
3.	Deteksi Ujaran Kebencian Pada Twitter Bahasa Indonesia Menggunakan Machine Learning: Reviu Literatur	(Dwitama, 2021)	Recurrent Neural Network	Dataset yang digunakan yaitu 13.169 dengan rincian 5.561 teks yang termasuk ke dalam teks ujaran kebencian dan 7.608 lainnya tidak termasuk ujaran kebencian. Hasilnya mendapatkan akurasi sebesar 91%.
4.	Tinjauan Pustaka Sistematis:	(Pratiwi dkk., 2022)	Neural Network	Model CNN yang diperoleh berhasil memberikan hasil yang

	Klasifikasi Ujaran Kebencian Pada Sosial Media Dengan Algoritma Neural Network			sangat baik dengan akurasi tinggi untuk klasifikasi ujaran kebencian multilabel pada social media pada dataset dengan akurasi 98,07%. Media social yang banyak digunakan terkait hate speech adalah Twitter. Hasil penelitian ini merekomendasikan analisis lebih lanjut dari penerapan model CNN untuk menangani kasus ujaran kebencian di social media.
5.	Klasifikasi Ujaran Kebencian Menggunakan Metode FeedForward Neural Network (IndoBERT)	(Dharmawan dkk., 2023)	FeedForward neural network dan IndoBERT	Metode FeedForward neural network dengan IndoBERT berhasil melakukan klasifikasi dengan nilai akurasi terbaik sebesar 89,52%. Model memiliki kecenderungan memberi hasil klasifikasi “ujaran kebencian”.
6.	Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan CNN-LSTM	(Ridwan & Muzakir, 2022)	Convolutional Neural Network dan Long Short-Term Memory	Pada penelitian ini diperoleh hasil mengenai model klasifikasi terbaik untuk ujaran kebencian di media sosial Twitter. Model yang dibangun dengan kombinasi CNN-LSTM dan Skip-gram sebagai word embedding-nya berhasil memperoleh nilai akurasi pada tes uji coba terbaik dengan tingkat akurasi sebesar 69.1%. Pada tes uji coba juga didapatkan beberapa permasalahan dalam klasifikasi, yang dimana untuk sebagian tweet yang

				ada terdapat test data tidak dapat diklasifikasikan dengan baik. Hal ini disebabkan oleh ketidakseimbangan data.
7.	Identifikasi Ujaran Kebencian pada Twitter Menggunakan Metode Convolutional Neural Network (CNN)	(Adhari & Sibaroni, 2023)	Convolutional Neural Network	Data yang digunakan berasal dari hastag seperti: #Tolak3Periode, #PembajakKonstitusi, #JokowiKejarSetoran, dan #BubarkanKhilafatulMuslimi dengan menggunakan Twitter API dengan library Tweepy yang diambil dari tanggal 03 Agustus 2021 hingga 15 Maret 2022. Hasil dari pengumpulan data ini sebanyak 3.365 data. Pelabelan data dilakukan oleh 2 dengan tujuan untuk mengurangi subjektivitas dalam melakukan pelabelan. Hasil kombinasi hyperparameter tuning CNN terbaik ada di learning rate 0,001 dan dropout 0,3 yang menghasilkan akurasi 81%.

BAB III METODE PENELITIAN

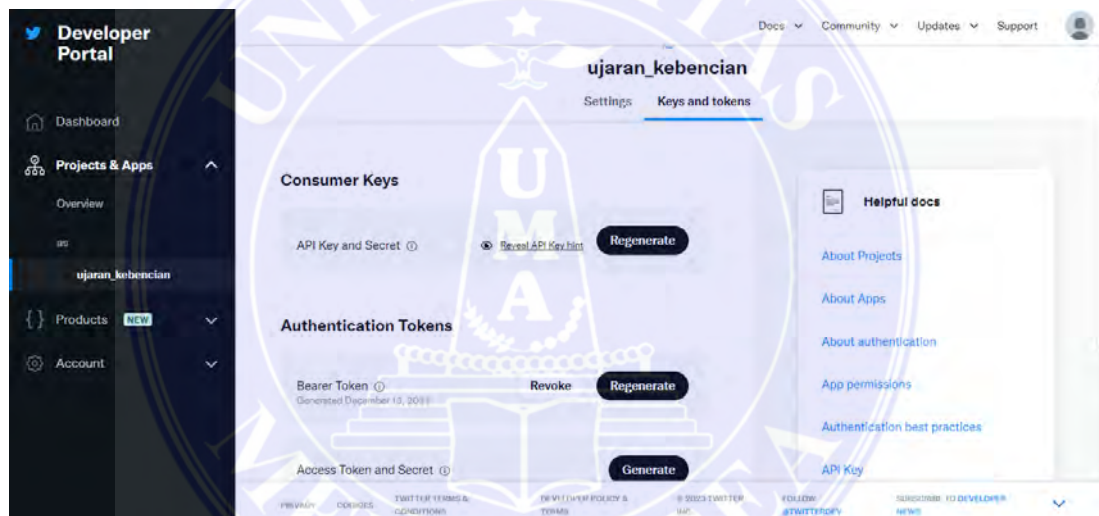
Bab ini akan mengulas beberapa aspek, seperti alur atau tahapan yang akan diterapkan dalam penelitian. Tahapan penelitian ini akan mencakup pengumpulan data, *text preprocessing*, algoritma RBFN, *confusion matrix*, dan hasil performa. Alur penelitian dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Alur Penelitian

3.1 Pengumpulan Data

Pengumpulan data pada penelitian ini adalah bersifat primer dengan menggunakan API (*Application Programming Interface*) *Twitter* dengan cara mendaftar terlebih dahulu ke forum portal *developer.twitter.com*. API berperan sebagai sistem kontak yang dibuat dengan *Twitter*. Untuk dapat mengakses API, diperlukan *bearer token* yang bertindak sebagai kunci agar *Twitter* dapat memahami apa yang dilakukan aplikasi serta bagaimana melakukannya dengan mendaftarkan aplikasi *Twitter* API. Kemudian menggunakan bahasa pemrograman *Python* untuk mencari data dari *Twitter* berlandaskan kata kunci. Website *developer Twitter* dapat dilihat pada Gambar 3.2.



