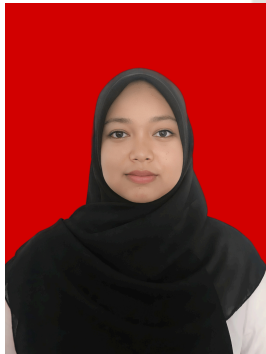


PENDEKATAN *LEXICON BASED FEATURES* UNTUK IDENTIFIKASI *CYBERBULLYING* MENGGUNAKAN ALGORITMA KLASIFIKASI

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi

Oleh:



RIMELDA ADHA

11850324476



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU

2023

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN
PENDEKATAN *LEXICON BASED FEATURES* UNTUK
IDENTIFIKASI *CYBERBULLYING* MENGGUNAKAN
ALGORITMA KLASIFIKASI

TUGAS AKHIR

Oleh:

RIMELDA ADHA
11850324476

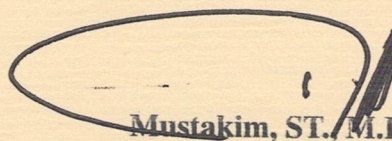
Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 18 Juli 2023

Ketua Program Studi



Eki Saputra, S.Kom., M.Kom.
NIP. 198307162011011008

Pembimbing



Mustakim, ST., M.Kom.
NIK. 130511023

LEMBAR PENGESAHAN

PENDEKATAN *LEXICON BASED FEATURES* UNTUK IDENTIFIKASI *CYBERBULLYING* MENGGUNAKAN ALGORITMA KLASIFIKASI

TUGAS AKHIR

Oleh:

RIMELDA ADHA

11850324476

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 06 Juli 2023

Pekanbaru, 06 Juli 2023

Mengesahkan,

Ketua Program Studi



Dekan

Dr. Hartono, M.Pd.

NIP. 196403011992031003



Eki Saputra, S.Kom., M.Kom.


NIP. 198307162011011008

DEWAN PENGUJI:

Ketua : Prof. Fitra Lestari Norhiza, ST., M.Eng., Ph.D. 

Sekretaris : Mustakim, ST., M.Kom.

Anggota 1 : Eki Saputra, S.Kom., M.Kom.

Anggota 2 : Tengku Khairil Ahsyar, S.Kom., M.Kom. 

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada peneliti. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin peneliti dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

UIN SUSKA RIAU

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 06 Juli 2023
Yang membuat pernyataan,

RIMELDA ADHA
NIM. 11850324476

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSEMBAHAN

Alhamdulillah, segala puji bagi Allah *Subhanahu Wa Ta'ala* yang telah melimpahkan rahmat, dan karunia-Nya memberikan kekuatan, kesabaran, dan ilmu pengetahuan dalam menyelesaikan Tugas Akhir ini. Shalawat dan salam senantiasa saya sampaikan kepada Nabi Muhammad *Shallallahu 'Alaihi Wa Salam* dengan mengucapkan *Allahuma Sholli'ala Sayyidina Muhammad Wa'ala 'Ali Sayyidina Muhammad*.

Sebagai tanda bakti, hormat, dan rasa terima kasih yang tidak terhingga saya persembahkan karya ini kepada orang tua saya **Santiago** dan **Sulastri**. Untuk ayahanda tercinta, terima kasih atas kasih sayang, nasihat, dan do'a yang telah diberikan. Untuk Ibunda tercinta, terima kasih atas dukungan dan kasih sayang yang selalu diberikan kepada saya sampai saat ini. Terima kasih telah membesarkan saya dengan baik, terima kasih telah menjadi orang tua yang suportif dan pengertian, senantiasa memanjatkan do'a, memberikan dukungan, semangat, solusi, dan nasihat sehingga saya dapat menyelesaikan Tugas Akhir ini. Saya sadar bahwa seluruh jasa ayahanda dan ibunda tidak bisa dibandingkan dengan apapun. Semoga Allah *Subhanahu Wa Ta'ala* merahmati.

Saya juga menyampaikan terima kasih yang tidak terhingga kepada **Nenny** yang telah merawat saya dari kecil, selalu setia mendengarkan keluh kesah saya, dan selalu mendukung setiap keputusan yang saya ambil. Selanjutnya untuk Adik-adik saya **Elno Budiman** dan **Sabina Az Zahra** yang selalu menemani dan mendengarkan keluh kesah. Salam rindu dan doa untuk Almarhum Mama dan Akek. Semoga Allah *Subhanahu Wa Ta'ala* merahmati.

Untuk segala bentuk kebaikan yang saya pernah terima, semoga Allah *Subhanahu Wa Ta'ala* membalas pemberinya dengan kebaikan, kebahagiaan, dan pahala yang berlipat ganda.

-Tidak ada balasan untuk kebaikan, selain kebaikan (pula)-
(QS. Ar Rahman : 60)

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

*Assalamu'alaikum Warahmatullahi Wabarakatuh. Alhamdulillah, puji syukur peneliti ucapkan pada Allah Subhanahu Wa Ta'ala karena atas izin-Nya peneliti dapat menyelesaikan penelitian sekaligus penulisan Laporan Tugas Akhir ini dengan judul Pendekatan *Lexicon Based Features* untuk Identifikasi *Cyberbullying* Menggunakan Algoritma Klasifikasi. Shalawat dan salam teriring kepada Nabi Muhammad Shallallahu 'Alaihi Wa Salam, Allahuma Sholli'ala Sayyidina Muhammad Wa'ala 'Ali Sayyidina Muhammad.*

Pada kesempatan ini peneliti ingin menyampaikan ucapan terima kasih kepada Allah *Subhanahu Wa Ta'ala* yang telah memberikan kekuatan dan ilmu pengetahuan, serta ucapan terima kasih khususnya kepada Ayahanda Santiagus dan Ibunda Sulastri. Tidak lupa pula peneliti mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. H. Hairunas, M.Ag sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd sebagai Dekan Fakultas Sains dan Teknologi.
3. Bapak Eki Saputra, S.Kom., M.Kom sebagai Ketua Program Studi Sistem Informasi sekaligus sebagai Penguji I Tugas Akhir peneliti yang telah memberikan banyak saran dan perbaikan dalam penyelesaian Tugas Akhir.
4. Ibu Siti Monalisa, ST., M.Kom sebagai Sekretaris Program Studi Sistem Informasi.
5. Bapak Tengku Khairil Ahsyar, S.Kom., M.Kom selaku Kepala Laboratorium Program Studi Sisem Informasi sekaligus Penguji II Tugas Akhir peneliti yang telah memberikan ide dan pandangan baru untuk penelitian dan penyelesaian Tugas Akhir.
6. Ibu Dr. Rice Novita, S.Kom., M.Kom sebagai Pembimbing Akademik peneliti yang telah dengan sabar memberikan arahan, motivasi, dan solusi dari awal perkuliahan sampai peneliti menyelesaikan penelitian, dan laporan ini.
7. Bapak Prof. Fitra Lestari Norhiza, ST., M.Eng., Ph.D sebagai Ketua Sidang Tugas Akhir peneliti yang telah bersedia membantu dan meluangkan waktu dalam Sidang Tugas Akhir peneliti.
8. Bapak Mustakim, ST., M.Kom selaku Pembimbing Tugas Akhir yang telah bersedia membimbing peneliti dalam penelitian, memberikan masukan dan saran yang membantu dalam penyelesaian Tugas Akhir ini.
9. Keluarga besar peneliti, Papap, Ibu, Nenny, Acik, Angah, Mamah dan Teteh, terima kasih atas segala bentuk dukungan yang diberikan kepada peneliti

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

selama menempuh pendidikan sarjana.

10. Bapak dan Ibu dosen Program Studi Sistem Informasi yang telah memberikan banyak ilmu dan arahan yang bermanfaat selama perkuliahan, dan segenap Pegawai dan Staf Program Studi Sistem Informasi yang telah mendukung dalam proses administrasi Tugas Akhir.
11. Kepada Nana Nurhaliza, Mulyana Widyastuti, dan Ummi Soleha, terima kasih sudah menjadi teman yang bisa dipercaya dan diandalkan, teman yang menerima, teman selayaknya keluarga.
12. Untuk seluruh teman-teman Program Studi Sistem Informasi Angkatan 2018 khususnya teman-teman seperjuangan dari Kelas B. Terima kasih untuk tahun-tahunnya, penghuni Grup Calon S.Kom Sukses.
13. Keluarga besar Puzzle Research Data Technology (PREDATECH) khususnya Bapak dan Ibu Dosen Pembimbing yang selalu sedia membantu dan membimbing. Kepada kakak, abang dan teman-teman yang tergabung. Terima kasih untuk seluruh pengalaman berharga, bantuan, ilmu, nasihat dan bimbingannya.
14. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu dalam penyelesaian Tugas Akhir ini.

Peneliti menyadari bahwa masih terdapat kekurangan pada penulisan Tugas Akhir ini. Oleh karena itu, peneliti mengharapkan kritik dan saran yang membangun dari pembaca. Peneliti berharap laporan ini dapat bermanfaat bagi pembaca. *Wassalamu'alaikum Warahmatullahi Wabarakatuh.*

Pekanbaru, 18 Juli 2023
Peneliti,

RIMELDA ADHA
NIM. 11850324476

UIN SUSKA RIAU

PENDEKATAN *LEXICON BASED FEATURES* UNTUK IDENTIFIKASI *CYBERBULLYING* MENGGUNAKAN ALGORITMA KLASIFIKASI

RIMELDA ADHA

NIM: 11850324476

Tanggal Sidang: 06 Juli 2023

Periode Wisuda:

Program Studi Sistem Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Salah satu bentuk perkembangan internet adalah *social content* yaitu Youtube. Namun, banyak interaksi penggunaannya tidak mempertimbangkan etika komunikasi dunia maya dapat menimbulkan *cyberbullying*. Adapun isu yang sering menjadi perbincangan antara lain *buzzer* politik dan isu agama yang dapat dianalisis untuk mengidentifikasi komentar *cyberbullying* menggunakan teknik klasifikasi pada *text mining* dan LBF. Pada penelitian ini dilakukan pendekatan LBF untuk mengidentifikasi *cyberbullying* menggunakan Algoritma PNN dan KNN. Pada penerapan menggunakan data pelabelan *lexicon*, nilai Akurasi yang paling baik diraih oleh PNN dan nilai Presisi yang paling baik diraih oleh KNN. Sedangkan pada penerapan menggunakan data pelabelan pakar, nilai Akurasi dan nilai Presisi yang paling baik diraih oleh Algoritma KNN. Nilai Akurasi tertinggi KNN sebesar 81,84%. Nilai Akurasi tertinggi PNN yaitu 78,01%. Selanjutnya nilai Presisi KNN tertinggi sebesar 89,22%, sedangkan nilai Presisi tertinggi PNN sebesar 61,77%. Berdasarkan perbandingan nilai parameter Akurasi dan Presisi, Algoritma KNN memiliki performa yang lebih baik daripada Algoritma PNN dalam mengklasifikasikan *cyberbullying*.

Kata Kunci: *Cyberbullying*, Klasifikasi, LBF, KNN, PNN.

UIN SUSKA RIAU

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEXICON-BASED FEATURES APPROACH TO IDENTIFY CYBERBULLYING USING CLASSIFICATION ALGORITHM

RIMELDA ADHA
NIM: 11850324476

Date of Final Exam: July 06th 2023
Graduation Period:

Department of Information System
Faculty of Science and Technology
State Islamic University of Sultan Syarif Kasim Riau
Soebrantas Street, No. 155, Pekanbaru

ABSTRACT

One form of internet development social content is YouTube. However, many user interactions do not consider the ethics of the cyber space communication can lead to cyberbullying. The issues that are often discussed include political buzzers and religious issues which can be analyzed to identify cyberbullying comments using classification techniques in text mining and LBF. then on this research is done the LBF approach to identifying cyber bullying using the algorithm PNN and KNN. on the application data using the results of labeling lexicon the best accuracy value achieved by PNN, but the most accurate value obtained by KNN. While in the application using data labelling results experts accuracy values and precision values are best achieved by the KNN algorithm. The highest accuracy value of KNN is 81.84%. the highest acurasai value of PNN is 78.01%. Based on a comparison of the Accuracy and Precision parameter values, the KNN Algorithm has better performance than the PNN Algorithm in classifying cyberbullying.

Keywords: *Cyberbullying, Classification, LBF, KNN, PNN.*

DAFTAR ISI

<p>© Hak cipta milik UIN Suska Riau</p> <p style="color: green; font-weight: bold;">State Islamic University of Sultan Syarif Kasim Riau</p>	
<p>Hak Cipta Dilindungi Undang-Undang</p> <p>1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:</p> <p style="margin-left: 20px;">a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.</p> <p style="margin-left: 20px;">b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.</p> <p>2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.</p>	
	<p>LEMBAR PERSETUJUAN ii</p> <p>LEMBAR PENGESAHAN iii</p> <p>LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL iv</p> <p>LEMBAR PERNYATAAN v</p> <p>LEMBAR PERSEMBAHAN vi</p> <p>KATA PENGANTAR vii</p> <p>ABSTRAK ix</p> <p><i>ABSTRACT</i> x</p> <p>DAFTAR ISI xi</p> <p>DAFTAR GAMBAR xv</p> <p>DAFTAR TABEL xvi</p> <p>DAFTAR SINGKATAN xvii</p> <p>1 PENDAHULUAN 1</p> <p style="margin-left: 20px;">1.1 Latar Belakang 1</p> <p style="margin-left: 20px;">1.2 Rumusan Masalah 4</p> <p style="margin-left: 20px;">1.3 Batasan Masalah 4</p> <p style="margin-left: 20px;">1.4 Tujuan Penelitian 5</p> <p style="margin-left: 20px;">1.5 Manfaat Penelitian 5</p> <p style="margin-left: 20px;">1.6 Sistematika Penulisan 5</p> <p>2 LANDASAN TEORI 7</p> <p style="margin-left: 20px;">2.1 <i>Data Mining</i> 7</p> <p style="margin-left: 20px;">2.2 <i>Text Mining</i> 8</p> <p style="margin-left: 20px;">2.3 <i>Text Preprocessing</i> 9</p> <p style="margin-left: 20px;">2.4 <i>Lexicon Based Features (LBF)</i> 10</p> <p style="margin-left: 20px;">2.5 <i>Term Frequency-Invers Document Frequency (TF-IDF)</i> 11</p> <p style="margin-left: 20px;">2.6 <i>Cyberbullying</i> 12</p>

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.7	<i>Buzzer</i> Politik	13
2.8	Isu Agama di Indonesia	13
2.9	<i>K-Fold Cross Validation</i>	13
2.10	K-Nearest Neighbor (KNN)	13
2.11	Probabilistic Neural Network (PNN)	14
2.12	<i>Confusion Matrix</i>	15
2.13	Penelitian Terdahulu	16
METODOLOGI PENELITIAN		18
3.1	Tahap Perencanaan	19
3.1.1	Wawancara Studi Kasus	19
3.1.2	Identifikasi Masalah	19
3.1.3	Menentukan Tujuan	19
3.1.4	Menentukan Batasan	19
3.1.5	Studi Literatur	19
3.2	Tahap Pengumpulan Data	19
3.3	Tahap <i>Text Preprocessing</i>	20
3.3.1	<i>Cleaning</i>	20
3.3.2	<i>Tokenizing</i>	20
3.3.3	<i>Filtering</i>	20
3.3.4	<i>Stemming</i>	20
3.4	Pelabelan Data	20
3.4.1	Pelabelan Data menggunakan LBF	20
3.4.2	Pelabelan Data oleh Pakar	21
3.5	Pembobotan Data menggunakan TF-IDF	21
3.6	Penerapan Algoritma KNN	21
3.6.1	Klasifikasi KNN menggunakan Data Pelabelan LBF	21
3.6.2	Klasifikasi KNN menggunakan Data Pelabelan Pakar	21
3.6.3	Perbandingan Akurasi dan Presisi Algoritma KNN	21
3.7	Penerapan Algoritma PNN	21
3.7.1	Klasifikasi PNN menggunakan Data Pelabelan LBF	21
3.7.2	Klasifikasi PNN menggunakan Data Pelabelan Pakar	22
3.7.3	Perbandingan Akurasi dan Presisi Algoritma PNN	22
3.8	Perbandingan Performa Algoritma KNN dan PNN	22
3.9	Visualisasi Kata pada Data Komentar	22
3.10	Hasil dan Dokumentasi	22
3.10.1	Kesimpulan dan Saran	22

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.10.2	Dokumentasi	22
4	ANALISIS DAN HASIL	23
4.1	Analisis Studi Kasus	23
4.2	Tahap Pengumpulan Data	24
4.3	Tahap <i>Text Preprocessing</i>	26
4.3.1	<i>Cleaning</i>	26
4.3.2	<i>Tokenizing</i>	27
4.3.3	<i>Filtering</i>	28
4.3.4	<i>Stemming</i>	29
4.4	Pelabelan Data	30
4.4.1	Pelabelan Data Menggunakan LBF	30
4.4.2	Pelabelan Data oleh Pakar	32
4.5	Pembobotan Data menggunakan TF-IDF	33
4.6	Penerapan Algoritma KNN	33
4.6.1	Klasifikasi KNN menggunakan Data Pelabelan LBF	33
4.6.2	Klasifikasi KNN menggunakan Data Pelabelan Pakar	35
4.6.3	Perbandingan Akurasi dan Presisi Algoritma KNN	36
4.7	Penerapan Algoritma PNN	36
4.7.1	Klasifikasi PNN menggunakan Data Pelabelan LBF	36
4.7.2	Klasifikasi PNN menggunakan Data Pelabelan Pakar	37
4.7.3	Perbandingan Akurasi dan Presisi Algoritma PNN	38
4.8	Perbandingan Performa Algoritma KNN dan PNN	38
4.9	Visualisasi Kata pada Data Komentar	39
4.9.1	Visualisasi Kata Kelas <i>Cyberbullying</i> pada Data LBF	39
4.9.2	Visualisasi Kata Kelas <i>Non-Cyberbullying</i> pada Data LBF	40
4.9.3	Visualisasi Kata Kelas <i>Cyberbullying</i> pada Data Pakar	40
4.9.4	Visualisasi Kata Kelas <i>Non-Cyberbullying</i> pada Data Pakar	41
5	PENUTUP	43
5.1	Kesimpulan	43
5.2	Saran	44

DAFTAR PUSTAKA

LAMPIRAN A DATA MENTAH A - 1

LAMPIRAN B TEXT PREPROCESSING B - 1

© Hak cipta milik UIN Suska Riau	LAMPIRAN C DATA HASIL PELABELAN <i>LEXICON</i>	C - 1
	LAMPIRAN D DATA HASIL PELABELAN PAKAR	D - 1
	LAMPIRAN E SURAT KETERANGAN WAWANCARA PAKAR	E - 1
	LAMPIRAN F SURAT KETERANGAN RISET	F - 1
	LAMPIRAN G DRAF WAWANCARA PAKAR	G - 1
	LAMPIRAN H DRAF WAWANCARA INSTANSI	H - 1
	LAMPIRAN I FOTO DOKUMENTASI WAWANCARA	I - 1

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR GAMBAR

2.1	Struktur <i>Probabilistic Neural Network</i>	14
3.1	Alur Metodologi Penelitian	18
4.1	Jumlah Data dari Video	24
4.2	Perbandingan Akurasi dan Presisi Algoritma KNN	36
4.3	Perbandingan Akurasi dan Presisi Algoritma PNN	38
4.4	Perbandingan performa Algoritma KNN dan PNN	39
4.5	Visualisasi Kata Kelas <i>Cyberbullying</i> pada Data LBF	39
4.6	Visualisasi Kata Kelas <i>Non-Cyberbullying</i> pada Data LBF	40
4.7	Visualisasi Kata Kelas <i>Cyberbullying</i> Pada Data Pakar	40
4.8	Visualisasi Kata Kelas <i>Non-Cyberbullying</i> pada Data Pakar	41

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR TABEL

2.1	<i>Confusion Matrix</i>	15
2.2	Penelitian Terdahulu	16
4.1	Data Kanal Youtube	24
4.2	Data Awal	25
4.3	Data Hasil <i>Cleaning</i>	26
4.4	Data Hasil <i>Tokenizing</i>	27
4.5	Data Hasil <i>Filtering</i>	28
4.6	Data Hasil <i>Stemming</i>	30
4.7	<i>Indonesian Sentiment Lexicon</i>	31
4.8	Hasil Pelabelan Data menggunakan LBF	31
4.9	Hasil Pelabelan Data oleh Pakar STKIP Rokania	32
4.10	Hasil pembobotan data dengan TF-IDF	33
4.11	Akurasi KNN pada Data LBF	34
4.12	Presisi KNN pada Data LBF	34
4.13	Akurasi KNN pada Data Pakar	35
4.14	Presisi KNN pada Data Pakar	35
4.15	Akurasi dan Presisi PNN pada Data LBF	36
4.16	Akurasi dan Presisi KNN pada Data Pakar	37

Hak Cipta Dilindungi Undang-Undang

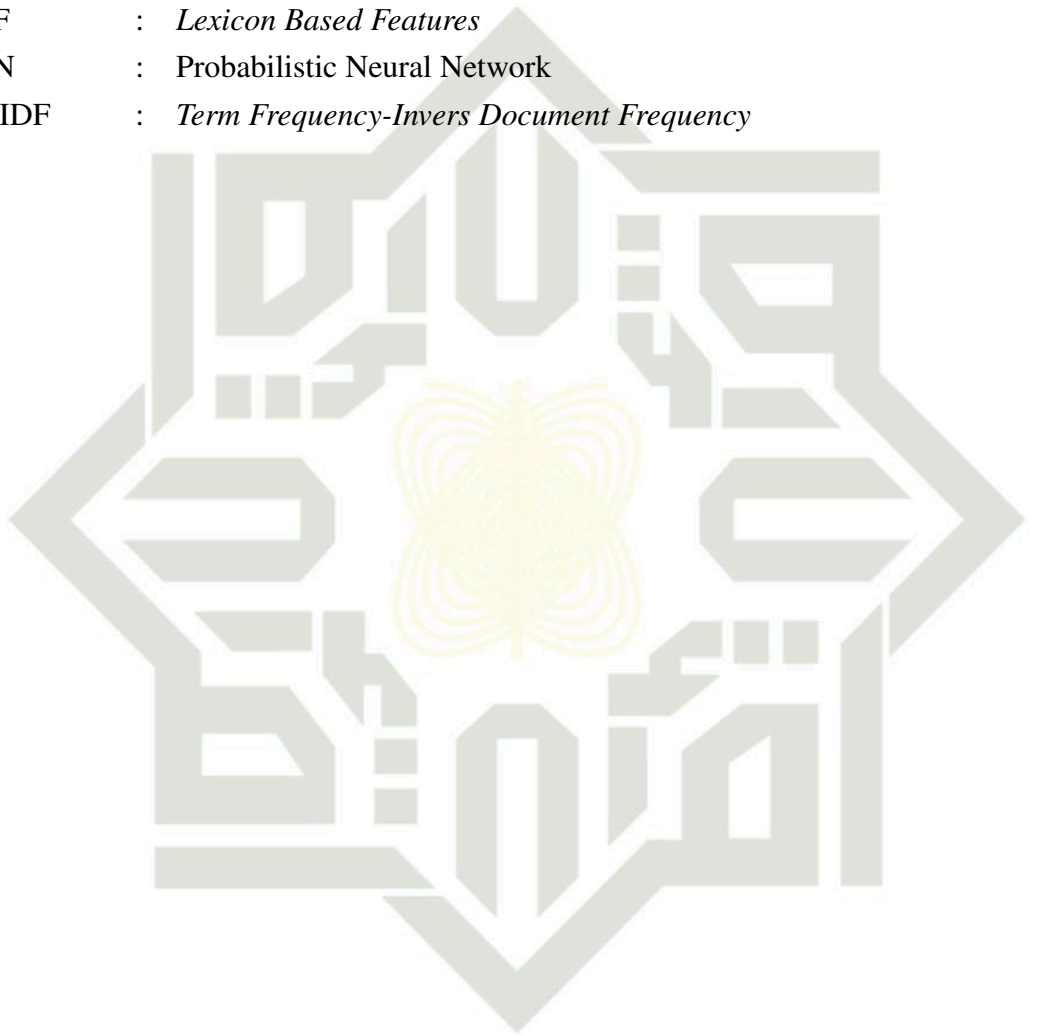
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR SINGKATAN

APJII	:	Asosiasi Penyelenggara Jasa Internet Indonesia
Diskominfo	:	Dinas Komunikasi dan Informatika
InSet	:	<i>Indonesian Sentiment Lexicon</i>
KBBI	:	Kamus Besar Bahasa Indonesia
KNN	:	K-Nearest Neighbor
LBF	:	<i>Lexicon Based Features</i>
PNN	:	Probabilistic Neural Network
TF-IDF	:	<i>Term Frequency-Invers Document Frequency</i>



UIN SUSKA RIAU

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Masyarakat digital atau *digital native* merupakan transformasi paradigma yang terjadi di masyarakat dengan kemudahan akses informasi dan interaksi pada era internet (Supratman, 2018). Dikutip dari Laporan Survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2019 sampai 2020 *Quartal 2* terdapat 196,7 juta pengguna internet dari 266,9 juta jiwa atau berkisar 73,7% dari total penduduk Indonesia. Dinas Komunikasi dan Informatika atau Diskominfo sebagai *leading sector* penggunaan dan pelayanan teknologi serta penyebaran informasi kepada publik, bertugas mengembangkan pelayanan dibidang komunikasi dan informasi seperti layanan konseling, layanan sosial, dan layanan informasi dan teknologi (Latifah dan Nurlukman, 2023). Diskominfo juga dapat melakukan sosialisasi untuk meningkatkan kecakapan literasi digital masyarakat terkait kejahatan *cyber* (Noval, Soecipto, dan Jamaludin, 2022). Permasalahan terkait kejahatan *cyber* diimbangi kurangnya sumber daya manusia dalam praktik pencegahan digital, koordinasi pihak terkait dalam melaksanakan upaya preventif, serta kurang maksimalnya sistem atau aplikasi *cyber* terintegrasi sebagai alat pencegahan kejahatan *cyber* di masyarakat umum (Sakban dan Bidaya, 2021). Adapun salah satu jenis kejahatan siber yang dibahas dalam Undang Undang Nomor 19 Tahun 2016 pada penjelasan pasal 45B tentang Informasi dan Transaksi Elektronik adalah *cyberbullying*.

Pemahaman dasar dari *bully* melibatkan intimidator dan korban yang dirundung baik secara fisik dan verbal yang memungkinkan untuk dilakukan secara langsung, dan tidak langsung. *Cyberbullying* atau perundungan siber diartikan sebagai tindakan dengan sengaja dan berulang mengirim pesan atau komentar yang bersifat ofensif kepada individu dan atau kelompok melalui media teknologi komunikasi (Margono, 2019). Latitude News memaparkan bahwa Indonesia adalah negara dengan tingkat *bullying* tertinggi setelah Jepang, di mana tindakan tersebut paling banyak dilakukan melalui jejaring sosial (Özel, Saraç, Akdemir, dan Aksu, 2017). Melalui survei terhadap lebih dari 10.000 orang yang 66% dari mereka adalah pengguna Youtube, ada 21% yang mengalami *cyberbullying*. Platform Youtube memiliki mekanisme interaksi penilaian video yang dapat dilakukan dengan rating, pengguna juga dapat memberikan komentar mereka terhadap konten yang disajikan (Khomsah, Aribowo, dkk., 2020). Namun, banyaknya interaksi pengguna yang tidak memahami dan atau mengesampingkan etika komunikasi dunia maya

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dapat menimbulkan tindakan *cyberbullying* (Candra dan Rozana, 2020). *Cyberbullying* di Youtube dapat dilihat dari komentar-komentar pada konten yang biasanya memuat pembahasan sensitif (Shalmarona, Chandhika, dan Soetikno, 2021). Salah satu topik yang menuai pro dan kontra di kalangan netizen Indonesia adalah tentang isu *buzzer* politik dan agama.

Individu yang memiliki pengaruh di media sosial dapat disebut sebagai *buzzer*. *buzzer* memiliki pengaruh yang kuat pula terhadap para pengikutnya sehingga dinilai efektif untuk melakukan kerja sama iklan pemasaran produk dengan peluang keberhasilan *marketing* yang tinggi bagi suatu *brand* (Juditha, 2019). Pada tahun 2012, peran *buzzer* yang semula ditujukan untuk *marketing* mulai bergeser ke bidang politik dan mengarah pada isu kepentingan publik atau tokoh tertentu. Hal tersebut memancing opini serta memengaruhi sudut pandang dan tanggapan netizen terhadap isu politik di sosial media (Ade, 2022) (Mustika, 2019). Politisi yang menggunakan *buzzer* politik untuk menggalakkan kampanye di media sosial dapat menimbulkan risiko perpecahan yang disebabkan oleh *hoax*, dan ujaran kebencian yang disebarkan antar pihak politisi yang bersaing (Felicia dan Loisa, 2019).

Disisi lain, agama adalah salah satu bagian penting dalam masyarakat Indonesia (Devit dan Arman, 2021). Perkembangan media sosial juga digunakan untuk menginternalisasi nilai-nilai agama dalam kehidupan. Media sosial telah mendampingi umat beragama dalam proses interaksi dan penyebaran berita dalam komunitas (Karim dan Wajdi, 2019). Cepatnya penyebaran dakwah dan isu keagamaan di media sosial menyorot atensi dari pengguna, reaksi yang ditimbulkan dapat berupa dukungan, kekecewaan, bahkan mengarah pada perundungan (Indainanto, 2020). Penyebaran pola teks dan ringkasan informasi komentar netizen tentang *buzzer* politik, dan isu keagamaan dapat dianalisis untuk mengidentifikasi komentar yang memuat unsur *cyberbullying* menggunakan *text mining* (Prabowo, Azizah, dkk., 2020).

Text mining adalah proses penambangan informasi dan pengetahuan dari data teks yang merupakan data tidak terstruktur, mulai dari pengumpulan, pengolahan, dan analisa hubungan pada masing-masing informasi (Heryanto dan Pramudita, 2020). Data teks yang dapat dengan mudah dipahami dengan manusia adalah sumber pengetahuan yang tidak tenilai sehingga diperlukan metode dan algoritma untuk memprosesnya secara efektif (Allahyari dkk., 2017), informasi yang berkualitas juga dapat diidentifikasi dengan memerhatikan *pattern* dan *trend* dari kegiatan pembelajaran pola statistik (Deolika, Kusrini, dan Luthfi, 2019). Contoh penelitian tentang *text mining* dilakukan oleh Syarif dkk (2019) tentang identifikasi *cyberbullying* pada komentar Instagram terkait pemilihan presiden Indonesia tahun

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2019 menggunakan metode *Lexicon Based Features* (LBF) dan *Naïve Bayes Classifier* diperoleh nilai Akurasi yang lebih baik pada metode *Naïve Bayes Classifier* 97%, Presisi 94%, *recall* 100%, dan *F-score* 97% (Syarif, Herdiani, dan Astuti, 2019). Selanjutnya penelitian (Hidayat, Mustakim, Fauzi, dan Syukra, 2019) tentang implementasi algoritma *K-Nearest Neighbors* (KNN) dan *Probabilistic Neural Network* (PNN) untuk menganalisis opini masyarakat terhadap toko *online* di Indonesia menggunakan data *tweet* yang diperoleh dari Twitter. Hasil yang diperoleh yaitu pada toko *online* Lazada, Akurasi KNN lebih baik dibandingkan dengan PNN yakni sebesar 71,57%. Begitu juga pada data Blibli juga menunjukkan bahwa Akurasi KNN lebih baik dibandingkan dengan PNN yakni 68,29%, sehingga disimpulkan bahwa pada kasus ini Algoritma KNN lebih baik dibanding PNN.

Penelitian ini menggunakan teknik Klasifikasi dengan dua pendekatan yaitu *learning based* dengan *machine learning* dan *lexicon based*. Pendekatan *Learning Based* akan menggunakan teknik klasifikasi yang merupakan teknik *supervised learning* menggunakan sumber data berupa set data pelatihan dan set data uji yang mengelompokkan data menjadi beberapa kelas juga melibatkan prediksi hasil berdasarkan input (Omar, 2019) (Kumar, Akhsi, Dabas, dan Hooda, 2020). Teknik ini juga dapat diartikan sebagai proses mengidentifikasi model yang memberi gambaran dan membedakan kelas dari setiap data (Nikmatun dan Waspada, 2019). LBF adalah salah satu metode tekstual dengan pendekatan berbasis leksikal di mana sentimen suatu kalimat ditentukan menggunakan sumber bahasa berupa kamus yang kemudian dilakukan analisis kecocokan lalu perhitungan frekuensi kemunculan suatu kata pada dokumen (Huang, Xie, Rao, Feng, dan Wang, 2020). Selain bersifat sederhana dan praktis untuk data yang berasal dari media sosial, pendekatan ini tidak membutuhkan data yang memiliki kelas melainkan melakukan penilaian berdasarkan polaritas (Hernikawati dkk., 2021).

PNN merupakan algoritma jaringan saraf *feed forward* turunan dari *Bayesian Network* (Alweshah, Qadoura, Hammouri, Azmi, dan AlKhalailah, 2020) yang difungsikan pada pembelajaran kuantitas vektor yang *layer* nya terdiri atas input, proses, dan *output* serta dapat digunakan untuk pengenalan pola diskret dari *dataset*. Dibandingkan dengan algoritma *neural network* lainnya, pada PNN perhitungan *probability density function* dilakukan dengan *eliminator parzen* untuk menekan kemungkinan terjadinya kesalahan klasifikasi *data set* (Jain dan Kaushal, 2018). Selanjutnya algoritma klasifikasi yang banyak digunakan adalah KNN karena tingkat efektivitas, kemudahan interpretasi, dan rendahnya waktu perhitungan (Moldagulova dan Sulaiman, 2017). Dalam hal pengenalan pola, KNN mengklasifikasikan data berdasarkan sampel ketetangaan dan pengukuran jarak dalam

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pada kategori tertentu (Chen, Zhou, Da Li, Zhang, dan Huo, 2020) (Senthilnayaki, Venkatalakshmi, dan Kannan, 2019). Pemilihan ruang dan penggunaan data latih serta nilai k juga sangat berpengaruh terhadap tingkat akurasi hasil klasifikasi algoritma ini (Azam, Ahmed, Sabah, Hussain, dkk., 2018).