Gambar 3. 2 Developer Portal Twitter

Setelah itu dapat mengakses kode *bearer token* yang ada di *developer Twitter* dengan menekan *regenerate* maka akan menghasilkan
 “AAAAAAAAAAAAAAAAAAAAAAAAAPjTrQEAAAAhPs%2BWYy%2F7bIT7CVm5gX9vKptx4E%3DLxd4dEjQBvfarnL4cAyLCczglUe4h9AZQjT7rDJoCYLingxUFW”.

```
# crawling data twitter menggunakan API V2

import tweepy # library
import pandas as pd # library

# Menggunakan token dari twitter API V2
bearer_token = 'AAAAAAAAAAAAAAAAAApJTrQEAAAAAHps%2BWYy%2F7bIT7CVm5gX9vKptx4EX3DLxd4dEjQBvFarnL4cAylCcZglUe4h9AZQjT7rDJoCYLingxUFW'

client = tweepy.Client(bearer_token)

# kata kunci
kata_kunci = "ujaran kebencian -filter:retweets"

tweets = client.search_all_tweets(kata_kunci, max_results=600)

users_locs = [[tweet.text, tweet.user.username, tweet.created_at] for tweet in tweets]

tweet_text = pd.DataFrame(data = users_locs, columns = ["full_text", "username", "datetime"])

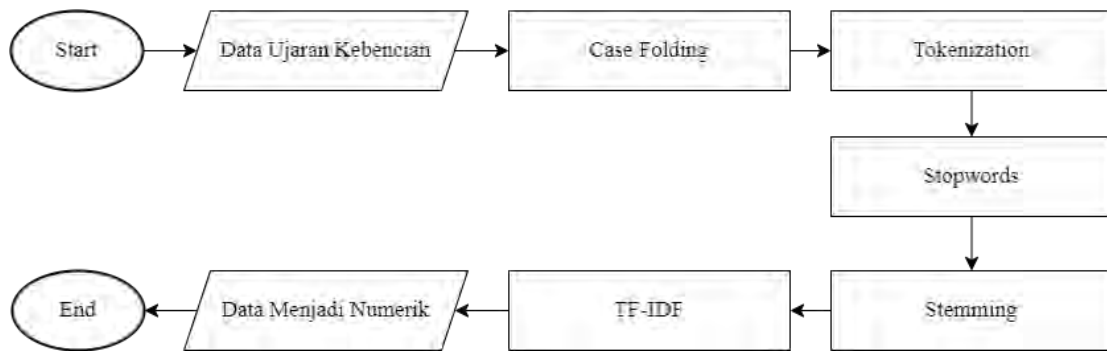
tweet_text.to_csv('data_ujaran_kebencian.csv', index=False) # simpan file csv
```

Gambar 3. 3 *Crawling Data*

Setelah mendapatkan tokennya langkah berikutnya adalah melakukan *crawling data* dengan *Python* seperti Gambar 3.3. Adapun kata kunci yang digunakan dalam proses *crawling* ialah “ujaran kebencian” dengan jumlah data sebesar 600 *record*, atribut yang didapatkan yaitu berupa teks, nama pengguna, dan tanggal. Data tersebut kemudian di simpan dengan format CSV.

3.2 Text Preprocessing

Tahapan pembersihan data adalah langkah penting dalam *Natural Language Processing* (NLP) yang melibatkan pembersihan dan transformasi data teks mentah menjadi format yang sesuai untuk analisis. Meskipun langkah-langkahnya dapat bervariasi tergantung pada tugas dan sifat data, berikut adalah tahapan dalam pembersihan teks pada penelitian ini:



Gambar 3. 4 Alur *Text Preprocessing*

3.2.1 *Case Folding*

Tahapan ini berfungsi mengubah seluruh teks menjadi huruf kecil. Ini memastikan keseragaman dan membantu dalam memperlakukan kata secara *case-insensitive*. Adapun tahapan tersebut dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Tahapan *Case Folding*

Teks	<i>Case Folding</i>
caci maki dan ujaran kebencian kepadanya. #BersamaIndonesiaMaju #PrabowoGibranIstimewa #PrabowoGemoy Prabowo Subianto	caci maki dan ujaran kebencian kepadanya bersamaIndonesiamaju prabowogibranIstimewa prabowogemoy prabowo subianto
ngga ada mutu balesannya cma ujaran kebencian semata. harga diri akun ini mana sih? atau emang udah dibeli harga dirinya	ngga ada mutu balesannya cma ujaran kebencian semata harga diri akun ini mana sih atau emang udah dibeli harga dirinya

3.2.2 *Tokenization*

Tahapan ini berfungsi membagi teks menjadi kata-kata atau token individu. Tokenisasi adalah proses membagi teks menjadi unit-unit kecil, seperti kata-kata atau frasa. Adapun tahapan tersebut dapat dilihat pada Tabel 3.2.

Tabel 3. 2 Tahapan *Tokenization*

<i>Case Folding</i>	<i>Tokenization</i>
caci maki dan ujaran kebencian kepadanya bersamaIndonesiamaju	“caci”, “maki”, “dan”, “ujaran”, “kebencian”, “kepadanya”,

prabowogibranIstimewa prabowogemoy prabowo subianto	“bersamaindonesiamaju”, “prabowogibranistimewa”, “prabowogemoy”, “prabowo”, “subianto”
ngga ada mutu balesannya cma ujaran kebencian semata harga diri akun ini mana sih atau emang udah dibeli harga dirinya	“ngga”, “ada”, “mutu”, “balesannya”, “cma”, “ujaran”, “kebencian”, “semata”, “harga”, “diri”, “akun”, “ini”, “mana”, “sih”, “atau”, “emang”, “udah”, “dibeli”, “harga”, “dirinya”

3.2.3 Stopwords

Tahapan ini berfungsi Menghilangkan kata-kata umum yang tidak banyak berkontribusi pada makna keseluruhan teks (contohnya: "dan", "yang", "adalah", "di"). Adapun tahapan *stopwords* dapat dilihat pada Tabel 3.3.