Berdasarkan pemaparan diatas juga didukung oleh beberapa penelitian terdahulu, maka pada penelitian ini akan dilakukan pendekatan LBF untuk mengidentifikasi *cyberbullying* menggunakan algoritma PNN dan KNN.

1.2 Rumusan Masalah

Berdasarkan pemaparan pada latar belakang dan didukung beberapa penelitian terdahulu, dapat dirumuskan permasalahan penelitian yaitu bagaimana menerapkan algoritma PNN dan KNN untuk mengidentifikasi *cyberbullying* pada data komentar Youtube terkait *buzzer* politik dan isu agama di Indonesia dengan pendekatan LBF.

1.3 Batasan Masalah

Agar penelitian terarah, sistematis, dan sesuai dengan tujuan penelitian yang ingin dicapai, maka diperlukan batasan masalah. Berikut batasan masalah pada penelitian ini yaitu:

1. Data yang digunakan dalam penelitian ini adalah data komentar dalam bahasa Indonesia dari aplikasi Youtube dengan domain *buzzer* politik dan isu agama di Indonesia.
2. Data komentar yang digunakan berasal dari empat kanal terverifikasi yang difilter berdasarkan jumlah tayangan (*views*) per Oktober 2022, sebagai berikut:
 - (a) KOMPASTV, Rocky Gerung Ungkap Ciri-ciri Buzzer – ROSI.
Link <https://youtu.be/A5XMmenxNog>.
 - (b) Indonesia Lawyers Club, Saor Siagian: Apa yang Dilakukan Kepala Daerah Selalu Ditarik ke Politik dan Pakai Buzzer — ILC tvOne.
Link <https://youtu.be/G2HIc9Jaag4>.
 - (c) Yaqut Cholil Qoumas Ditunjuk Jadi Menteri Agama, PDI Perjuangan: Masalah Toleransi Harus Selesai.
Link <https://www.youtube.com/watch?v=upwCqGMB3QM>.
 - (d) Massa Protes Pernyataan Menteri Agama yang Dianggap Lukai Umat Muslim Indonesia #iNewsMalam 01/03.
Link https://www.youtube.com/watch?v=_fm54Ctr-dA.
3. Dilakukan tahap *text preprocessing* secara umum yaitu *cleaning*, *tokenizing*, *filtering*, dan *stemming*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. *Tools* yang digunakan untuk pengambilan dan pengolahan data adalah bahasa pemrograman Python dan aplikasi pengolah angka Microsoft Excel.
5. Teknik yang digunakan adalah teknik Klasifikasi dengan pelabelan oleh pakar dan pendekatan LBF, sedangkan algoritma yang digunakan yaitu PNN dan KNN.
6. Dalam analisis komentar digunakan kamus *sentiment lexicon* berbahasa Indonesia (InSet) dan diklasifikasikan menjadi dua kelas yaitu *cyberbullying* dan *non-cyberbullying*.
7. Tahap pembagian data latih dan data uji menggunakan 10 *K-Fold Cross Validation*.

1.4 Tujuan Penelitian

Adapun tujuan dari Tugas Akhir ini adalah:

1. Menerapkan Algoritma PNN dan KNN untuk mengidentifikasi *cyberbullying* dengan pendekatan LBF.
2. Mengetahui hasil perbandingan akurasi dan presisi dari Algoritma PNN dan KNN dalam mengklasifikasikan *cyberbullying*.
3. Memperoleh hasil visualiasi data komentar baik pada kelas *cyberbullying* dan *non-cyberbullying*.

1.5 Manfaat Penelitian

Adapun manfaat dari Tugas Akhir ini adalah:

1. Memperoleh hasil penerapan pendekatan LBF pada data komentar Youtube terkait *buzzer* politik, dan isu agama di Indonesia.
2. Memperoleh hasil penerapan algoritma PNN dan KNN pada data komentar Youtube terkait *buzzer* politik, dan isu agama di Indonesia.
3. Mendapatkan hasil perbandingan akurasi antara algoritma PNN dan KNN dari data komentar yang digunakan untuk mengetahui algoritma dengan performa paling baik.
4. Memperoleh hasil visualisasi data komentar yang dikategorikan *cyberbullying* dan *non-cyberbullying*.
5. Memberikan pengetahuan dan bahan pertimbangan bagi *stakeholder* dan instansi terkait untuk dapat melakukan penyampaian, dan atau penyebaran informasi kepada publik tentang *cyberbullying* serta menggunakan hasil analisis komputasi sebagai salah satu rekomendasi pada pengembangan aplikasi.

1.6 Sistematika Penulisan

Sistematika penulisan Laporan Tugas Akhir ini adalah sebagai berikut:

Hak Cipta Dilindungi Undang-Undang

BAB 1. PENDAHULUAN

BAB 1 pada Tugas Akhir ini berisi tentang: (1) Latar Belakang; (2) Rumusan Masalah; (3) Batasan Masalah; (4) Tujuan; (5) Manfaat; (6) Sistematika Penulisan.

BAB 2. LANDASAN TEORI

BAB 2 pada Tugas Akhir ini berisi tentang: (1) *Data Mining*; (2) *Text Mining*; (3) *Text Preprocessing*; (4) *Lexicon Based Features*; (5) *Term Frequency-Invers Document Frequency*; (6) *Cyberbullying*; (7) *Buzzer Politik*; (8) Isu Agama di Indonesia; (9) *K-Fold Cross Validation*; (10) *K-Nearest Neighbor*; (11) *Probabilistic Neural Network (PNN)*; (12) *Confusion Matrix*; (13) Penelitian Terdahulu.

BAB 3. METODOLOGI PENELITIAN

BAB 3 pada Tugas Akhir ini berisi tentang: (1) Tahap Perencanaan; (2) Tahap Pengumpulan Data; (3) Tahap Pra Proses Teks; (4) Tahap Hasil dan Analisis; (5) Tahap Dokumentasi.

BAB 4. ANALISIS DAN HASIL

BAB 4 pada Tugas Akhir ini berisi tentang: (1) Analisis Studi Kasus; (2) Analisa Kebutuhan Data; (3) Tahap Pra Proses Teks; (4) Pelabelan; (5) Pembobotan; (6) Penerapan Algoritma Klasifikasi; (7) Perbandingan Akurasi dan Presisi Algoritma KNN dan PNN; (8) Visualisasi Data.

BAB 5. PENUTUP

BAB 5 pada Tugas Akhir ini dipaparkan tentang kesimpulan dari Tugas Akhir dan saran untuk penelitian yang dapat dilakukan selanjutnya berdasarkan hasil yang didapat pada penelitian Tugas Akhir ini.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 2

LANDASAN TEORI

2.1 Data Mining

Data mining atau penambangan data adalah suatu proses analisis matematis dan statistik untuk mengetahui pola atau tren yang ada dalam suatu data dimana pola atau tren tersebut dapat didefinisikan dalam bentuk urutan prioritas, prediksi, dan pengelompokan (Kumar, Jain, Sumika, dan Chauhan, 2019). *Data mining* juga dapat diartikan sebagai proses analisis hubungan yang tidak diduga serta peringkasan data dengan teknik tertentu sehingga dapat dipahami dan digunakan oleh pemilik data (Widaningsih, 2019). Informasi dan pengetahuan yang diperoleh dari analisis pola dan tren data tersebut dapat digunakan sebagai acuan penilaian dan atau pembantu keputusan (Godoy, Diego, Tagumasi, Lerit, dan Costales, 2022). *Data mining* ditujukan untuk menanggapi analisis data dalam jumlah yang besar, data dengan dimensi yang tinggi, serta data heterogen (Fajrin, Maulana, Informatika, Batam, dan Soeprpto, 2018).

Terdapat beberapa teknik yang dapat digunakan dalam *data mining*, setiap teknik tersebut memiliki aturan dan metode yang berbeda serta pemecahan jenis masalah yang berbeda pula (Godoy dkk., 2022). Adapun teknik dalam *data mining* antara lain:

1. Asosiasi
Teknik ini pada dasarnya bertujuan menemukan pola yang didasarkan pada hubungan tiap variabel dalam satu waktu. Asosiasi disebut juga dengan teknik relasi karena teknik ini menggunakan hubungan antar item serta frekuensi kemunculan item berbeda pada data.
2. Klasifikasi
Klasifikasi atau pengkategorian digunakan untuk mengklasifikasi satu data baru yang belum diketahui kelasnya ke dalam suatu kelas yang telah didefinisikan pada data sebelumnya.
3. Klasterisasi
Teknik Klasterisasi merupakan proses pengelompokan atau segmentasi dengan menganalisis sejumlah atribut untuk menemukan kemiripan antar data sehingga ditemukan sekelompok data yang memiliki korelasi.
4. Prediksi
Teknik Prediksi digunakan untuk menganalisis kejadian masa lalu dengan kombinasi *data mining* termasuk didalamnya analisis tren, klasifikasi, pencocokan pola historis, dan teknik relasi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5. *Neural Network*

Neural Network atau jaringan saraf adalah teknik yang menggunakan kumpulan relasi neuron di satu atau beberapa lapisan. Teknik ini dianggap sebagai permodelan teknik prediksi yang kuat dan kompleks sebagai bagian dari kecerdasan buatan.

Beberapa jenis teknik baru juga telah dikembangkan seperti generalisasi, karakterisasi, evolusi, *pattern match*, dan visualialisasi data yang dapat digunakan baik untuk basis data relasional, transaksional, *object-oriented*, basis data spasial dan aktif, serta sistem informasi global (Chowdary dan Radhika, 2018).

2.2 *Text Mining*

Penambahan teks atau *text mining* mengacu pada ekstraksi informasi secara cepat dan efisien dari objek yang tidak dan atau semi terstruktur pada sebuah dokumen teks yang dapat digunakan baik oleh pemerintah maupun perusahaan untuk pemilihan solusi, peningkatan aktivitas dan strategi, serta pengambilan keputusan (Gao, Tan, dan Li, 2020) (Xie, Fu, Jin, Zhao, dan Cao, 2020). Proses ekstraksi informasi dalam *text mining* melalui beberapa analisis dengan menggunakan *Natural Language Processing* (NLP) (Lee, 2019), informasi lisan sekalipun dapat direpresentasikan ke dalam bentuk tekstual dan dapat digunakan *text mining* dalam hal interpretasi dan analisis kalimat, kata, frasa, dan pernyataan (Hassani, Beneki, Unger, Mazinani, dan Yeganegi, 2020). *Text mining* mencakup teknik linguistik, statistik, dan *machine learning* atau pembelajaran mesin (Pejić Bach, Krstić, Seljan, dan Turulja, 2019). Proses utama pada *text mining* dimulai dari ekstraksi, pembagian dan pembobotan, dilanjutkan dengan klasifikasi atau menentukan label kelas, dan visualisasi informasi (Fauziyyah, 2020).

Adapun *text mining* diklasifikasikan menjadi empat jenis, antara lain (Xie dkk., 2020):

1. *Contex-Based Text Mining*

Metode ini mengekstrak informasi dengan pendekatan analisis berdasarkan konteks informasi, dengan melakukan perhitungan nilai representasi dari *keyword* di setiap konteks atau mengekstrak fitur implisit yang relevan dengan kata atau opini dan kesamaan dalam konteks.

2. *Semantic-Based Text Mining*

Text mining berbasis semantik mengidentifikasi makna bahasa antar dokumen teks agar diperoleh hasil yang akurat.

3. *Sentiment Analysis-Based Text Mining*

Metode dengan dasar analisis sentimen ini mengidentifikasi opini, arah sen-

timent, dan kecenderungan emosi. Saat ini, *sentiment analysis-based text mining* adalah salah satu penelitian substansial pada NLP.

4. *Prediction-Based Text Mining*

Pada teknik prediksi digunakan untuk menganalisis kejadian masa lalu dengan kombinasi *data mining* termasuk didalamnya analisis tren, klasifikasi, pencocokan pola historis, dan teknik relasi.

5. *Content-Based Text Mining*

Metode *text mining* ini berfokus pada konten tekstual untuk optimalisasi aliran informasi *microblog*. Ciri khas metode ini adalah karakteristik yang menunjukkan minat konten dengan penggabungan *tag*, perhitungan dan pemeringkatan bobot aliran informasi.

2.3 *Text Preprocessing*

Pra-pemrosesan teks atau *text preprocessing* merupakan suatu tahapan sebelum analisis yang melibatkan transformasi teks, aglomerasi semantik yang bertujuan untuk mengurangi *sparsity* dan meningkatkan prediksi, meningkatkan informasi semantik, serta dapat membersihkan *dataset* dari *noise* dan menerjemahkan kata umum maupun slang (HaCohen-Kerner, Miller, dan Yigal, 2020) (Hickman, Thapa, Tay, Cao, dan Srinivasan, 2022). *Text preprocessing* data mentah melalui normalisasi, segmentasi, dan penguraian tata bahasa, di mana data teks akan diubah dalam bentuk yang logis dan dapat digunakan untuk membangun struktur gramatikal teks (Kononova dkk., 2021).

Text preprocessing merupakan tahapan penting dalam pengaplikasian *text mining*. Berikut proses utama *text preprocessing* (Rosid, Fitriani, Astutik, Mulloh, dan Gozali, 2020):

1. *Case Folding*

Proses *case folding* bertujuan untuk mengubah semua karakter pada dokumen teks menjadi karakter dengan format seragam (Atimi dan Pratama, 2022). Salah satunya mengubah dokumen teks ke format *lowercase* atau huruf kecil.

2. *Tokenizing*

Tokenisasi merupakan bagian dari segmentasi atau pemisahan dokumen teks menjadi kata-kata terpisah yang disebut token.

3. *Stop-Word Removal*

Proses ini bertujuan menghapus kata-kata seperti kata umum, tanda baca, dan konjungsi yang tidak memiliki makna yang berpengaruh pada kategorisasi (Alsaiddi, Sadeq, dan Abdullah, 2020).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. *Stemming*

Stemming adalah proses di mana dilakukan pemecahan ke dalam kata dasar atau penghapusan imbuhan yang bertujuan untuk mengurangi ukuran kosakata (Anandarajan dkk., 2019).

2.4 *Lexicon Based Features (LBF)*

Pendekatan berbasis leksikal menggunakan kamus berisi polaritas kata, artinya kata yang muncul didokumen akan dibandingkan dengan kata dalam kamus kemudian dilakukan perhitungan skor sentimen dengan total polaritas yang terkandung (Isabelle, Maharani, dan Asror, 2019). Pada pendekatan ini, tidak diperlukan data latih karena pelabelan disimpulkan berdasarkan akumulasi polaritas kata atau frasa dalam kalimat yang terbagi atas dua metode yaitu (Taj, Shaikh, dan Meghji, 2019):

1. *Dictionary-Based Methods*, atau metode berbasis kamus yang digunakan untuk mengidentifikasi kata positif dan kata negatif.
2. *Corpus-Based Methods*, yaitu metode dengan penggunaan korpus data di mana pelabelan dilakukan pada beberapa level mulai dari kata, frasa, kalimat, hingga eksplorasi analisis tingkat dokumen.

Penentuan nilai polaritas dalam suatu kalimat dilakukan dengan penentuan polaritas kata menggunakan Persamaan 2.1 dan Persamaan 2.2 (Syarif dkk., 2019).

$$\frac{1}{nk} \sum_{i=1}^{nk} \text{PositiveSentence}(i) \tag{2.1}$$

$$\frac{1}{nk} \sum_{i=1}^{nk} \text{NegativeSentence}(i) \tag{2.2}$$

Keterangan:

Nk : Total kata dalam kalimat

i : Skor positif/negatif dari kata i berdasarkan leksikon

Selanjutnya, dilakukan perhitungan nilai polaritas dokumen dengan perhitungan skor rata-rata polaritas setiap kalimat menggunakan Persamaan 2.3 dan Persamaan 2.4 (Isabelle dkk., 2019).

$$\frac{1}{ns} \sum_{i=1}^{ns} \text{PositiveSentence}(i) \tag{2.3}$$

$$\frac{1}{nk} \sum_{i=1}^{ns} \text{NegativeSentence}(i) \tag{2.4}$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Keterangan:

n_s : Total kalimat dalam dokumen

s_i : Skor positif/negatif dari kalimat i

Untuk menentukan sentiment kalimat i digunakan persamaan 2.5 (Syarif dkk., 2019).

$$SentenceSentimen = \begin{cases} positif & \text{if } S_{positif} > S_{negatif} \\ negatif & \text{if } S_{positif} < S_{negatif} \end{cases} \quad (2.5)$$

Pendekatan berbasis leksikal bersifat dinamis, namun terkait penggunaan leksikon untuk penelitian dengan dokumen berbahasa Indonesia masih minim dilakukan karena kurangnya sumber leksikon yang dibangun dalam Bahasa Indonesia, sehingga selain sumber daya untuk proses klasifikasi diperlukan juga sumber untuk proses penerjemahan (Pratama, Utami, dan Sunyoto, 2019). Kurangnya sumber leksikon tersebut dapat menghambat perkembangan pemahaman bahasa alami berbasis komputer (*Computer-based Natural Language Understanding*) karena harus dilakukan penyesuaian atau rekonstruksi dari bahasa asli agar mencerminkan makna dari perspektif bahasa terjemahan untuk mendukung fungsi algoritma dan analisis (Wijayanti dan Arisal, 2021).