Tabel 3. 3 Tahapan *Stopwords*

<i>Tokenization</i>	<i>Stopwords</i>
“caci”, “maki”, “dan”, “ujaran”, “kebencian”, “kepadanya”, “bersamaindonesiamaju”, “prabowogibranistimewa”, “prabowogemoy”, “prabowo”, “subianto”	“caci”, “maki”, “ujaran”, “kebencian”, “bersamaindonesiamaju”, “prabowogibranistimewa”, “prabowogemoy”, “prabowo”, “subianto”
“ngga”, “ada”, “mutu”, “balesannya”, “cma”, “ujaran”, “kebencian”, “semata”, “harga”, “diri”, “akun”, “ini”, “mana”, “sih”, “atau”, “emang”, “udah”, “dibeli”, “harga”, “dirinya”	“balesannya”, “ujaran”, “kebencian”, “semata”, “harga”, “diri”, “akun”, “dibeli”, “harga”, “dirinya”

3.2.4 Stemming

Tahapan ini berfungsi mengurangi kata-kata ke bentuk dasar atau akar untuk normalisasi variasi (contohnya: "berlari" menjadi "lari", "lebih baik" menjadi "baik"). Adapun tahapan *stemming* dapat dilihat pada Tabel 3.4.

Tabel 3. 4 Tahapan *Stemming*

<i>Stopwords</i>	<i>Stemming</i>
“caci”, “maki”, “ujaran”, “kebencian”, “bersamaindonesiamaju”, “prabowogibranistimewa”, “prabowogemoy”, “prabowo”, “subianto”	“caci”, “maki”, “ujaran”, “benci”, “indonesia”, “prabowo”, “gibran”, “prabowo”, “prabowo”, “subianto”
“balesannya”, “ujaran”, “kebencian”, “semata”, “harga”, “diri”, “akun”, “dibeli”, “harga”, “dirinya”	“bales”, “ujaran”, “benci”, “semata”, “harga”, “diri”, “akun”, “beli”, “harga”, “diri”

3.2.5 *Term-Frequency Inverse Document Frequency (TF-IDF)*

TF-IDF digunakan dalam pemrosesan teks dan temu balik informasi untuk mengevaluasi seberapa penting suatu kata dalam sebuah dokumen terhadap kumpulan dokumen yang lebih besar. Skor TF-IDF lebih tinggi untuk kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen-dokumen lain. Ini memungkinkan kita untuk menemukan kata-kata yang secara relatif unik dan memberikan kontribusi besar terhadap pemahaman konten suatu dokumen dalam konteks kumpulan dokumen yang lebih besar. Adapun tahapan TF-IDF dapat dilihat pada Tabel 3.5.

Tabel 3. 5 Tahapan TF-IDF

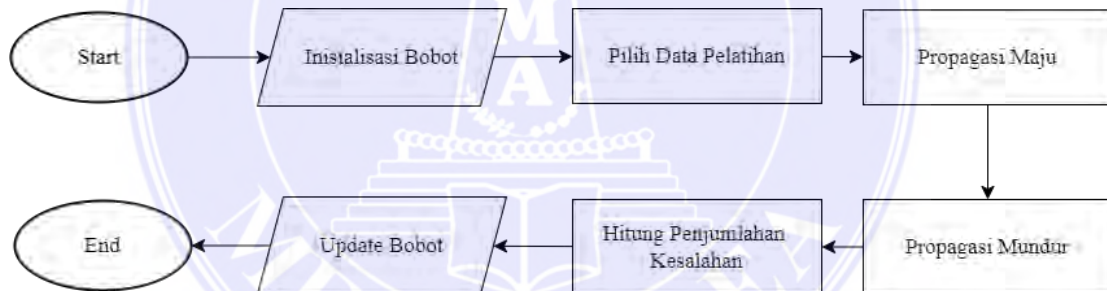
<i>Term</i>	TF		df	IDF	TF-IDF	
	N1	N2			D1	D2
caci	1	0	1	0,301	0,301	0
maki	1	0	1	0,301	0,301	0
ujaran	1	1	2	0	0	0
benci	1	1	2	0	0	0
indonesia	1	0	1	0,301	0,301	0
prabowo	1	0	1	0,301	0,301	0
gibran	1	0	1	0,301	0,301	0
subianto	1	0	1	0,301	0,301	0
bales	0	1	1	0,301	0	0,301
semata	0	1	1	0,301	0	0,301
harga	0	1	1	0,301	0	0,301
diri	0	1	1	0,301	0	0,301

akun	0	1	1	0,301	0	0,301
beli	0	1	1	0,301	0	0,301

Selanjutnya menghitung nilai TF-IDF menggunakan kata per kata tanpa mengulangi kata yang sama. Kolom TF (berisi N1 dan N2) yang merupakan jumlah data yang tersedia pada Tabel 3.4. kemudian kolom df merupakan jumlah dokumen atau data yang mengandung kata tersebut, contoh kata benci tersedia pada data N1 dan N2. Selanjutnya pada kolom IDF menggunakan rumus persamaan (2.1), dan terakhir kolom TF-IDF adalah hasil perkalian dari kolom TF (yang berisi N1 dan N2) dengan IDF.

3.3 Perhitungan Manual Radial Basis Function Neural Network

Selanjutnya, menghitung manual algoritma RBFN menggunakan *backpropagation* dengan inputan dari hasil TF-IDF. *Flowchart* dari algoritma RBFN dapat dilihat pada Gambar 3.5.