Indonesia Sentiment Lexicon (InSet) adalah salah satu kamus leksikal yang digunakan sebagai tinjauan literatur sistematis dalam analisis dokumen berbahasa Indonesia yang memuat 3.609 kata positif dan 6.609 kata negatif dimana setiap katanya telah memiliki skor polaritas pada rentang -5 sampai +5 (Asri dan Wahyu, 2021). Selain memuat kata-kata baku, kamus leksikal InSet juga memuat kata-kata tidak baku yang karena dibangun menggunakan data dari media sosial Twitter yang dipilih karena paling mewakili media sosial di Indonesia (Firdaus, Asror, dan Herdiani, 2021).

2.5 Term Frequency-Invers Document Frequency (TF-IDF)

TF-IDF merupakan suatu metode yang digunakan untuk melakukan pembobotan kata yang didasarkan pada kemunculan kata (*term*) terhadap dokumen dengan mengalikan frekuensi *term* (TF) dengan *invers* dari frekuensi dokumen (IDF) (Simatupang dan Utomo, 2019). TF-IDF memiliki asumsi bahwa terdapat beberapa *term* yang sering muncul dibanding *term* lain pada dokumen secara umum. Semakin sering suatu *term* muncul, maka akan memengaruhi nilai bobot. TF-IDF dapat dihitung menggunakan Persamaan 2.6, Persamaan 2.7, dan Persamaan 2.8 (Riadi dan Widiandana, 2019).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$W(d,t) = TF(d,t) \quad (2.6)$$

Keterangan:

$TF(d,t)$: Frekuensi *term t* pada setiap dokumen

$$idf_t = \log 10\left(\frac{N}{df_t}\right) + 1 \quad (2.7)$$

Keterangan:

N : Jumlah total dari seluruh dokumen

idf_t : Jumlah dokumen yang memuat *term target*

$$w = tf_t \times idf_t \text{ atau } w = tf_t \times \log 10\left(\frac{N}{df_t}\right) + 1 \quad (2.8)$$

Keterangan:

w : Bobot atau hasil perkalian TF dan IDF

tf_t : Jumlah *term* pada setiap kalimat

idf_t : Jumlah dokumen yang memuat *term* pada *query*

2.6 Cyberbullying

Cyberbullying jika didefinisikan dari segi sastra merupakan interaksi melibatkan perilaku yang disengaja, kejam, dan repetitif dalam komunikasi sosial melalui media elektronik (Rosa dkk., 2019) yang bertujuan untuk mengganggu, melakukan penghinaan, dan pencemaran identitas sosial (Mutma, 2020). Jenis tindakan yang dilakukan bermacam-macam mulai dari pesan, tayangan video, dan atau komentar ofensif yang dapat menimbulkan konsekuensi psikologis kepada korban (Lozano-Blasco, Cortés-Pascual, dan Latorre-Martínez, 2020). Melalui survei terhadap lebih dari 10.000 orang yang 66% nya adalah pengguna Youtube, ada 21% dari mereka yang mengalami *cyberbullying*. *Cyberbullying* di Youtube dapat dilihat dari komentar-komentar pada konten yang biasanya memuat pembahasan sensitif (Shalmarona dkk., 2021). Adapun bentuk-bentuk *cyberbullying* yaitu (Rachmayanti dan Candrasari, 2022):

1. *Called Name*, yaitu pemberian nama atau panggilan yang konotasinya buruk.
2. *Image of Victim Spread*, atau ancaman penyebaran foto atau video korban.
3. *Threatened Physical Harm*, terkait ancaman keamanan dan keselamatan fisik.
4. *Opinion Slammed*, yaitu tanggapan atau pendapat yang merendahkan.

Tindakan *cyberbullying* terus terjadi di media sosial karena adanya kebe-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

basan publik untuk memberikan komentar yang bisa saja mengandung unsur-unsur *cyberbullying* (Pardede, Miftahuddin, dan Kahar, 2020).

2.7 Buzzer Politik

Buzzer dalam Bahasa Inggris berarti lonceng atau perangkat yang mengeluarkan bunyi guna menyebarkan sinyal tertentu. *Buzzer* akan melakukan promosi atau kampanye terhadap suatu produk atau *brand* sebagai bagian dari usaha meningkatkan penjualan produk tersebut (Mustika, 2019). *Buzzer* yang pada dasarnya adalah bentuk strategi promosi dari suatu produk, sejak tahun 2012 mengalami pergeseran peran sebagai agen *branding* tokoh politik di sosial media (Faulina, Chatra, dan Sarmiati, 2020). Politisi yang menggunakan *buzzer* politik untuk menggalakkan kampanye di media sosial dapat menimbulkan risiko perpecahan yang disebabkan oleh *hoax* dan ujaran kebencian yang disebarkan antar pihak politisi yang bersaing (Felicia dan Loisa, 2019).

2.8 Isu Agama di Indonesia

Ranah kehidupan beragama di Indonesia mengalami pergeseran dengan hadirnya koneksi sosial virtual yang sehingga terjadi perubahan representasi prana-ta keagamaan melalui situs, kanal dan mmeunculkan fenomena *online religion* (Fakhruroji, Rustandi, Busro, dkk., 2020). Indonesia merupakan negara dengan masyarakat heterogen yang menganut beberapa agama dimana Islam sebagai anutan mayoritas rakyatnya, dan dengan adanya perbedaan ini banyak terjadi konflik atau perselisihan antar umat beragama di Indonesia (Savitri dan Taher, 2018). Perselisihan terkait isu agama sering disebabkan oleh perbedaan pemikiran yang berujung pada klaim kebenaran dan saling serang (Jarir, 2019).

2.9 K-Fold Cross Validation

K-Fold Cross Validation merupakan salah satu teknik secara validasi silang dalam *subset* untuk melakukan pembagian data menjadi data latih dan data uji. Keuntungan dari penggunaan *K-Fold Cross Validation* dibanding jenis lain dari teknik validasi silang adalah semua data akan digunakan sebagai data uji dan data latih secara bergantian (Mustakim, 2020).

2.10 K-Nearest Neighbor (KNN)

KNN mengklasifikasikan data berdasarkan ketetanggaan yang paling dekat atau *nearest neighbor* yang merupakan salah satu metode tertua dari penalaran kelas, jarak (Xing dan Bei, 2019). Ketika ada data baru yang tidak diketahui kelasnya, KNN akan mengidentifikasi nilai *k* yang paling dekat dengan nilai data baru (Taunk, De, Verma, dan Swetapadma, 2019). Jika nilai *k* merupakan jumlah tetangga ter-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dekat dengan suatu data, maka penentuan klasifikasi sangat dipengaruhi oleh nilai k tersebut (Rizaldi dan Mustakim, 2020). Jarak data dan ketetanggannya dihitung menggunakan *Euclidean Distance* seperti pada Persamaan 2.9 (Hidayat dkk., 2019).

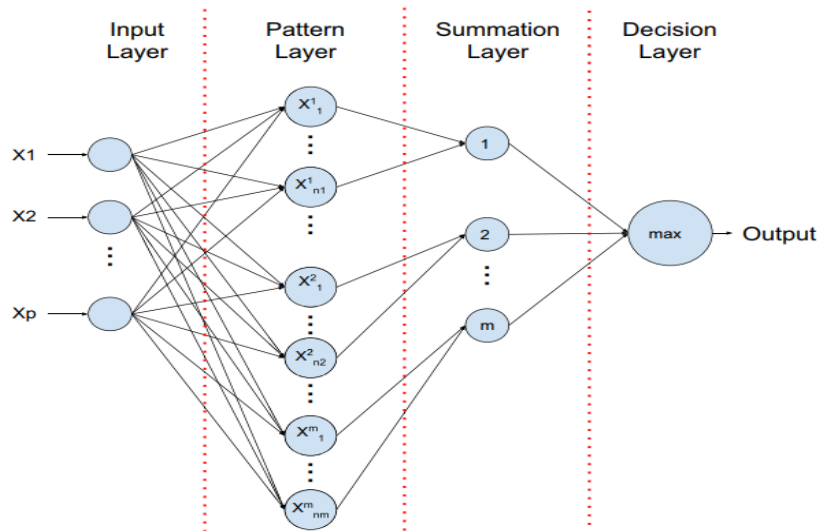
$$d(X_i, X_j) = \sqrt{\sum_{r=1}^n (a_r(X_i) - a_r(X_j))^2} \quad (2.9)$$

Keterangan:

- X_i : Data latih i
- X_j : Data uji i

2.11 Probabilistic Neural Network (PNN)

PNN merupakan algoritma jaringan saraf yang menggunakan teori *bayesian classification* (Adyati, Nasution, dan Wahyuningsih, 2019). Algoritma ini memiliki model pembelajaran *one pass* dan struktur paralel yang tinggi dengan waktu pembelajaran yang relatif singkat (Rady dan Anwar, 2019). PNN melakukan klasifikasi berdasarkan perhitungan jarak dan fungsi probabilitas peluang. Selain itu PNN juga cenderung insensitif terhadap data *outlier* (Siregar, Lubis, dan Ginting, 2021). Adapun struktur dasar PNN seperti pada Gambar 2.1 (Mohebali, Tahmassebi, Meyer-Baese, dan Gandomi, 2020).



Gambar 2.1. Struktur *Probabilistic Neural Network*

Berdasarkan Gambar 2.1 diatas, struktur PNN memiliki empat *layer* yang terdiri atas (Zhang, Chen, dan He, 2022):

1. *Input Layer*: neuron pada *layer* ini merepresentasikan neuron input variable prediktor dan tidak ada melakukan proses komputasi.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. *Pattern Layer*: sebagian besar perhitungan berlangsung pada *layer* ini kedekatan data latih dengan vektor input dihitung menggunakan Persamaan 2.10.

$$W_{i,j}(x) = \frac{1}{2\pi^d/2\sigma^d} \exp \left[\frac{|x-x_{ij}|^2}{2\pi^2} \right] \quad (2.10)$$

3. *Summation Layer*: disebut juga lapisan penjumlahan, *layer* ini menerima masukan dari setiap data latih. Adapun perhitungannya dapat dilakukan menggunakan Persamaan 2.11.

$$P_i(x) = \frac{1}{2\pi^d/2\sigma^d} \frac{1}{n_i} \sum_j^{n_i} \exp \left[\frac{|x-x_{ij}|^2}{2\pi^2} \right] \quad (2.11)$$

Penjumlahan *output* pada setiap *layer* dengan pola kelas yang sama dilakukan menggunakan persamaan 2.12

$$P(X | C_j) = \sum_{i=1}^N (gx) \quad (2.12)$$

4. *Output Layer*: penentuan kelas dari *input* sesuai pola berdasarkan *output* dari semua neuron pada *layer*. Adapun perhitungan *output layer* dapat menggunakan Persamaan 2.13.

$$Output(z) = \max P(X | C_j) \quad (2.13)$$

2.12 Confusion Matrix

Confusion matrix merupakan tabel yang digunakan untuk mengukur ketepatan dari suatu metode klasifikasi (Kumar dkk., 2020). Tabel *confusion matrix* untuk 2 kelas dapat dilihat pada Tabel 2.1.

Tabel 2.1. *Confusion Matrix*

Klasifikasi	Kelas Prediksi	
	Yes	No
Yes	True Positive (TP)	False Negative (FN)
No	False Positive (FP)	True Negative (TN)

Accuracy, *Precision*, dan *Recall* merupakan aspek penting dalam mengevaluasi kinerja algoritma klasifikasi dengan persamaan yang digunakan sebagai berikut (Azam dkk., 2018).

1. *Accuracy*: merupakan persentase dari nilai prediksi yang akurat dari keselu-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

ruhan nilai. Akurasi dapat dihitung menggunakan Persamaan 2.14.

$$Accuracy = ((TP + TN)/(TP + FN + FP + TN)) \times 100\% \quad (2.14)$$

2. *Precision*: dapat dihitung dengan Persamaan 2.15.

$$Precision = (TP/(TP + FP)) \times 100\% \quad (2.15)$$

3. *Recall*: dapat dihitung dengan Persamaan 2.16.

$$Recall = (TP/(TP + FN)) \times 100\% \quad (2.16)$$

2.13 Penelitian Terdahulu

Adapun penelitian terdahulu terkait metode yang digunakan dalam penelitian Tugas Akhir dapat dilihat pada Tabel 2.2.

Tabel 2.2. Penelitian Terdahulu

No	Peneliti dan Tahun	Judul	Hasil
1	Syarif dkk. (2019)	Identifikasi <i>Cyberbullying</i> Pada Komentar Instagram menggunakan metode <i>Lexicon-Based</i> dan <i>Naïve Bayes Classifier</i> (NBC).	<i>Lexicon-Based</i> dan <i>Naïve Bayes Classifier</i> dapat digunakan untuk pengklasifikasian komentar <i>cyberbullying</i> dan <i>non-cyberbullying</i> dengan hasil metode NBC memiliki performa yang lebih baik dengan akurasi 97%, presisi 94%, <i>recall</i> 100% dan <i>F1-Score</i> 97%.
2	Hidayat dkk. (2019)	Implementasi Algoritma <i>K-Nearest Neighbor</i> dan <i>Probabilistic Neural Network</i> untuk Analisis Opini Masyarakat Terhadap Toko Online di Indonesia.	Dengan metode pembagian data 10 K pada <i>K-Fold Cross Validation</i> didapat hasil perbandingan akurasi KNN lebih tinggi dibandingkan PNN yaitu 71,57% dan 68,29%.
3	Hertini dan Ismawati (2023)	<i>Techno Prevention</i> Sebagai Kebijakan <i>Non Penal</i> Dalam Upaya Penanggulangan Kejahatan <i>Cyberbullying</i> .	<i>Cyberbullying</i> merupakan salah satu kejahatan yang menjadi fenomena di Indonesia. Dilihat dari kebijakan <i>criminal</i> upaya penanggulangan <i>cyberbullying</i> perlu ditempuh dengan pendekatan teknologi. Untuk mewujudkan poin tersebut harus didukung oleh adanya aplikasi-aplikasi yang bisa membatasi tindak kriminalitas yang mengarah pada <i>cyberbullying</i> .

Tabel 2.2 Penelitian Terdahulu (Tabel lanjutan...)

No	Peneliti dan Tahun	Judul	Hasil
4	Fauziah, Yuwono, dan Aribowo (2021)	<i>Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review.</i>	Tho tahun 2021 <i>Lexicon VADER</i> memberikan kinerja lebih baik dari <i>SentiNetWord</i> , sementara pada penelitian lain performa Sentimen Leksikon Indonesia (<i>InSet</i>) lebih tinggi dari <i>SentiNetWord</i> dengan nilai <i>F-Measure</i> 0,598. Penelitian Musfiroh pada 2021 yang melakukan pengujian dengan 80% data latih dan 20% data uji menghasilkan akurasi 79,2%, presisi 72,9%, <i>recall</i> 62,8%. Penelitian Firdaus tentang sentiment umpan balik siswa dengan <i>InSet</i> menghasilkan akurasi 90,9%.
5	Waroka dkk. (2022)	Klasifikasi Ulasan Aplikasi Shopee Menggunakan Algoritma <i>Probabilistic Neural Network</i> dan <i>K-Nearest Neighbor</i> .	Dengan metode pembagian data <i>K-Fold Cross Validation</i> didapat bahwa akurasi KNN lebih tinggi dibanding PNN yaitu 77,85% untuk ulasan aplikasi, dan 85,71% untuk ulasan produk.
6	Balakrishna dan Rao (2019)	<i>Tomato Plant Leaves Disease Classification Using KNN and PNN.</i>	<i>Precision</i> dan <i>recall</i> menunjukkan Presisi daun terinfeksi adalah 422/500, untuk daun sehat 78/100. Sementara itu <i>recall</i> dari daun terinfeksi 422/422, untuk daun sehat 78/78. Kinerja klasifikasi PNN lebih baik dibandingkan KNN pada seluruh presentasi pelatihan.
7	Faruk dan Nafi'iyah (2020)	Klasifikasi Kanker Kulit Berdasarkan Fitur Tekstur, Fitur Warna Citra Menggunakan SVM dan KNN.	Dapat disimpulkan bahwa proses klasifikasi kanker kulit paling baik pada algoritma KNN dengan $K = 5$ akurasi 70,61%, sedangkan akurasi SVM 69,85% dan <i>Naïve Bayes</i> 67,27%.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

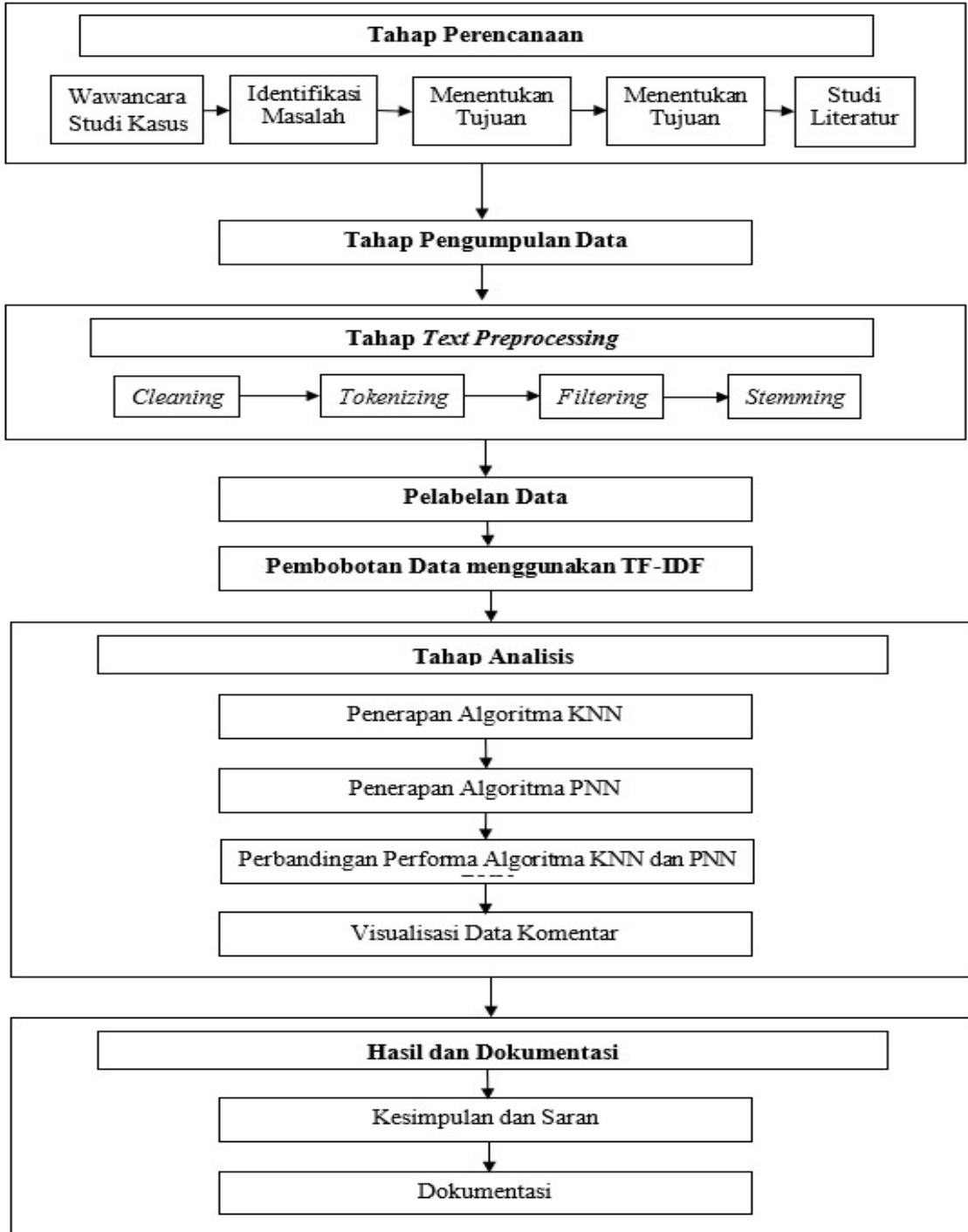
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 3

METODOLOGI PENELITIAN

Metodologi penelitian Tugas Akhir ini terdiri dari 7 proses utama. Adapun metodologi yang diterapkan pada Tugas Akhir dapat dilihat pada Gambar 3.1.



Gambar 3.1. Alur Metodologi Penelitian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.1 Tahap Perencanaan

Tahapan perencanaan merupakan tahap inisiasi ide awal penelitian yang kegiatannya meliputi identifikasi masalah, menentukan tujuan, menentukan batasan masalah, dan studi literatur.

3.1.1 Wawancara Studi Kasus

Sebagai awal tahap perencanaan, dilakukan wawancara dengan *stake holder* untuk menemukan arah rumusan permasalahan awal dalam penelitian.

3.1.2 Identifikasi Masalah

Identifikasi permasalahan dalam penelitian dilakukan berdasarkan pengamatan penulis terhadap perkembangan komunikasi dan interaksi dunia maya dan hasil wawancara yang dilakukan Dinas Komunikasi dan Informatika Kota Pekanbaru. Masalah yang peneliti amati yaitu maraknya aksi *bullying* yang terdapat di kolom komentar pada *platform* media sosial, khususnya terkait isu atau pembahasan tertentu, namun belum ada penyebaran informasi atau sosialisasi kepada publik dari dinas terkait yang dalam hal ini Diskominfo Kota Pekanbaru terkait *cyberbullying*.

3.1.3 Menentukan Tujuan

Untuk memperjelas arah dan fokus penelitian, pada tahap ini ditentukan tujuan dari penelitian. Pada tahap ini penulis menentukan metode yang sesuai dan algoritma yang akan diterapkan untuk memperoleh hasil yang ditentukan.

3.1.4 Menentukan Batasan

Batasan masalah dapat membantu peneliti menentukan batas dan lingkup penelitian baik dari studi kasus, data, teknik pengolahan, maupun algoritma yang digunakan.

3.1.5 Studi Literatur

Setelah permasalahan, tujuan, dan batasan masalah diidentifikasi maka pada tahap ini dilakukan studi literatur terkait pemahaman materi yang akan dibahas dalam penelitian. Adapun sumber studi meliputi publikasi ilmiah seperti Jurnal, Skripsi atau Thesis, Buku, dan Sumber tervalidasi lainnya.

3.2 Tahap Pengumpulan Data

Setelah mencari referensi penelitian terkait dan studi literatur, diputuskan bahwa data yang akan digunakan pada penelitian ini adalah data komentar media sosial Youtube yang diambil dari beberapa video dengan ketentuan yang telah ditentukan pada batasan masalah. Teknik yang digunakan untuk memperoleh data komentar video Youtube adalah *scraping* menggunakan salah satu *tools* pada ba-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

hasa pemrogramman Python yaitu *youtube-coment-downloader*. Pencarian video dilakukan dengan menggunakan kata kunci *buzzer* politik dan isu agama di Indonesia. Setelah dilakukan *scraping*, pada tahap ini data yang diperoleh dikonversikan ke format yang lebih rapi menggunakan Microsoft Excel.

3.3 Tahap *Text Preprocessing*

Sebelum data diolah menggunakan algoritma, dilakukan tahap pra proses teks atau *text preprocessing* untuk memastikan data yang akan digunakan sudah sesuai baik secara nilai maupun formatif. Berikut pemaparan tahapan *text preprocessing* yang digunakan.

3.3.1 *Cleaning*

Setelah data diimpor, dilakukan penyeragaman data dalam *lowercase* pembersihan data dengan menghapus karakter selain huruf seperti *hashtag* (#), *link*, dan tanda baca ((/), (;), (:), (“)).

3.3.2 *Tokenizing*

Pada tahap *tokenizing* atau tokenisasi, dilakukan pemisahan kata dari kalimat menjadi bentuk satuan kata (token).

3.3.3 *Filtering*

Filtering adalah tahap pengambilan kata-kata penting dari hasil tokenisasi serta penghilangan kata umum dan tidak memiliki makna seperti kata hubung dan kata depan. Pada tahap ini juga dilakukan *spell correction*.

3.3.4 *Stemming*

Pada proses *stemming*, kata di ekstrak ke dalam bentuk kata dasar atau dapat diartikan sebagai penghapusan imbuhan pada kata. Selain itu dilakukan pengelompokan untuk kata-kata dengan makna dasar yang sama namun memiliki bentuk berbeda.

3.4 Pelabelan Data

Pada tahap ini, pelabelan data dilakukan dengan 2 pendekatan, sehingga setelah melalui proses ini akan digunakan dua *dataset* penelitian dengan teknik pelabelan yang berbeda untuk di terapkan pada algoritma yang telah ditentukan. Pendekatan yang diterapkan yaitu:

3.4.1 Pelabelan Data menggunakan LBF

Pendekatan LBF diterapkan menggunakan Kamus InSet dan pelabelan oleh Pakar Bahasa untuk memberi label/kelas pada data. Pelabelan menggunakan LBF

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

merupakan teknik perhitungan sentimen kalimat dengan kamus yang tidak memerlukan data latih karena didasarkan pada akumulasi polaritas kata atau frasa (Tajdikk., 2019).

3.4.2 Pelabelan Data oleh Pakar

Selain menggunakan pendekatan LBF, dilakukan juga pelabelan data komentar oleh Pakar Bahasa Indonesia untuk melabelkan data. Dalam hal ini, pelabelan mengacu pada pemahaman atau perspektif pakar terhadap arti, makna, tata bahasa, kaidah kebahasaan, serta arah sentimen.

3.5 Pembobotan Data menggunakan TF-IDF

Sebelum dilakukan perhitungan menggunakan algoritma, dilakukan pembobotan pada data menggunakan TF-IDF. Adapun tahap *text preprocessing* hingga tahap pembobotan dilakukan pada data yang sama, yang membedakan 2 *dataset* yang digunakan hanya terletak pada teknik pelabelan dan penerapan algoritma.

3.6 Penerapan Algoritma KNN

Pada penerapan Algoritma KNN baik pada data pelabelan LBF maupun data pelabelan pakar, ditentukan jumlah dan nilai dari parameter k yang digunakan dengan merujuk pada penelitian terdahulu.

3.6.1 Klasifikasi KNN menggunakan Data Pelabelan LBF

Dilakukan penerapan Algoritma KNN dengan nilai k yang sudah ditentukan pada data pelabelan LBF.

3.6.2 Klasifikasi KNN menggunakan Data Pelabelan Pakar

Dilakukan penerapan Algoritma KNN dengan nilai k yang sudah ditentukan pada data pelabelan pakar.

3.6.3 Perbandingan Akurasi dan Presisi Algoritma KNN

Setelah Algoritma KNN diterapkan pada setiap *dataset*, diperoleh hasil perbandingan Akurasi dan Presisi dari Algoritma KNN.

3.7 Penerapan Algoritma PNN

Pada penerapan Algoritma PNN baik pada data pelabelan LBF maupun data pelabelan pakar, ditentukan jumlah dan nilai dari parameter penghalus *spread* yang digunakan dengan merujuk pada penelitian terdahulu.

3.7.1 Klasifikasi PNN menggunakan Data Pelabelan LBF

Dilakukan penerapan Algoritma PNN dengan nilai *spread* yang sudah ditentukan pada data pelabelan LBF.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.7.2 Klasifikasi PNN menggunakan Data Pelabelan Pakar

Dilakukan penerapan Algoritma PNN dengan nilai *spread* yang sudah ditentukan pada data pelabelan pakar.

3.7.3 Perbandingan Akurasi dan Presisi Algoritma PNN

Setelah Algoritma PNN diterapkan pada *dataset* pelabelan LBF dan pakar, diperoleh perbandingan Akurasi dan Presisi dari Algoritma PNN.

3.8 Perbandingan Performa Algoritma KNN dan PNN

Untuk mengetahui Algoritma yang memiliki performa terbaik diukur dengan Akurasi dan Presisi. Maka diperoleh perbandingan Akurasi dan Presisi Algoritma KNN dan Algoritma PNN yang diterapkan pada data pelabelan LBF dan data pelabelan pakar.

3.9 Visualisasi Kata pada Data Komentar

Visualisasi kata ditampilkan dalam bentuk *word cloud* berdasarkan tingkat frekuensi kata yang sering muncul pada setiap kelas pada *dataset*.

3.10 Hasil dan Dokumentasi

3.10.1 Kesimpulan dan Saran

Bagian ini memuat pemaparan hasil analisis dari penerapan setiap algoritma terhadap data dengan pelabelan LBF maupun pelabelan Pakar. Hasil analisis yang diperoleh merujuk kepada rumusan masalah dan tujuan penelitian yang ingin dicapai

3.10.2 Dokumentasi

Hasil penelitian dimuat ke dalam dokumen yang hasil akhirnya berupa Laporan Tugas Akhir. Selain itu, *sample data*, bukti dokumentasi administrasi dan wawancara studi kasus dapat dilihat pada Lampiran.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 5 PENUTUP

5.1 Kesimpulan

Kesimpulan dari penerapan pendekatan LBF untuk identifikasi *cyberbullying* menggunakan algoritma KNN dan PNN sebagai berikut:

1. Pada penerapan menggunakan data hasil pelabelan *lexicon* nilai akurasi tertinggi KNN 72,19%, selanjutnya nilai presisi tertinggi KNN 5 sebesar 84,07%. Penerapan menggunakan data hasil pelabelan pakar nilai akurasi tertinggi KNN bernilai 81,84%, dan presisi KNN tertinggi sebesar 89,22%.
2. Nilai tertinggi akurasi algoritma PNN pada data *lexicon* yaitu sebesar 73,33%, dan nilai presisi tertinggi yaitu 75,53%. Nilai tertinggi akurasi algoritma PNN pada data pakar sebesar 78,01%, dan untuk nilai presisi tertinggi yakni 61,77%.
3. Pada data penerapan menggunakan data hasil pelabelan *lexicon* nilai akurasi yang paling baik diraih oleh algoritma PNN, namun nilai presisi yang paling baik diraih oleh algoritma KNN. Sedangkan pada penerapan menggunakan data hasil pelabelan pakar nilai akurasi dan nilai presisi paling baik diraih oleh algoritma KNN.
4. Terdapat 35,12% data yang dikategorikan ke dalam kelas *cyberbullying* atau dengan 5 fekuensi tertinggi pada kelas *cyberbullying* data *lexicon*, 5 kata dengan frekuensi tertinggi yaitu “salah”, “pecat”, “perintah”, “rakyat”, dan “hina”. Sedangkan Kelas *non-cyberbullying* data *lexicon* mendominasi *dataset* dengan jumlah 64,87% dengan 5 fekuensi tertinggi pada kelas *non-cyberbullying* data *lexicon* yaitu “negara”, “suara”, “agama”, “menag”, dan “bang”.
5. Sebanyak 21,11% dari data pakar dikategorikan ke dalam kelas *cyberbullying*. Berdasarkan Gambar 4.6, diketahui 5 kata dengan frekuensi kemunculan paling tinggi antara lain “istana”, “jubir”, “negara”, “dungu”, dan “agama”. Sementara pakar melabelkan 78,88% data kedalam kelas *non-cyberbullying*, di mana diketahui kata dengan frekuensi kemunculan terbanyak antara lain kata “rakyat”, “islam”, “allah”, “suara”, dan “tidak”.
6. Hasil pengkategorian dan visualisasi kelas komentar dapat dijadikan rekomendasi dan bahan pertimbangan bagi Diskominfo khususnya Divisi Informasi dan Komunikasi Public (IKP) untuk menyusun dan memberikan informasi serta penyampaian edukasi kepada masyarakat tentang *cyberbullying*. Selanjutnya pada hasil penerapan algoritma KNN dan PNN, dari pe-

nilai menggunakan parameter akurasi dan presisi, dalam penerapan data hasil pelabelan *lexicon* dan pakar algoritma KNN memiliki performa paling baik sehingga algoritma KNN dapat menjadi salah satu rekomendasi algoritma yang dapat diterapkan pada pengembangan aplikasi di Diskominfo Kota Pekanbaru.

5.2 Saran

Saran yang ingin peneliti sampaikan untuk dapat dikembangkan pada penelitian selanjutnya sebagai berikut:

1. Penelitian selanjutnya dapat menggunakan *dataset* yang lebih banyak dan membandingkan penggunaan beberapa kamus *lexicon* untuk memperoleh akurasi dan performa yang lebih baik.
2. Penelitian selanjutnya dapat dilakukan dengan meningkatkan atau mengembangkan tahapan pra proses teks yang lebih mumpuni untuk memaksimalkan penerapan algoritma.
3. Penelitian yang dapat dilakukan selanjutnya adalah menerapkan algoritma yang lebih bervariasi untuk memaksimalkan klasifikasi dan mencapai performa yang lebih baik.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR PUSTAKA

- Ade, F. (2022). *Peran buzzer dalam proses pembentukan opini publik di new media (studi kasus pada akun media sosial buzzer)* (Unpublished doctoral dissertation). Universitas Andalas.
- Adyati, R. D., Nasution, Y. N., dan Wahyuningsih, S. (2019). Klasifikasi probabilistic neural network (pnn) pada data diagnosa penyakit demam berdarah dengue (dbd) tahun 2018. Dalam *Prosiding seminar nasional matematika dan statistika* (hal. 15–21).
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., dan Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*.
- Alsaidi, S. A., Sadeq, A. T., dan Abdullah, H. S. (2020). English poems categorization using text mining and rough set theory. *Bulletin of Electrical Engineering and Informatics*, 9(4), 1701–1710.
- Alweshah, M., Qadoura, M. A., Hammouri, A. I., Azmi, M. S., dan AlKhalailah, S. (2020). Flower pollination algorithm for solving classification problems. *Int J Adv Soft Comput Appl*, 12(1), 15–34.
- Anandarajan, M., Hill, C., Nolan, T., Anandarajan, M., Hill, C., dan Nolan, T. (2019). Cluster analysis: modeling groups in text. *Practical Text Analytics: Maximizing the Value of Text Data*, 93–115.
- Asri, J. S., dan Wahyu, S. (2021). Analisis sentimen menerapkan lexicon-learning based untuk melihat opini masyarakat mengenai protokol kesehatan dan perkembangan vaksin covid-19 di indonesia menggunakan dataset twitter. *Proceeding KONIK (Konferensi Nasional Ilmu Komputer)*, 5, 530–536.
- Atimi, R. L., dan Pratama, E. E. (2022). Implementasi model klasifikasi sentimen pada review produk lazada indonesia. *Jurnal Sains Dan Informatika*, 8(1), 88–96.
- Azam, M., Ahmed, T., Sabah, F., Hussain, M. I., dkk. (2018). Feature extraction based text classification using k-nearest neighbor algorithm. *IJCSNS Int. J. Comput. Sci. Netw. Secur*, 18(12), 95–101.
- Balakrishna, K., dan Rao, M. (2019). Tomato plant leaves disease classification using knn and pnn. *International Journal of Computer Vision and Image Processing (IJCVIP)*, 9(1), 51–63.
- Candra, R. M., dan Rozana, A. N. (2020). Klasifikasi komentar bullying pada instagram menggunakan metode k-nearest neighbor. *IT Journal Research and Development*, 5(1), 45–52.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Chen, Z., Zhou, L. J., Da Li, X., Zhang, J. N., dan Huo, W. J. (2020). The lao text classification method based on knn. *Procedia Computer Science*, 166, 523–528.