Gambar 3. 5 Alur Algoritma RBFN

1. Fase Propagasi Maju :

a.) Hitung keluaran unit tersembunyi (z_j) :

Z_j merupakan nilai dari hasil keluaran *hidden layer*.

$$Z_{net1} = (0,301 * 0,02) + (0,301 * 0,071) + (0 * 0,030) + (0 * (-0,003)) + (0,301 * (-0,210)) + (0,301 * 0,121) + (0,301 * 0,111) + (0,301 * 0,012) + (0,301 * 0,012)$$

$$(-0,101)) + (0,301 * 0,040) + (0,301 * (-0,121)) + (0,301 * 0,025) + (0,301 * (-0,020)) + (0,301 * 0,107) = 0,016$$

$$\begin{aligned} z_{net_2} &= (0,301 * (-0,100)) + (0,301 * 0,170) + (0 * (-0,034) + (0 * 0,103) + \\ & (0,301 * (-0,002)) + (0,301 * 0,090) + (0,301 * 0,001) + (0,301 * (-0,100)) + \\ & (0,301 * 0,011) + (0,301 * 0,002)) + (0,301 * (-0,022)) + (0,301 * (-0,01)) + \\ & (0,301 * 0,020) + (0,301 * 0,113) = 0,052 \end{aligned}$$

Menghitung *Sigmoid* :

Fungsi Aktivasi *Sigmoid Biner* Menghasilkan nilai *output* pada interval 0 hingga 1.

$$f(z_{net_1}) = \frac{1}{1 + e^{-0,016}} = 0,504$$

$$f(z_{net_2}) = \frac{1}{1 + e^{-0,052}} = 0,513$$

b.) Hitung keluaran unit y_k :

y_k merupakan nilai fungsi aktivitas pada lapisan keluaran.

$$\begin{aligned} y_{net_k} &= w_{k0} + \sum_{j=1}^p z_j w_{kj} = 0,010 + (0,504)(0,1) + (0,513)(0,012) \\ &= 0,070 \end{aligned}$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-0,070}} = 0,517$$

w_{kj} merupakan nilai bobot yang menghubungkan masukan node 2 dengan lapisan keluaran 2.

2. Fase Propagasi Mundur :

Hitung faktor δ di unit keluaran y_k

$$\begin{aligned}\delta_k &= (t_k - y_k)f'(y_{net_k}) = (t_k - y_k)y_k(1 - y_k) \\ &= (1 - 0,517)0,517(1 - 0,517) = 0,120\end{aligned}$$

$$\Delta w_{10} = \alpha \delta_k z_0 = 0,2 (0,120) (1) = 0,024$$

$$\Delta w_{11} = \alpha \delta_k z_1 = 0,2 (0,120) (0,504) = 0,012$$

$$\Delta w_{12} = \alpha \delta_k z_2 = 0,2 (0,120) (0,513) = 0,012$$

δ_k adalah nilai *error* lapisan keluaran.

T_k adalah nilai keluaran yang diharapkan.

3. Hitung penjumlahan kesalahan dari unit tersembunyi (δ) :

Faktor kesalahan δ di unit tersembunyi :

$$\delta_1 = (0,120)(0,04) = 0,004$$

$$\delta_2 = (0,120)(-0,02) = -0,002$$

$$\delta_1 = \delta_{net_{z1}}(1 - z_1) = (0,6)(0,504)(1 - 0,504) = 0,149$$

$$\delta_2 = \delta_{net_{z2}}(1 - z_2) = (-0,4)(0,513)(1 - 0,513) = -0,099$$

Suku perubahan bobot ke unit tersembunyi :

$$\Delta v_{10} = \alpha \delta_1 = (0,2)(0,149)(1) = 0,029$$

$$\Delta v_{20} = \alpha \delta_2 = (0,2)(-0,099)(1) = -0,019$$

$$\Delta v_{11} = \alpha \delta_1 = (0,2)(0,149)(0,301) = 0,008$$

$$\Delta v_{21} = \alpha \delta_2 = (0,2)(-0,099)(0,301) = -0,005$$

$$\Delta v_{12} = \alpha \delta_1 = (0,2)(0,149)(0) = 0$$

$$\Delta v_{22} = \alpha \delta_2 = (0,2)(-0,099)(0) = 0$$

4. Hitung semua perubahan bobot :

Perubahan bobot unit keluaran :

$$w_{10}(\text{baru}) = w_{10}(\text{lama}) + \Delta w_{10} = 0,3 + 0,024 = 0,324$$

$$w_{11}(\text{baru}) = w_{11}(\text{lama}) + \Delta w_{11} = 0,1 + 0,012 = 0,112$$

$$w_{12}(\text{baru}) = w_{12}(\text{lama}) + \Delta w_{12} = -0,5 + 0,012 = -0,488$$

Perubahan bobot unit tersembunyi :

$$v_{10}(\text{baru}) = v_{10}(\text{lama}) + \Delta v_{10} = -0,01 + 0,029 = 0,019$$

$$v_{20}(\text{baru}) = v_{20}(\text{lama}) + \Delta v_{20} = 0,01 + (-0,019) = -0,009$$

$$v_{11}(\text{baru}) = v_{11}(\text{lama}) + \Delta v_{11} = 0,02 + 0,008 = 0,028$$

$$v_{21}(\text{baru}) = v_{21}(\text{lama}) + \Delta v_{21} = 0,03 + (-0,005) = 0,025$$

$$v_{12}(\text{baru}) = v_{12}(\text{lama}) + \Delta v_{12} = 0,03 + 0 = 0,03$$

$$v_{22}(\text{baru}) = v_{22}(\text{lama}) + \Delta v_{22} = 0,01 + 0 = 0,01$$

Berdasarkan fase nomor 4 dapat disimpulkan bahwa klasifikasi pada data sampel di Tabel 3.1 sebagai berikut :

1. Ditentukan berdasarkan perubahan bobot unit tersembunyi $V_{10}(\text{baru}) = 0,019$, $V_{11}(\text{baru}) = 0,028$, dan $V_{12}(\text{baru}) = 0,03$. Hasil tersebut akan ditentukan berdasarkan yang paling dominan apakah mengarah ke *Hate Speech* (HS) atau *Non-Hate Speech* (N-HS). Maka data sampel nomor 1 menjadi kelas **N-HS**.
2. Ditentukan berdasarkan perubahan bobot unit tersembunyi $V_{20}(\text{baru}) = -0,009$, $V_{21}(\text{baru}) = 0,025$, dan $V_{22}(\text{baru}) = 0,01$. Hasil tersebut akan ditentukan

berdasarkan yang paling dominan apakah mengarah ke *Hate Speech* (HS) atau *Non-Hate Speech* (N-HS). Maka data sampel nomor 2 menjadi kelas **N-HS**.