Chowdary, B., dan Radhika, Y. (2018). A survey on applications of data mining techniques. *International Journal of Applied Engineering Research*, 13(7), 5384–5392.

Deolika, A., Kusrini, K., dan Luthfi, E. T. (2019). Analisis pembobotan kata pada klasifikasi text mining. (*JurTI*) *Jurnal Teknologi Informasi*, 3(2), 179–184.

Devit, D., dan Arman, Z. (2021). Pertanggungjawaban pidana pelaku penistaan agama melalui media sosial ditinjau dari undang-undang informasi dan transaksi elektronik. *SCIENTIA JOURNAL: Jurnal Ilmiah Mahasiswa*, 3(1).

Fajrin, A. A., Maulana, A., Informatika, T., Batam, U., dan Soeprpto, J. (2018). Penerapan data mining untuk analisis pola pembelian konsumen dengan algoritma fpgrowth pada data transaksi penjualan spare part motor. *Kumpulan jurnaL Ilmu Komputer (KLIK)*, 5(1), 27–36.

Fakhruroji, M., Rustandi, R., Busro, B., dkk. (2020). Bahasa agama di media sosial: Analisis framing pada media sosial islam populer. *Jurnal Bimas Islam*, 13(2), 203–234.

Faruk, M., dan Nafi'iyah, N. (2020). Klasifikasi kanker kulit berdasarkan fitur tekstur, fitur warna citra menggunakan svm dan knn. *Telematika*, 13(2), 100–109.

Faulina, A., Chatra, E., dan Sarmiati, S. (2020). Peran buzzer dan konstruksi pesan viral dalam proses pembentukan opini publik di new media. *JRTI (Jurnal Riset Tindakan Indonesia)*, 7(1), 1.

Fauziah, Y., Yuwono, B., dan Aribowo, A. S. (2021). Lexicon based sentiment analysis in indonesia languages: A systematic literature review. Dalam *Rsf conf. ser. eng. technol* (Vol. 1, hal. 363–367).

Fauziyyah, A. (2020). *Gautama*. DH.

Fazrin, F., Pratiwi, O. N., dan Andreswari, R. (2023). Perbandingan algoritma k-nearest neighbor dan logistic regression pada analisis sentimen terhadap vaksinasi covid-19 pada media sosial twitter dengan pelabelan vader dan textblob. *eProceedings of Engineering*, 10(2).

Felicia, F., dan Loisa, R. (2019). *Peran buzzer politik dalam aktivitas kampanye di media sosial twitter. koneksi*, 2 (2), 352.

Firdaus, R., Asror, I., dan Herdiani, A. (2021). Lexicon-based sentiment analysis of indonesian language student feedback evaluation. *Indonesia Journal on Computing (Indo-JC)*, 6(1), 1–12.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Gao, X., Tan, R., dan Li, G. (2020). Research on text mining of material science based on natural language processing. Dalam *Iop conference series: materials science and engineering* (Vol. 768, hal. 072094).

Godoy, C. H., Diego, N. J. R., Tagumasi, R. E., Lerit, J. C., dan Costales, J. A. (2022). Cybersecurity scientometric analysis: Mapping of scientific articles using scopus api for data mining and webscraping. Dalam *2022 5th international conference on data science and information technology (dsit)* (hal. 1–6).

HaCohen-Kerner, Y., Miller, D., dan Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PloS one*, *15*(5), e0232525.

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., dan Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, *4*(1), 1.

Hernikawati, D., dkk. (2021). Kecenderungan tanggapan masyarakat terhadap vaksin sinovac berdasarkan lexicon based sentiment analysis (the trend of public response to sinovac vaccine based on lexicon based sentiment analysis). *Jurnal Iptekkom (Jurnal Ilmu Pengetahuan & Teknologi Informasi)*, *23*(1), 21–31.

Hertini, M. F., dan Ismawati, S. (2023). Techno prevention sebagai kebijakan non penal dalam upaya penanggulangan kejahatan cyberbullying. *Jurnal Komunikasi Hukum (JKH)*, *9*(1), 863–870.

Heryanto, A., dan Pramudita, R. (2020). Opini media sosial facebook terhadap produk hijab menggunakan metode text mining. *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS: Journal of Information System*, *4*(2), 168–177.

Hickman, L., Thapa, S., Tay, L., Cao, M., dan Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, *25*(1), 114–146.

Hidayat, A., Mustakim, M., Fauzi, M. Z., dan Syukra, I. (2019). Implementasi algoritma k-nearest neighbor dan probabilistic neural network untuk analisis opini masyarakat terhadap toko online di indonesia. Dalam *Seminar nasional teknologi informasi komunikasi dan industri* (hal. 254–261).

Huang, M., Xie, H., Rao, Y., Feng, J., dan Wang, F. L. (2020). Sentiment strength detection with a context-dependent lexicon-based convolutional neural network. *Information Sciences*, *520*, 389–399.

Indainanto, Y. I. (2020). Normalisasi kekerasan seksual wanita di media online.

Hak Cipta Dilindungi Undang-Undang

Jurnal Komunikasi, 14(2), 105–118.

Indrawan, D., Cakrawijaya, S. R., Wicaksono, B. D., Erni, E., dan Gata, W. (2021). Prediksi jumlah penonton video youtube menggunakan model deep neural network (dnn). *JISICOM (Journal of Information System, Informatics and Computing)*, 5(1), 94–98.

Irfan, M. (2020). Fenomena cyber-bullying dalam teknologi media baru (instagram) perspektif ilmu komunikasi. *Jurnal Public Relations (J-PR)*, 1(1), 15–22.

Isabelle, G., Maharani, W., dan Asror, I. (2019). Analysis on opinion mining using combining lexicon-based method and multinomial naïve bayes. Dalam *2018 international conference on industrial enterprise and system engineering (i-coiese 2018)* (hal. 214–219).

Jain, K., dan Kaushal, S. (2018). A comparative study of machine learning and deep learning techniques for sentiment analysis. Dalam *2018 7th international conference on reliability, infocom technologies and optimization (trends and future directions)(icrito)* (hal. 483–487).

Jarir, J. (2019). Solusi konflik agama di media sosial. *TOLERANSI: Media Ilmiah Komunikasi Umat Beragama*, 10(2), 106–116.

Juditha, C. (2019). Buzzer di media sosial pada pilkada dan pemilu indonesia. Dalam *Seminar nasional komunikasi dan informatika*.

Karim, A., dan Wajdi, F. (2019). Propaganda and da'wah in digital era (a case of hoax cyber-bullying against ulama). *KARSA: Jurnal Sosial dan Budaya Keislaman*, 27(1), 172–205.

Khomsah, S., Aribowo, A. S., dkk. (2020). Text-preprocessing model youtube comments in indonesian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), 648–654.

Kononova, O., He, T., Huo, H., Trewartha, A., Olivetti, E. A., dan Ceder, G. (2021). Opportunities and challenges of text mining in materials research. *Iscience*, 24(3).

Kumar, Akhsi, A., Dabas, V., dan Hooda, P. (2020). Text classification algorithms for mining unstructured data: a swot analysis. *International Journal of Information Technology*, 12, 1159–1169.

Kumar, Jain, J., Sumika, S., dan Chauhan, K. (2019). Knowledge discovery from data mining techniques. *International Journal of Engineering Research & Technology (IJERT)*, 7(12), 1–3.

Latifah, L., dan Nurlukman, A. D. (2023). Inovasi pelayanan publik elektronik sistem aplikasi bantuan kota (sabakota) kota tangerang. *Publik: Jurnal Manajemen Sumber Daya Manusia, Administrasi dan Pelayanan Publik*, 10(1),

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

167–180.

Lee, J. Y. (2019). Deep learning research trend analysis using text mining. *The International journal of advanced culture technology*, 7(4), 295–301.

Lozano-Blasco, R., Cortés-Pascual, A., dan Latorre-Martínez, M. P. (2020). Being a cybervictim and a cyberbully—the duality of cyberbullying: A meta-analysis. *Computers in human behavior*, 111, 106444.

Margono, H. (2019). *Analysis of the Indonesian cyberbullying through data mining: the effective identification of cyberbullying through characteristics of messages* (Unpublished doctoral dissertation). Victoria University.

Mohebali, B., Tahmassebi, A., Meyer-Baese, A., dan Gandomi, A. H. (2020). Probabilistic neural networks: a brief overview of theory, implementation, and application. *Handbook of probabilistic models*, 347–367.

Moldagulova, A., dan Sulaiman, R. B. (2017). Using knn algorithm for classification of textual documents. Dalam *2017 8th international conference on information technology (icit)* (hal. 665–671).

Mustakim, M. (2020). Efektivitas pembelajaran daring menggunakan media online selama pandemi covid-19 pada mata pelajaran matematika. *Al Asma: Journal of Islamic Education*, 2(1), 1–12.

Mustika, R. (2019). Pergeseran peran buzzer ke dunia politik di media sosial. *Diakom: Jurnal Media Dan Komunikasi*, 2(2), 144–151.

Mutma, F. S. (2020). Deskripsi pemahaman cyberbullying di media sosial pada mahasiswa. *Jurnal Common*, 4(1), 32–55.

Nikmatun, I. A., dan Waspada, I. (2019). Implementasi data mining untuk klasifikasi masa studi mahasiswa menggunakan algoritma k-nearest neighbor. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 10(2), 421–432.

Noval, S. M. R., Soecipto, S., dan Jamaludin, A. (2022). Modus operandi dan strategi pencegahan kejahatan perdagangan seksual anak secara daring. *Undang: Jurnal Hukum*, 5(2), 419–451.

Omar, S. H. (2019). *Deep learning approach for student performance prediction in e-learning* (Unpublished doctoral dissertation). Universiti Teknologi Malaysia.

Özel, S. A., Saraç, E., Akdemir, S., dan Aksu, H. (2017). Detection of cyberbullying on social media messages in turkish. Dalam *2017 international conference on computer science and engineering (ubmk)* (hal. 366–370).

Pardede, J., Miftahuddin, Y., dan Kahar, W. (2020). Deteksi komentar cyberbullying pada media sosial berbahasa inggris menggunakan naïve bayes classification. *Jurnal Informatika*, 7(1), 46–54.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Pejić Bach, M., Krstić, Ž., Seljan, S., dan Turulja, L. (2019). Text mining for big data analysis in financial sector: A literature review. *Sustainability*, 11(5), 1277.
- Prabowo, W. A., Azizah, F., dkk. (2020). Sentiment analysis for detecting cyberbullying using tf-idf and svm. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(6), 1142–1148.
- Pratama, B. T., Utami, E., dan Sunyoto, A. (2019). A comparison of the use of several different resources on lexicon based Indonesian sentiment analysis on app review dataset. Dalam *2019 international conference of artificial intelligence and information technology (icaiit)* (hal. 282–287).
- Price, M., dan Dagleish, J. (2010). Cyberbullying: Experiences, impacts and coping strategies as described by Australian young people. *Youth studies Australia*, 29(2), 51–59.
- Rachmayanti, A., dan Candrasari, Y. (2022). Perilaku cyberbullying di Instagram: Perilaku cyberbullying di Instagram. *Linimasa: Jurnal Ilmu Komunikasi*, 5(1), 1–12.
- Rady, E.-H. A., dan Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. *Informatics in Medicine Unlocked*, 15, 100178.
- Riadi, I., dan Widiandana, P. (2019). Mobile forensics for cyberbullying detection using term frequency-inverse document frequency (tf-idf). *vol*, 5, 68–76.
- Rizaldi, S. T., dan Mustakim, M. (2020). Perbandingan teknik pembagian data untuk klasifikasi sarana akses air pada algoritma k-nearest neighbor dan naïve Bayes classifier. Dalam *Seminar nasional teknologi informasi komunikasi dan industri* (hal. 130).
- Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P. C., Carvalho, J. P., Oliveira, S., ... Trancoso, I. (2019). Automatic cyberbullying detection: A systematic review. *Computers in Human Behavior*, 93, 333–345.
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., dan Gozali, H. A. (2020). Improving text preprocessing for student complaint document classification using sastrawi. Dalam *Iop conference series: Materials science and engineering* (Vol. 874, hal. 012017).
- Sakban, A., dan Bidaya, Z. (2021). Desain pola integrasi cyber dalam mengurangi kejahatan cyberbullying. *CIVICUS: Pendidikan-Penelitian-Pengabdian Pendidikan Pancasila dan Kewarganegaraan*, 9(1), 38–46.
- Savitri, S. F., dan Taher, A. (2018). Fenomena pemberitaan hoax terkait isu agama (studi perseteruan Ahok vs front pembela Islam (FPI) di media sosial Instagram). *Jurnal Ilmiah Mahasiswa Fakultas Ilmu Sosial & Ilmu Politik*, 3(1).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Senthilnayaki, B., Venkatalakshmi, K., dan Kannan, A. (2019). Intrusion detection system using fuzzy rough set feature selection and modified knn classifier. *Int. Arab J. Inf. Technol.*, 16(4), 746–753.
- Shalmarona, S., Chandhika, J., dan Soetikno, N. (2021). Cyberbullying in youtube: The role of psychological loneliness towards the emotional regulation. Dalam *International conference on economics, business, social, and humanities (icebsh 2021)* (hal. 269–274).
- Simatupang, M. P., dan Utomo, D. P. (2019). Analisa testimonial dengan menggunakan algoritma text mining dan term frequency-inverse document frequency (tf-idf) pada toko allmeeart. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1).
- Siregar, S. J., Lubis, A. I., dan Ginting, E. F. (2021). Penerapan neural network dalam klasifikasi citra permainan batu kertas gunting dengan probabilistic neural network. *Building of Informatics, Technology and Science (BITS)*, 3(3), 420–425.
- Sukma, F. A., Akbar, R. K., Azizah, N. N., dan Juliani, G. P. (2019). Konsep dan implementasi akad qardhul hasan pada perbankan syariah dan manfaatnya. *Amwaluna: Jurnal Ekonomi dan Keuangan Syariah*, 3(2), 148–162.
- Supratman, L. P. (2018). Penggunaan media sosial oleh digital native. *Jurnal Ilmu Komunikasi*, 15(1), 47–60.
- Syarif, R. D., Herdiani, A., dan Astuti, W. (2019). Identifikasi cyberbullying pada komentar instagram menggunakan metode lexicon-based dan naïve bayes classifier (studi kasus: Pemilihan presiden indonesia tahun 2019). *eProceedings of Engineering*, 6(2).
- Taj, S., Shaikh, B. B., dan Meghji, A. F. (2019). Sentiment analysis of news articles: A lexicon based approach. Dalam *2019 2nd international conference on computing, mathematics and engineering technologies (icomet)* (hal. 1–5).
- Taunk, K., De, S., Verma, S., dan Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. Dalam *2019 international conference on intelligent computing and control systems (iccs)* (hal. 1255–1260).
- Waroka, L. W. L., dkk. (2022). Klasifikasi ulasan aplikasi shopee menggunakan algoritma probabilistic neural network dan k-nearest neighbor: Shopee application review classification using probabilistic neural network algorithm and k-nearest neighbor. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 2(1), 49–58.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Widaningsih, S. (2019). Perbandingan metode data mining untuk prediksi nilai dan waktu kelulusan mahasiswa prodi teknik informatika dengan algoritma c4, 5, naïve bayes, knn dan svm. *Jurnal Tekno Insentif*, 13(1), 16–25.

Wijayanti, R., dan Arisal, A. (2021). Automatic indonesian sentiment lexicon curation with sentiment valence tuning for social media sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 20(1), 1–16.

Xie, X., Fu, Y., Jin, H., Zhao, Y., dan Cao, W. (2020). A novel text mining approach for scholar information extraction from web content in chinese. *Future Generation Computer Systems*, 111, 859–872.

Xing, W., dan Bei, Y. (2019). Medical health big data classification based on knn classification algorithm. *IEEE Access*, 8, 28808–28819.