Setelah mendapatkan masing – masing kelasnya, selanjutnya adalah menggabungkan teks sampel dengan kelas tersebut. Maka dapat dilihat pada Tabel 3.6.

Tabel 3. 6 Hasil Klasifikasi Menggunakan Algoritma RBFN

No	Teks	Kelas
1.	caci maki dan ujaran kebencian kepadanya. #BersamaIndonesiaMaju #PrabowoGibranIstimewa #PrabowoGemoy Prabowo Subianto	N-HS
2.	ngga ada mutu balesannya cma ujaran kebencian semata. harga diri akun ini mana sih? atau emang udah dibeli harga dirinya	N-HS

3.4 Confusion Matrix

Sub – bab ini akan mengevaluasi hasil klasifikasi menggunakan *confusion matrix*, *matrix* ini memberikan gambar tentang seberapa baik model tersebut dapat memprediksi kelas – kelas target. Hasilnya dapat dilihat pada Tabel 3.7.

Tabel 3. 7 Hasil *Confusion Matrix*

Data Aktual	Data Prediksi	
	HS	N-HS
HS	0	0
N-HS	0	2

$$\text{Akurasi} = \frac{0 + 2}{0 + 2 + 0 + 0} \times 100\% = \mathbf{100\%}$$

$$\text{Presisi} = \frac{0}{0 + 0} \times 100\% = \mathbf{0\%}$$

$$\text{Recall} = \frac{0}{0 + 0} \times 100\% = \mathbf{0\%}$$

Dengan menggunakan algoritma RBFN, penelitian ini dapat mengklasifikasikan ujaran kebencian di *Twitter* dengan mengikuti langkah-langkah yang mencakup pengumpulan data, *text preprocessing*, algoritma RBFN, *confusion matrix*, dan hasil performa. Dengan pendekatan ini, didapatkan hasil akurasi sebesar 100% dalam mengklasifikasikan data sampel diatas. Model tersebut dapat membantu dalam deteksi dan penanganan ujaran kebencian secara otomatis di platform *Twitter*.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian tentang klasifikasi ujaran kebencian di Twitter menggunakan algoritma *Radial Basis Function Network* (RBFN), dapat disimpulkan sebagai berikut:

1. Penelitian ini berhasil mengimplementasikan algoritma RBFN untuk klasifikasi ujaran kebencian di *Twitter* dengan melalui tahapan *text preprocessing* yang meliputi *case folding*, *tokenization*, penghapusan *stopwords*, *stemming*, dan representasi numerik dengan TF-IDF.
2. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, dan recall. Model menunjukkan performa yang baik, dengan akurasi tertinggi sebesar 87,7% pada pembagian data 70% *training* dan 30% *testing*.
3. Algoritma RBFN terbukti mampu melakukan klasifikasi ujaran kebencian dengan baik pada dataset yang digunakan, meskipun terdapat sedikit penurunan kinerja seiring dengan perubahan proporsi data uji.

5.2 Saran

Adapun saran yang dapat dilakukan pada penelitian selanjutnya adalah:

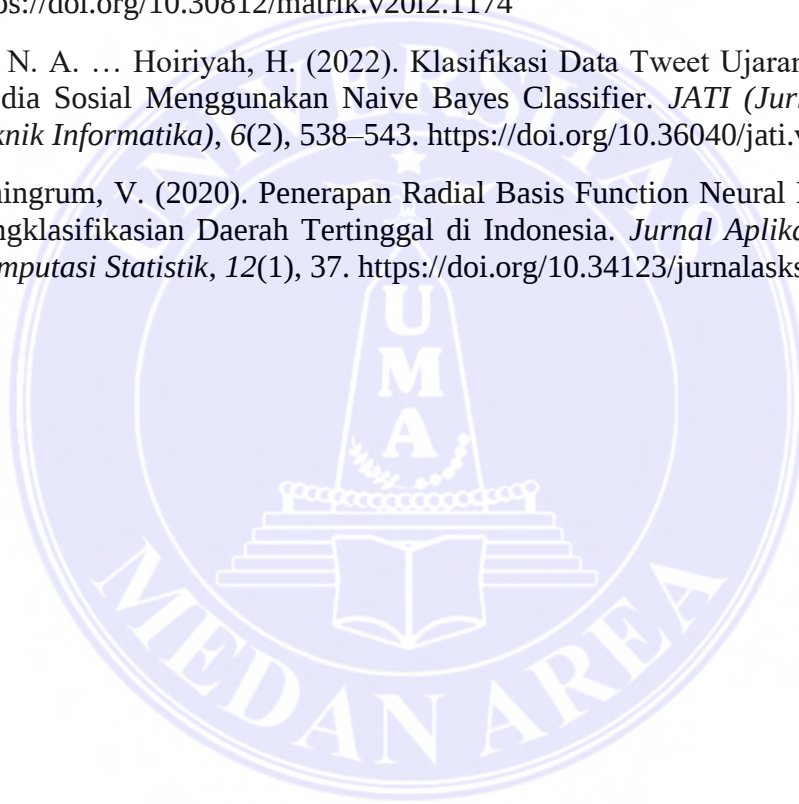
1. Melakukan pengujian lebih lanjut dengan dataset yang lebih besar dan lebih bervariasi agar hasil klasifikasi bisa lebih general dan representatif.
2. Melakukan eksperimen dengan teknik pembagian data yang berbeda, seperti *k-fold cross-validation*, dapat memberikan gambaran yang lebih akurat tentang performa model.
3. Melakukan penyetelan *hyperparameter* yang lebih mendalam atau menggunakan metode optimasi *hyperparameter* seperti *Grid Search* atau *Random Search* bisa dilakukan untuk meningkatkan performa model.
4. Membandingkan performa RBFN dengan algoritma lain seperti *Support Vector Machine* (SVM) atau *Neural Network* lainnya dapat memberikan perspektif tentang efektivitas metode yang digunakan.
5. Hasil dari penelitian ini dapat diimplementasikan dalam sebuah aplikasi yang digunakan untuk mendeteksi ujaran kebencian secara otomatis di media sosial, yang dapat bermanfaat bagi masyarakat luas atau pihak berwenang.