Xu, G., Meng, Y., Qiu, X., Yu, Z., dan Wu, X. (2019). Sentiment analysis of comment texts based on bilstm. *Ieee Access*, 7, 51522–51532.

Zhang, Z., Chen, F., dan He, Y. (2022). Design and implementation of english teaching analysis system based on bp neural network algorithm. Dalam *International conference on applications and techniques in cyber intelligence* (hal. 455–461).

LAMPIRAN A

DATA MENTAH

Data mentah yang digunakan pada penelitian ini dapat dilihat pada Tabel berikut.

No	Username	Komputer
1	Luthfi Rahman	Buzzer itu nyata gw pernah nganterin temen gw buat jadi buzzer pilpres kemaren, dia di gaji 5jtaan. Jadi kalo istana bilang gk ada buzzer itu omong kosong
2	Cahya Agus Permana Channel	Salut sama pemikiran bung Rocky trims tlah menolong negara ini dengan kritik2
3	Fadlan Afriyandy	Fadrul ngeles. Rocky gerung mantap. Ini yg membuat semangat nonton tv bersemi kembali. Undang trus rocky gerung di debat2. Slm akal sehat
4	Abdul Muin	"Sudah jelas buzzer itu merusak demokrasi ! Jubir istana...begitu saja ngga bisa mengucapkan ! Itulah....Dungu !"
5	Yoppy Silitonga	Ini lucu banget. Asli gue suka cara Rocky mancing si Fadjoel. "Buzzer ngerusak demokrasi apa enggak?" Hahaha
6	Muhammad Abdi	BUZZER meyakinkan pada dunia bahwa dirinya tak ada. Itu adalah tipuan terhebat yang dilakukan IBLIS""."
7	Kurniawan rahmad-hani Dhani	Demokrasi menurun jauh di bawah pimpinan Jokowi, masih mending zaman SBY
8	Pande Lonthor	Mantap Roky Gerung, kau luar biasa cara berpikir dan menganalisa
9	bambang kurnaedi	Hebat jg rocky,semua bidang tau,salut lima jempol deh wkwkwk
10	Abdul Muin	"Jelas terlihat...siapa yg berargumen palsu karna duit...dan siapa yg berargumen asli sesuai hati nurani !"
11	bunda mala	Ngakak abis melihat kedunguan nya, joos RG
12	Asep Soo-man Tri	Ciri-ciri buzzer,adalah keras kepala,otak beku, selalu motong bicara orang dan minta di bayar
13	Ami Tabacan	FADJROEL SEDANG MEMUTAR BALIK KAN PAKTA DAN ISU....
15	Aghnie Hartanto	He's cOmeback !!! Kapan nih ILC berani ngundang RG lagi ?!
...
7259	rYĤUETrErH	Selama ibukota masih di jakarta disitulah demo akan selalu ada sebagai rutinitas ..
7260	Abu putin Al-Moskowi	Ini yg demo bukan di jakarta aja
7261	rYĤUETrErH	@Abu putin Al-Moskowi iya. Demonya cuma palingan di daerah jawa barat dan jakarta. Mungkin juga di sumbar

Seluruh data mentah yang digunakan dalam penelitian ini dapat dilihat pada link berikut <https://bit.ly/datamentah2023>.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN B

TEXT PREPROCESSING

Pada lampiran ini dimuat data hasil pra proses teks yang meliputi beberapa tahap. Adapun data hasil proses *cleaning* dapat dilihat pada Tabel .

No	Komentar
1	buzzer itu nyata pernah nganterin temen buat jadi buzzer pilpres kemaren dia gaji jtaan jadi kalo istana bilang ada buzzer itu omong kosong
2	salut sama pemikiran bung rocky trims tlah menolong negara ini dengan kritik
3	fadrul ngeles rocky gerung mantap ini membuat semangat nonton bersemi kembali undang trus rocky gerung debat slm akal sehat
4	sudah jelas buzzer itu merusak demokrasi jubir istanabegitu saja ngga bisa mengucapkan itulahdungu
5	ini lucu banget asli gue suka cara rocky mancing fadjroel buzzer ngerusak demokrasi apa enggak hahaha
6	Rbuzzer meyakinkan pada dunia bahwa dirinya tak ada itu adalah tipuan terhebat yang dilakukan iblis
7	demokrasi menurun jauh bawah pimpinan jokowi masih mending zaman sby
8	mantap roky gerung kau luar biasa cara berpikir dan menganalisa
9	hebat rockysemua bidang tausalut lima jempol deh wkwkwk
10	jelas terlihatsiapa berargumen palsu karna duitdan siapa berargumen asli sesuai hati nurani
11	ngakak abis melihat kedunguan nya joos
12	ciriciri buzzeradalah keras kepalaotak beku selalu motong bicara orang dan minta bayar
13	fadjroel sedang memutar balik kan pakta dan isu
14	rocky berfikir dan berbicara secara systemmatis runtut dan berlogika tapi fajrul hanya membantah dari pada pikiran rocky
15	hes comeback kapan nih ilc berani ngundang lagi
....
7250	selama ibukota masih jakarta disitulah demo akan selalu ada sebagai rutinitas
7251	ini demo bukan jakarta aja
7252	abu putin almoskowi iya demonya cuma palingan daerah jawa barat dan jakarta mungkin juga sumbar

Seluruh data hasil *cleaning* dalam penelitian ini dapat dilihat pada *link* berikut <https://bit.ly/datahasilcleaning2023>.

Proses selanjutnya adalah *tokenizing*. Berikut data hasil proses *tokenizing* dapat dilihat pada Tabel.

No	Komentar
1	['buzzer', 'itu', 'nyata', 'pernah', 'mengantarkan', 'teman', 'buat', 'jadi', 'buzzer', 'pil-pres', 'kemarin', 'dia', 'gaji', 'jtaan', 'jadi', 'kalo', 'istana', 'bilang', 'ada', 'buzzer', 'itu', 'omong', 'kosong']
2	['salut', 'sama', 'pemikiran', 'bung', 'rocky', 'Terima Kasih', 'Telah', 'menolong', 'ne-gara', 'ini', 'dengan', 'kritik']
3	['fadrul', 'ngeles', 'rocky', 'gerung', 'mantap', 'ini', 'membuat', 'semangat', 'menonton', 'bersemi', 'kembali', 'undang', 'terus', 'rocky', 'gerung', 'debat', 'salam', 'akal', 'sehat']
4	['sudah', 'jelas', 'buzzer', 'itu', 'merusak', 'demokrasi', 'jubir', 'Istana Begitu', 'saja', 'enggak', 'bisa', 'mengucapkan', 'itulahdungu']
5	['ini', 'lucu', 'banget', 'asli', 'gue', 'suka', 'cara', 'rocky', 'mancing', 'fadjroel', 'buzzer', 'ngerusak', 'demokrasi', 'apa', 'enggak', 'hahaha']
6	['buzzer', 'meyakinkan', 'pada', 'dunia', 'bahwa', 'dirinya', 'tak', 'ada', 'itu', 'adalah', 'tipuan', 'terhebat', 'yang', 'dilakukan', 'iblis']
7	['demokrasi', 'menurun', 'jauh', 'bawah', 'pimpinan', 'jokowi', 'masih', 'mending', 'za-man', 'surabaya']
8	['mantap', 'roky', 'gerung', 'kau', 'luar', 'biasa', 'cara', 'berpikir', 'dan', 'menganalisa']
9	['hebat', 'rockysemua', 'bidang', 'tausalut', 'lima', 'jempol', 'deh', 'wkwkwk']
10	['jelas', 'terlihat siapa', 'berargumen', 'palsu', 'karena', 'duitdan', 'siapa', 'berargumen', 'asli', 'sesuai', 'hati', 'nurani']
11	['mengakak', 'habis', 'melihat', 'kedunguan', 'nya', 'joos']
12	['ciciciri', 'buzzeradalah', 'keras', 'kepalaotak', 'beku', 'selalu', 'motong', 'bicara', 'o-rang', 'dan', 'meminta', 'bayar']
13	['fadjroel', 'sedang', 'memutar', 'balik', 'kan', 'pakta', 'dan', 'isu']
14	['rocky', 'berpikir', 'dan', 'berbicara', 'secara', 'systemmatis', 'runtut', 'dan', 'berlogika', 'tapi', 'fajrul', 'hanya', 'membantah', 'dari', 'pada', 'pikiran', 'rocky']
15	['hes', 'comeback', 'kapan', 'nih', 'ilc', 'berani', 'ngundang', 'lagi']
....
7214	['selama', 'ibukota', 'masih', 'jakarta', 'disitulah', 'demo', 'akan', 'selalu', 'ada', 'seba-gai', 'rutinitas']
7215	['ini', 'demo', 'bukan', 'jakarta', 'saja']
7216	['abu', 'putin', 'almoskowi', 'iya', 'demonya', 'cuma', 'palingan', 'daerah', 'jawa', 'barat', 'dan', 'jakarta', 'mungkin', 'juga', 'sumbar']

Data hasil *tokenizing* secara lengkap dapat dilihat pada *link* berikut <https://bit.ly/datatokenizing2023>.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Proses selanjutnya adalah *filtering*. Adapun data hasil proses *filtering* dapat dilihat pada Tabel.

No	Komentar
1	buzzer itu nyata pernah nganterin temen buat jadi buzzer pilpres kemaren dia gaji jtaan jadi kalo istana bilang ada buzzer itu omong kosong,buzzer nyata mengantarkan teman buzzer pilpres kemarin gaji jtaan kalo istana bilang buzzer omong kosong
2	salut sama pemikiran bung rocky trims tlah menolong negara ini dengan kritik,salut pemikiran rocky Terima Kasih Telah menolong negara kritik
3	fadrul ngeles rocky gerung mantap ini membuat semangat nonton bersemi kembali undang trus rocky gerung debat slm akal sehat,fadrul ngeles rocky mantap semangat menonton bersemi undang rocky debat salam akal sehat
4	sudah jelas buzzer itu merusak demokrasi jubir istanabegitu saja ngga bisa mengucapkan itulahdungu,buzzer merusak demokrasi jubir Istana Begitu itulahdungu
5	ini lucu banget asli gue suka cara rocky mancing fadjroel buzzer ngerusak demokrasi apa enggak hahaha,lucu banget asli gue suka rocky mancing fadjroel buzzer ngerusak demokrasi hahaha
6	buzzer meyakinkan pada dunia bahwa dirinya tak ada itu adalah tipuan terhebat yang dilakukan iblis,buzzer dunia tipuan terhebat iblis
7	demokrasi menurun jauh bawah pimpinan jokowi masih mending zaman s-by,demokrasi menurun pimpinan jokowi mending zaman surabaya
8	mantap roky gerung kau luar biasa cara berpikir dan menganalisa,mantap roky kau berpikir menganalisa
9	hebat rockysemua bidang tausalut lima jempol deh wkwkwk,hebat rockysemua bidang tausalut jempol deh wkwkwk
10	jelas terlihatsiapa berargumen palsu karna duitdan siapa berargumen asli sesuai hati nurani,terlihatsiapa berargumen palsu duitdan berargumen asli sesuai hati nurani
11	ngakak abis melihat kedunguan nya joos,mengakak habis kedunguan nya joos
12	ciriciri buzzeradalah keras kepalaotak beku selalu motong bicara orang dan minta bayar,ciriciri buzzeradalah keras kepalaotak beku motong bicara orang bayar
13	fadjroel sedang memutar balik kan pakta dan isu,fadjroel memutar pakta isu
14	rocky berfikir dan berbicara secara systemmatis runtut dan berlogika tapi fajrul hanya membantah dari pada pikiran rocky,rocky berpikir berbicara systemmatis runtut berlogika fajrul membantah pikiran rocky
15	hes comeback kapan nih ilc berani ngundang lagi,hes comeback nih ilc berani ngundang
....
7214	selama ibukota masih jakarta disitulah demo akan selalu ada sebagai rutinitas,ibukota jakarta disitulah demo rutinitas
7215	ini demo bukan jakarta aja,demo jakarta

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

7216 abu putin almoskowi iya demonya cuma palingan daerah jawa barat dan jakarta mungkin juga sumbar,abu putin almoskowi iya demonya palingan daerah jawa barat jakarta sumbar

Data hasil *filtering* secara lengkap dapat dilihat pada *link* berikut <https://bit.ly/datafiltering2023>.

Proses selanjutnya adalah *stemming*. Adapun data hasil proses *stemming* dapat dilihat pada Tabel.

No	Komentar
1	buzzer nyata antar teman buzzer pilpres kemarin gaji jtaan kalo istana bilang buzzer omong kosong
2	salut pikir rocky terima kasih telah tolong negara kritik
3	fadrul ngeles rocky mantap semangat tonton semi undang rocky debat salam akal sehat
4	buzzer rusak demokrasi jubir istana begitu itulahdunggu
5	lucu banget asli suka rocky mancing fadjroel buzzer ngerusak demokrasi hahaha
6	buzzer dunia tipu hebat iblis
7	demokrasi turun pimpin jokowi mending zaman surabaya
8	mantap roky pikir menganalisa
9	hebat rockysemua bidang tausalut jempol wkwkwk
10	terlihatsiapa argumen palsu duitdan argumen asli sesuai hati nurani
11	akak habis dunggu joos
12	ciriciri buzzeradalah keras kepalaotak beku motong bicara orang bayar
13	fadjroel putar pakta
14	rocky pikir bicara systemmatis runtut logika fajrul pikir rocky
15	comeback berani ngundang
....
7050	ibukota jakarta situ demo rutinitas
7051	demo jakarta
7052	putin almoskowi demo paling daerah jawa barat jakarta sumbar

Data hasil *stemming* secara lengkap dapat dilihat pada *link* berikut <https://bit.ly/datastemming2023>.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN C

DATA HASIL PELABELAN *LEXICON*

Data komentar dengan kelas data yang dilabelkan menggunakan *lexicon* dapat dilihat pada Tabel.

No	Komentar	Kelas
1	buzzer nyata antar teman buzzer pilpres kemarin gaji j-taan kalo istana bilang buzzer omong kosong	Cyberbullying
2	salut pikir rocky terima kasih telah tolong negara kritik	Non-Cyberbullying
3	fadrul ngeles rocky mantap semangat tonton semi undang rocky debat salam akal sehat	Non-Cyberbullying
4	buzzer rusak demokrasi jubir istana begitu itulahdunggu	Non-Cyberbullying
5	lucu banget asli suka rocky mancing fadjroel buzzer ngerusak demokrasi hahaha	Non-Cyberbullying
6	buzzer dunia tipu hebat iblis	Cyberbullying
7	demokrasi turun pimpin jokowi mending zaman surabaya	Non-Cyberbullying
8	mantap roky pikir menganalisa	Non-Cyberbullying
9	hebat rockysemua bidang tausalut jempol wkwkwk	Non-Cyberbullying
10	terlihat siapa argumen palsu duitdan argumen asli sesuai hati nurani	Cyberbullying
11	akak habis dunggu joos	Cyberbullying
12	ciriciri buzzeradalah keras kepalaotak beku motong bicara orang bayar	Non-Cyberbullying
13	fadjroel putar pakta	Non-Cyberbullying
14	rocky pikir bicara systemmatis runtut logika fajrul pikir rocky	Non-Cyberbullying
15	comeback berani ngundang	Cyberbullying
...
7050	ibukota jakarta situ demo rutinitas	Cyberbullying
7051	demo jakarta	Cyberbullying
7052	putin almoskowi demo paling daerah jawa barat jakarta sumbar	Cyberbullying

Seluruh data hasil pelabelan *lexicon* dapat dilihat pada *link* berikut <https://bit.ly/datapelabelanlexicon2023>.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN D

DATA HASIL PELABELAN PAKAR

Data komentar dengan kelas data yang dilabelkan oleh pakar bahasa dapat dilihat pada Tabel.

No	Komentar	Kelas
1	buzzer nyata antar teman buzzer pilpres kemarin gaji j-taan kalo istana bilang buzzer omong kosong	Non-Cyberbullying
2	salut pikir rocky terima kasih telah tolong negara kritik	Non-Cyberbullying
3	fadrul ngeles rocky mantap semangat tonton semi undang rocky debat salam akal sehat	Cyberbullying
4	buzzer rusak demokrasi jubir istana begitu itulahdunggu	Cyberbullying
5	lucu banget asli suka rocky mancing fadjroel buzzer ngerusak demokrasi hahaha	Non-Cyberbullying
6	buzzer dunia tipu hebat iblis	Non-Cyberbullying
7	demokrasi turun pimpin jokowi mending zaman surabaya	Cyberbullying
8	mantap roky pikir menganalisa	Non-Cyberbullying
9	hebat rockysemua bidang tausalut jempol wkwkwk	Non-Cyberbullying
10	terlihat siapa argumen palsu duitdan argumen asli sesuai hati nurani	Cyberbullying
11	akak habis dunggu joos	Cyberbullying
12	ciriciri buzzeradalah keras kepalaotak beku motong bicara orang bayar	Cyberbullying
13	fadjroel putar pakta	Cyberbullying
14	rocky pikir bicara systemmatis runtut logika fajrul pikir rocky	Cyberbullying
15	comeback berani ngundang	Non-Cyberbullying
...
7050	ibukota jakarta situ demo rutinitas	Non-Cyberbullying
7051	demo jakarta	Non-Cyberbullying
7052	putin almoskowi demo paling daerah jawa barat jakarta sumbar	Non-Cyberbullying

Seluruh data hasil pelabelan pakar dapat dilihat pada *link* berikut <https://bit.ly/datapelabelanpakar2023>.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN E

SURAT KETERANGAN WAWANCARA PAKAR



YAYASAN ROKAN RIAU RAYA
SEKOLAH TINGGI KEGURUAN DAN ILMU PENDIDIKAN ROKANIA
STKIP ROKANIA

Jl. Raya Pasir Pengaraian Km 15, Langkitin, Rambah Samo, Kab. Rokan Hulu Riau
Telp 082171718760 www.stkiprokania.ac.id, email : stkiprokania@gmail.com

Bukti Wawancara Pakar

Yang bertanda tangan dibawah ini :

Nama : Nuratika, S.Hum., M.Pd
Afiliasi : STKIP Rokania
Jabatan : Dosen Prodi Pendidikan Bahasa dan Sastra Indonesia
Hari/Tanggal : 30 Desember 2022

Menerangkan bahwa

Nama : Rimelda Adha
Afiliasi : Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim

Telah melakukan wawancara dan pengkajian data awal yang diperlukan dalam penelitian Tugas Akhir (Skripsi) pada tanggal 30 Desember 2022.