DAFTAR PUSTAKA

- Adhari, Z., & Sibaroni, Y. (2023). *Identifikasi Ujaran Kebencian pada Twitter Menggunakan Metode Convolutional Neural Network (CNN)*. 10(3), 3464–3474.
- Afrillia, Y. ... Siska, D. (2022). Analisis Sentimen Ciutan Twitter Terkait Penerapan Permendikbudristek Nomor 30 Tahun 2021 Menggunakan TextBlob dan Support Vector Machine. *G-Tech: Jurnal Teknologi Terapan*, 6(2), 387–394. <https://doi.org/10.33379/gtech.v6i2.1778>
- Ambarwati, I. (2023). Metode Radial Basis Function Neural Network (RBFNN) untuk Peramalan Kunjungan Wisatawan dengan Perbandingan Kombinasi Fungsi Pelatihan. *PRISMA, Prosiding Seminar Nasional Matematika 6. PRISMA, Prosiding Seminar Nasional Matematika*, 6, 687–693. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Anggreany, M. . (2022). *Confusion Matrix*. dari : [socs.binus.ac.id. https://socs.binus.ac.id/2020/11/01/confusion-matrix/](https://socs.binus.ac.id/2020/11/01/confusion-matrix/)
- Bara, E. A. B. ... Zahara Ginting, R. Z. (2022). Penelitian tentang Twitter. *Jurnal Edukasi Nonformal*, 3(2), 167–172.
- Chen, Y. ... Lin, H. (2019). Bioresource Technology Application of radial basis function artificial neural network to quantify interfacial energies related to membrane fouling in a membrane bioreactor. *Bioresource Technology*.
- Claudia, V. S., & Wijayanto, Y. R. (2020). Tindak Tutur Ujaran Kebencian (Hate Speech) Pada Komentar Forum Diskusi COVID-19 dalam Jejaring Sosial Facebook “Ini Kebumen.” *Prosiding Seminar Nasional Linguistik dan Sastra (SEMANTIKS)*, 533–542.
- Dharmawan, S. ... Perdana, N. J. (2023). Klasifikasi Ujaran Kebencian Menggunakan Metode FeedForward Neural Network (IndoBERT). *Jurnal Ilmu Komputer dan Sistem Informasi*, 11(1), 1–6. <https://doi.org/10.24912/jiksi.v11i1.24066>
- Dwitama, A. P. J. (2021). Deteksi Ujaran Kebencian Pada Twitter Bahasa Indonesia Menggunakan Machine Learning: Reviu Literatur. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, 1(1), 31–39. <https://doi.org/10.20885/snati.v1i1.5>
- Hakiem, M. ... Indriati. (2019). Klasifikasi Ujaran Kebencian pada Twitter Menggunakan Metode Naïve Bayes Berbasis N-Gram Dengan Seleksi Fitur Information Gain. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), 2443–2451. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4682>
- Kermani, M. Z. ... Hinkelmann, R. (2020). *Advanced Engineering Informatics Ensemble data mining modeling in corrosion of concrete sewer : A comparative study of network-based (MLPNN & RBFNN)*.

- Kusumasari, D., & Arifianto, S. (2020). Makna Teks Ujaran Kebencian Pada Media Sosial. *Jurnal Komunikasi*, 12(1), 1. <https://doi.org/10.24912/jk.v12i1.4045>
- Madhiarasan, M. (2020). Accurate prediction of different forecast horizons wind speed using a recursive radial basis function neural network. *Protection and Control of Modern Power Systems*, 5(1). <https://doi.org/10.1186/s41601-020-00166-8>
- Nikentari, N. ... #3, P. (2018). *JEPIN (Jurnal Edukasi dan Penelitian Informatika) Prediksi Kecepatan Angin Menggunakan Adaptive Neuro Fuzzy (ANFIS) dan Radial Basis Function Neural Network (RBFNN)*. 4(1), 70–75.
- Nugroho, K. . (2019). *Confusion Matrix Untuk Evaluasi Model Pada Supervised Learning*. Diakses dari : <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-padaunsupervised-machine-learning-bc4b1ae9ae3f>.
- Permatasari, D. I., & Subyantoro, S. (2020). Ujaran Kebencian Facebook Tahun 2017-2019. *Jurnal Sastra Indonesia*, 9(1), 62–70. <https://doi.org/10.15294/jsi.v9i1.33020>
- Prastowo, A. ... Ritha, N. (2021). Implementasi Algoritma Radial Basis Function Neural Network dalam Memprediksi Jumlah Angka Kelahiran. *Junal Sustainable*, 10, 1–6.
- Pratiwi, A. ... Tegal, S. (2022). Tinjauan Pustaka Sistematis: Klasifikasi Ujaran Kebencian Pada Sosial Media Dengan Algoritma Neural Network. *Anisa Pratiwi IJIR*, 3(1), 49–57.
- Puspaningrum, E. Y. ... Manggala, H. A. (2020). Penerapan Radial Basis Function Untuk Klasifikasi Jenis Tanah. *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, 15(1), 46–49. <https://doi.org/10.33005/scan.v15i1.1852>
- Ridwan, M., & Muzakir, A. (2022). Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan CNN-LSTM. *Teknomatika*, 12(02), 209–218.
- Roy, A., & Ojha, M. (2020). Twitter sentiment analysis using deep learning models. *2020 IEEE 17th India Council International Conference (INDICON)*, 1–6. <https://doi.org/10.1109/INDICON49873.2020.9342279>
- Safitri, M. (2020). Problem Ujaran Kebencian di Media Sosial Dalam Perspektif Al-Qur'an. *Angewandte Chemie International Edition*, 6(11), 951–952., 14(2), 5–24.
- Sandryan, M. K. ... Ratnawati, D. E. (2021). Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritma Backpropagation dan Term Frequency-Inverse Document Frequency. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(12), 5349–5355.
- Saputra, A. (2019). Survei Penggunaan Media Sosial Di Kalangan Mahasiswa Kota Padang Menggunakan Teori Uses and Gratifications. *Baca: Jurnal Dokumentasi Dan Informasi*, 40(2), 207. <https://doi.org/10.14203/j.baca.v40i2.476>

- Saskia, T. G. (2021). *Klasifikasi Hate Speech dan Abusive Language pada Twitter Bahasa Indonesia Dengan Metode Naive Bayes Classifier*. Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Septiawan, Y., & Chairani, C. (2023). Perbandingan Akurasi Metode Deteksi Ujaran Kebencian dalam Postingan Twitter Menggunakan Metode SVM dan Decision Trees yang Dioptimalkan dengan Adaboost. *Teknika*, 17(2), 287 – 299. <https://doi.org/10.30871/jaic.v5i2.3473>
- Sihombing, P. R., & Yuliati, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), 417–426. <https://doi.org/10.30812/matrik.v20i2.1174>
- Susanti, N. A. ... Hoiriyah, H. (2022). Klasifikasi Data Tweet Ujaran Kebencian Di Media Sosial Menggunakan Naive Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 538–543. <https://doi.org/10.36040/jati.v6i2.5174>
- Wahyuningrum, V. (2020). Penerapan Radial Basis Function Neural Network dalam Pengklasifikasian Daerah Tertinggal di Indonesia. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 12(1), 37. <https://doi.org/10.34123/jurnalasks.v12i1.250>



LAMPIRAN

Source code

```
# crawling data twitter menggunakan API V2

import tweepy # library

import pandas as pd # library

# Menggunakan token dari twitter API V2

bearer_token =
'AAAAAAAAAAAAAAAAAAAAAAAAAPjTrQEAAAAAhPs%2BWYy%2F7bIT7CVm
5gX9vKptx4E%3DLxd4dEjQBvfarnL4cAyLCczglUe4h9AZQjT7rDJoCYLingxUF
W'

client = tweepy.Client(bearer_token)

# kata kunci

kata_kunci = "ujaran kebencian -filter:retweets"

tweets = client.search_all_tweets(kata_kunci, max_results=600)

users_locs = [[tweet.text, tweet.user.username, tweet.created_at] for tweet in tweets]

tweet_text = pd.DataFrame(data = users_locs, columns = ["full_text", "username",
"datetime"])

tweet_text.to_csv('data_ujaran_kebencian.csv', index=False) # simpan file csv
```

LABELING DATA

Pembersihan Data

```
import pandas as pd

data = pd.read_csv('/content/data_ujaran_kebencian.csv')

data
```

1. Case Folding

```
import re

def caseFolding(teks):
```

```
teks = re.sub(r'@[A-Za-a0-9]+'+',',',teks)
teks = re.sub(r'#[A-Za-z0-9]+'+',',',teks)
teks = re.sub(r"http\S+",',',',teks)
teks = re.sub(r'[0-9]+'+',',',teks)
teks = re.sub(r"[-()\"#/@;:<>{}'+=~|.!?,_]", " ", teks)
teks = teks.strip(' ')
teks = teks.lower()
return teks
data['case_folding']= data['full_text'].apply(caseFolding)
data.head()
```

2. Tokenizaiton

```
import nltk
nltk.download('punkt_tab')
nltk.download('punkt')
from nltk.tokenize import word_tokenize
def tokenization(teks):
    teks = word_tokenize(teks)
    return teks
data['tokenization'] = data['case_folding'].apply(tokenization)
data.head()
```

3. Stopwords

```
from nltk.corpus import stopwords
nltk.download('stopwords')

daftar_stopword = stopwords.words('indonesian')
```

```
daftar_stopword = set(daftar_stopword)

def stopwords(words):

    return [word for word in words if word not in daftar_stopword]

data['stopwords'] = data['tokenization'].apply(stopwords)

data.head()
```

4. Stemming

```
!pip install sastrawi

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()

stemmer = factory.create_stemmer()

def stemming(text):

    text = ' '.join(stemmer.stem(word) for word in text if word in text)

    return text

data['stemming'] = data['stopwords'].apply(stemming)

data.head()

import nltk

nltk.download('vader_lexicon')

from nltk.sentiment.vader import SentimentIntensityAnalyzer

sia = SentimentIntensityAnalyzer()

polaritas = lambda x : sia.polarity_scores(x)

nilai_sia = data.stemming.apply(polaritas)

sia_data = pd.DataFrame(data = list(nilai_sia))

label = lambda x : 1 if x>0 else 0

sia_data['keterangan'] = sia_data.compound.apply(label)

dataset = data.join(sia_data.keterangan)
```

```
dataset.head()
```

5. TF-IDF

```
stem = dataset['stemming'].astype(str)

from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer()

vector = vectorizer.fit_transform(stem)

# membagi data untuk di seleksi

from sklearn.model_selection import train_test_split

X = vector

y = dataset.keterangan

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1,
random_state=0)
```

RBFClassifier

```
from RBFClassifier import RBFClassifier # panggil Libraries RBFNN

from sklearn.metrics import accuracy_score # panggil library

from sklearn.metrics import precision_score

from sklearn.metrics import recall_score

rbfnn = RBFClassifier()

rbfnn.fit(X_train, y_train)

y_pred = rbfnn.predict(X_test)

acc = accuracy_score(y_test, y_pred) # akurasi

pre = precision_score(y_test, y_pred, average="micro") # presisi

re = recall_score(y_test, y_pred, average="micro") # recall

print('Akurasi :', acc)

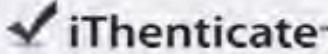
print('Presisi :', pre)
```

```
print('Recall :',re)

# confusion_matrix
from sklearn.metrics import confusion_matrix
print('Confusion Matrix :\n')
cm = confusion_matrix(y_test, y_pred)
print(cm)
```



1. Plagiasi

Similarity Report ID: oia:29477:96173922

PAPER NAME	AUTHOR
DANI GUNAWAN SITOMPUL_Skripsi Dani Gunawan Sitompul_aJlvXIFVIBNZmXI54vO4BgkFuqGZYcigsFMaYbH1.pdf	DANI GUNAWAN SITOMPUL

WORD COUNT	CHARACTER COUNT
6944 Words	41561 Characters
PAGE COUNT	FILE SIZE
43 Pages	998.6KB
SUBMISSION DATE	REPORT DATE
May 16, 2025 11:37 AM GMT+7	May 16, 2025 11:38 AM GMT+7

28% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 28% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database

Excluded from Similarity Report

- Submitted Works database
- Bibliographic material
- Cited material
- Abstract
- Small Matches (Less than 15 words)

Summary

2. Surat keterangan pembimbing



UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 (061) 7366878, 7360168, 7364348, 7366781, Fax. (061) 7366998 Medan 20223

Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, (061) 8225602, Fax. (061) 8226331 Medan 20122

Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : /FT/01.10/III/2025

21 Maret 2025

Lamp : -

Hal : Perpanjang SK Pembimbing Tugas Akhir

Yth. Pembimbing Tugas Akhir
Dr. Arnes Sembiring, ST, M. Kom
di
Tempat

Dengan hormat,
Sehubungan telah berakhirnya waktu masa berlaku SK pembimbing nomor 271/FT.6/01.10/VI/2024 tertanggal 14 Juni 2024 maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa berikut :

N a m a : Dani Gunawan

N P M : 188160010

Jurusan : Teknik informatika

Oleh karena itu kami mengharapkan kesediaan saudara :

Dr. Arnes Sembiring, ST, M. Kom (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

"Klasifikasi Ujaran Kebencian di Twitter menggunakan Algoritma Radial Basis Function Network (RBFN)"

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

Dekan,


Dr. Eng. Supriatno, ST, MT

3. Surat pengantar riset



UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781 Fax (061) 7366998 Medan 20223

Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602, Fax (061) 8226331 Medan 20122

Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 282 /FT.6/01.10/VI/2024

24 Juni 2024

Lamp : -

H a l : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Wakil Rektor Bidang Mutu Sumber Daya Manusia dan Perekonomian
Jln. Kolam No.1
Di
Medan

Dengan hormat, kami mohon kesediaan bapak kiranya berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	N A M A	N P M	PRODI
1	Dani Gunawan	188160010	Teknik informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di **Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area**.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul :

Klasifikasi Ujaran Kebencian di Twitter menggunakan Algoritma Radial Basis Function Network (RBFN).

Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya Surat ini.

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.



Dekan
Dr. Eng. Supriatno, ST, MT

Tembusan :

1. Ka. BPMPP

2. Mahasiswa

3. File

4. Surat izin riset



UNIVERSITAS MEDAN AREA
Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, 7366878, 7364348 📠 (061) 7369012 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602 📠 (061) 8226331 Medan 20122
Website: www.uma.ac.id E-Mail: univ_medanarea@uma.ac.id

Nomor : 947/UMA/B/01.7/VII/2024 03 Juli 2024
Lamp. : -
Hal : Izin Penelitian Dan Pengambilan Data Tugas Akhir

Kepada Yth.
Dekan Fakultas Teknik
Universitas Medan Area
di - Medan

Dengan hormat, sesuai dengan surat dari Dekan Fakultas Teknik Universitas Medan Area nomor : 282/FT.6/01.10/VI/2024 tertanggal 24 Juni 2024, perihal izin penelitian dan pengambilan data Tugas Akhir di Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area oleh mahasiswa sebagai berikut :

Nama : Dani Gunawan
NPM : 188160010
Program Studi : Teknik Informatika

Dengan Judul Penelitian "**Klasifikasi Ujaran Kebencian di Twitter Menggunakan Algoritma Radial Basis Function Network (RBFN)**".

Pada Prinsipnya disetujui yang bersangkutan untuk melaksanakan pengambilan data di Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area guna penyusunan Skripsi untuk kepentingan Ilmiah dan Akademik.

Demikian surat ini diterbitkan untuk dipergunakan sebagaimana mestinya.

Wakil Rektor Bidang Mutu SDM dan
Perekonomian,



Dr. Dedi Sahputra, S.Sos, MA

Tembusan :

1. Mahasiswa Ybs
2. File

5. Surat selesai riset



UNIVERSITAS MEDAN AREA

Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 42402994, Medan 20122
Website: www.uma.ac.id E-Mail: univ_medanarea@uma.ac.id

SURAT KETERANGAN
Nomor : 1191/UMA/B/01.7/VIII/2024

Rektor Universitas Medan Area dengan ini menerangkan bahwa :

Nama	: Dani Gunawan
No.Pokok Mahasiswa	: 188160010
Program Studi	: Teknik Informatika

Benar telah selesai Pengambilan Data di Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area dengan Judul Skripsi "**Klasifikasi Ujaran Kebencian di Twitter Menggunakan Algoritma Radial Basis Function Network (RBFN)**".

Dan kami harapkan Data tersebut kiranya dapat membantu yang bersangkutan dalam penyusunan Skripsi dan dapat bermanfaat bagi mahasiswa khususnya Fakultas Teknik Universitas Medan Area.

Demikian surat ini diterbitkan untuk dapat digunakan seperlunya

Medan, 14 Agustus 2024
a.n R e k t o r
Wakil Rektor Bidang Mutu SDM & Perekonomian,

Dr. Dedi Sahputra, S.Sos, MA

CC :
- Arsip

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 16/3/26

Access From (repository.uma.ac.id)16/3/26