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya. Atas perhatian dan Kerjasama diucapkan terima kasih.

Pekanbaru, 3 Januari 2023

Nuratika, S.Hum., M.Pd

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN F

SURAT KETERANGAN RISET



PEMERINTAH KOTA PEKANBARU
DINAS KOMUNIKASI INFORMATIKA STATISTIK DAN PERSANDIAN
Komp. Perkantoran Walikota Pekanbaru Lt. III
Jalan Abdul Rahman Hamid Kel. Tuah Negeri Kec. Tenayan Raya
PEKANBARU - RIAU

SURAT KETERANGAN

Nomor: B.PN.01/KOMINFO/14/2023

Berdasarkan Surat dari Universitas Islam Negeri Sultan Syarif Kasim Riau Nomor : B-5491/F.V/PP.00.9/06/2023 tanggal 21 Juni 2023 perihal Mohon Izin Penelitian dan Pengambilan Data Tugas Akhir/Skripsi.

Kepala Dinas Komunikasi Informatika Statistik dan Persandian Kota Pekanbaru dengan ini menyatakan bahwa :

Nama	: Rimelda Adha
Nim	: 11850324477
Fakultas	: Sains dan Teknologi
Jurusan	: Sistem Informasi
Jenjang	: S1
Universitas	: Universitas Islam Negeri Sultan Syarif Kasim Riau

Mahasiswi tersebut di atas benar telah melakukan penelitian pada Bidang Pengendalian Informasi dan Komunikasi Publik (PIKP) Dinas Komunikasi Informatika Statistik dan Persandian Kota Pekanbaru dengan judul : **“Text Mining/Pendekatan Lexion Based Features Untuk Identifikasi Cyberbullying Menggunakan Algoritma Klasifikasi”**.

Demikian surat keterangan ini diberikan agar dapat dipergunakan sebagaimana mestinya.



Pekanbaru, 26 Juni 2023



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN G

DRAF WAWANCARA PAKAR

Nama Pakar : Nuratika, S.Hum., M.Pd.

Afiliasi : STKIP Rokania

Jabatan : Dosen Program Studi Pendidikan Bahasa dan Sastra Indonesia

Tanggal Surat : 3 Januari 2023

Peneliti :

Sebelumnya saya ucapkan terima kasih atas kesempatan untuk melakukan wawancara pada hari ini Bu. Saya Rimelda Adha, mahasiswa Sistem Informasi UIN SUSKA Riau yang sedang mengerjakan Tugas Akhir.

Narasumber :

Saya Nuratika, S.Hum., M.Pd. Biasa dipanggil Nuratika Rusli.

Peneliti :

Izin bertanya bu apakah sebelum ini ibu pernah menjadi pakar pada penelitian yang sejenis ?

Narasumber :

Saya pernah beberapa kali menjadi pakar untuk penelitian seperti yang adik lakukan, khususnya untuk pelabelan teks.

Peneliti :

Menurut ibu, apa yang menjadi dasar atau acuan utama dalam pengkategorian teks dari sudut pandang ibu sebagai pakar ?

Narasumber :

Kalau dari saya, tentu saja acuan utama untuk mengkategorikan teks adalah kaidah pemahaman konteks kebahasaan dari teks tersebut. Baru bisa kita menentukan arah sentimen teks tersebut.

Peneliti :

Dari beberapa penelitian khususnya pada pelabelan teks yang pernah ibu lakukan, apa saja kategori teks yang pernah ibu labelkan ?

Narasumber :

Untuk sentimen, biasanya dikategorikan ke kelas positif, negatif, atau netral. Pelabelan jenis ini beberapa kali pernah saya terapkan untuk mahasiswa yang penelitiannya sentimen analisis. Kemudian selain itu, penelitian dibidang Bahasa biasanya berkaitan pengkategorian berdasarkan deskripsi teks.

Peneliti :

Pada penelitian yang sedang saya lakukan, saya mengkategorikan teks ke dalam dua kelas yang berbeda bu, yaitu *Cyberbullying* dan *Non-Cyberbullying*. Apa

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

perbedaan hal tersebut dengan melabelkan ke positif, negatif, atau netral ?

Narasumber :

Cyberbullying disini secara umum saya pahami sebagai pesan elektronik yang sifatnya intimidatif serta memuat unsur entah itu ancaman, julukan buruk, atau pencemaran nama dan martabat lainnya. Untuk pengkategorian teks ke kelas *Cyberbullying* dan *Non-Cyberbullying*, hal pertama tentu saja pada perbedaan arah atau kecondongan pada teks ya. Jika pada label sentiment biasa terdapat tiga label yang artinya bisa condong ke positif, condong ke negatif, atau tidak keduanya. Sementara pada pelabelan penelitian yang adik lakukan hanya ada dua kelas yang artinya, apakah teks tersebut memuat unsur *bullying* atau tidak. Jadi hanya ada dua jawaban, ya atau tidak. Jadi jika suatu teks memiliki sentiment negatif, belum tentu teks tersebut mengarah ke teks *bullying*. Menurut saya untuk pelabelannya sendiri menjadi lebih mengerucut, hanya ada 2 jawaban ya atau tidak.

Peneliti :

Dalam pengkategorian teks ke dalam kelas *Cyberbullying* dan *Non-Cyberbullying*, apakah teks yang dikategorikan *cyberbullying* pasti memiliki arah sentiment negatif ?

Narasumber :

Kebanyakan iya. Namun ada aspek kebahasaan yang biasanya saya jadikan pertimbangan, seperti contohnya istilah satire dan sarkasme. Kalimat satire dan sarkasme terkadang tidak selalu mengarah ke sentiment negatif. Jadi dalam mengkategorikan teks ke dalam *Cyberbullying* dan *Non-Cyberbullying* harus dipahami bukan hanya arah sentimen, tapi makna kata per kata, juga secara keseluruhan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN H

DRAF WAWANCARA INSTANSI

Nama Narasumber : Adil Trisema

Afiliasi : Dinas Komunikasi, Informatika, Statistik Dan Persandian Kota
Pekanbaru

Jabatan : Pengelolaan Pengaduan Publik Pelaksana Bidang Pengelola Informasi
dan Komunikasi Publik

Tanggal Surat : 26 Juni 2023

Pertanyaan :

Sebelumnya saya ucapkan terima kasih atas kesempatan untuk melakukan wawancara pada hari ini pak. Saya Rimelda Adha, mahasiswa Sistem Informasi UIN SUSKA Riau yang sedang mengerjakan Tugas Akhir. baik pak, mungkin di awal saya akan menanyakan terlebih dari terkait tupoksi atau tugas dan fungsi utama dari Dinas Komunikasi dan Informatika, Statistik Dan Persandian (Diskominfo) ?

Pertanyaan :

Sebelumnya saya ucapkan terima kasih atas kesempatan untuk melakukan wawancara pada hari ini pak. Saya Rimelda Adha, mahasiswa Sistem Informasi UIN SUSKA Riau yang sedang mengerjakan Tugas Akhir. baik pak, mungkin di awal saya akan menanyakan terlebih dari terkait tupoksi atau tugas dan fungsi utama dari Dinas Komunikasi dan Informatika, Statistik Dan Persandian (Diskominfo) ?

Narasumber :

baik, Dinas Komunikasi dan Informatika atau biasa lebih sering kita sebut KOMINFO istilahnya merupakan perpanjangan tangan dari Kementerian Komunikasi dan Informatika, yang mana tupoksi utamanya secara kebijakan adalah merumuskan kebijakan nasional, kebijakan pelaksanaan, dan kebijakan teknis bidang komunikasi dan informasi, termasuklah didalamnya jaringan telekomunikasi, penyiaran, layanan teknologi informasi dan komunikasi, maupun layanan multimedia dan desiminasi informasi. Untuk Diskominfo sendiri, seperti yang sudah saya sampaikan merupakan perpanjangan tangan dari KOMINFO yang secara teknisnya bekerja di daerah yang dalam hal ini kabupaten atau kota untuk membantu pimpinan wilayah mengurus urusan pemerintahan bidang komunikasi publik, pengelolaan informatika, e- government, persandian, dan statistic informasi dan beberapa perihal dinas terkait.

Pertanyaan :

Apakah bapak bisa beri contoh dari fungsi dan tugas utama yang sudah di paparkan diatas, apa saja contoh pelaksanaan kebijakan atau teknis ?

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

Narasumber :

Baik, mungkin maksudnya contoh pekerjaan terkait komunikasi dan informatika yang dilakukan oleh Diskominfo. Contohnya seperti yang saya sudah sebutkan yaitu pengelolaan E-Government, termasuk disana secara teknis menyiapkan kebutuhan data, informasi, maupun teknologi informasi yang diperlukan untuk mendukung berjalannya kebijakan.

Pertanyaan :

Selain tugas dan fungsi yang sudah bapak sebutkan, mengerucut kepada bidang Pengelolaan Informasi dan Komunikasi Publik. Apa tugas dan tupoksi Divisi ini pak ?

Narasumber :

Pada dasarnya tugas dan tupoksi utama dari bidang Pengelolaan Informasi dan Komunikasi Publik atau biasa disingkat PIKP adalah mengelola informasi yang akan disampaikan kepada publik guna memberitahu terkait program kerja atau kegiatan dari pemerintah.

Pertanyaan :

Dalam pengelolaan dan penyampaian informasi ke masyarakat, apa media informasi yang digunakan dan bentuk informasi yang dipublikasikan ?

Narasumber :

Baik, ada beberapa kategori konten yang dipublikasikan yaitu konten dari pemerintah pusat, dalam hal ini Kominfo RI dan Konten besutan diskominfo kota yang biasanya berisi informasi *campaign* dan kegiatan strategis lainnya. Untuk penyampaiannya kepada masyarakat dapat melalui beberapa jenis. Yang pertama *by text* yaitu media informasi berupa teks yang biasanya kami publikasikan di website dalam bentuk berita. Selanjutnya *by photo* yang bisa disajikan dalam bentuk *feeds*, lalu *by video* yang bisa diakses baik di media sosial Tik Tok, *reels* Instagram, dan Youtube.

Pertanyaan :

Bagaimana proses bisnis dalam penyampaian informasi dan komunikasi publik pak ?

Narasumber :

Untuk masukan atau input nya sendiri sudah pasti berita, dan informasi, kemudian dalam prosesnya dipublikasikan kemasyarakat dengan output yang diharapkan yaitu meningkatnya partisipasi masyarakat dalam pembangunan.

Pertanyaan :

Dari berita dan penelitian yang saya baca pak, dikatakan bahwa salah satu tujuan Kementerian Komunikasi dan Informatika Republik Indonesia terus berupaya Ger-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

akan kampanye Internet Sehat Aman yang bertujuan menekan pengaruh konten negatif serta dampak *cyberbullying*. Apa contoh program kerja Diskominfo dalam rangka meliterasi masyarakat ?

Narasumber :

Untuk sosialisasi secara langsung ke masyarakat dalam program kerja terakhir ini tidak ada bentuk program yang langsung ke lapangan. Meliterasi masyarakat juga merupakan salah satu tujuan kami, namun dalam proses penyampaian informasinya biasanya dilakukan melalui *event/pameran* dan media luar ruang berupa konten dengan desain.

Pertanyaan :

Baik pak. Penelitian yang sedang saya lakukan terkait dengan *cyberbullying* di media sosial ait dengan *cyberbullying* di media sosial Youtube yang topiknya tentang *buzzer* politik dan isu keagamaan di Indonesia. Nantinya akan diperoleh pengetahuan tentang perkembangan *cyberbullying* dan jenis-jenis komentar yang memuat unsur *cyberbullying*. Apakah Diskominfo juga dapat menyebarkan informasi berkaitan dengan masalah atau topik diluar program kerja yang telah ada atau ditentukan sebelumnya ?

Narasumber :

Tentu saja bisa. Seperti yang sudah saya sebutkan bahwa salah satu tujuan yang kami adalah meliterasi masyarakat. Untuk sumber informasi yang digunakan, ada yang memang diagendakan atau dalam agenda, dan ada yang tidak teragenda. Contoh misalnya literasi terkait internet sehat dilakukan secara tidak teragenda tapi *by case* atau *by momentum*. beberapa waktu lalu sedang viral kasus *hacker* Bjorka, nah berdasarkan kasus atau momentum tersebut secara tidak teragenda kami mempersiapkan, mengelola, dan mempublikasikan terkait internet sehat dan aman, seperti tips menjaga keamanan akun dan lain-lain. penelitian anda tentang *cyberbullying* dengan politik dan agama dapat menjadi konten bermanfaat dan dapat digunakan *by momentum*, contohnya saja sebentar lagi kita sudah memasuki tahun-tahun pemilu dan pasti media sosial akan berseliweran komentar massa. Penelitian yang anda kerjakan dapat kami gunakan sebagai bahan atau strategi untuk meliterasi masyarakat baik secara umum maupun pada momentum tertentu. Bahan literasi dapat kita terbitkan di website pemerintahan baik, dalam bentuk narasi pemberitaan, pamflet, dan poster setelah biasanya melalui SOP tertentu.

Pertanyaan :

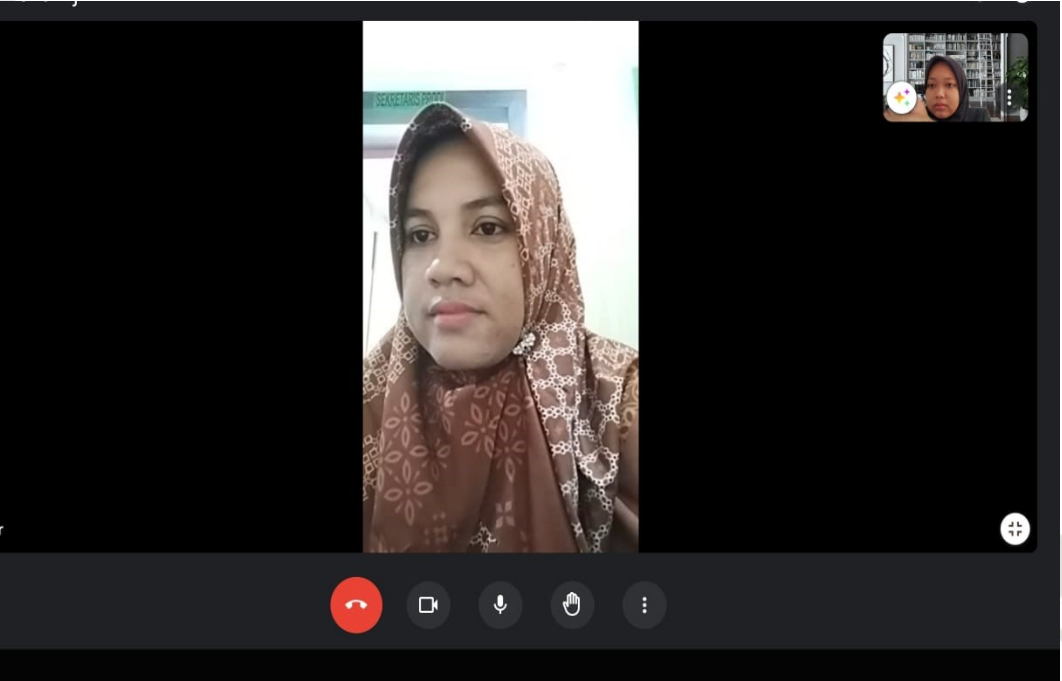
Baik pak. Saya ucapkan terima kasih telah meluangkan waktu untuk wawancara ditengah kesibukan bapak.

(Penutup)

LAMPIRAN I

FOTO DOKUMENTASI WAWANCARA

© Hak cipta milik UIN Suska Riau



State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



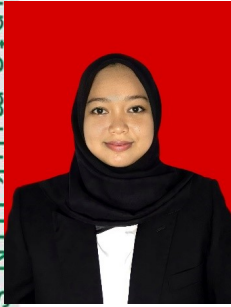
© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR RIWAYAT HIDUP



Penulis bernama Rimelda Adha lahir di Titian Tinggi, Kabupaten Indragiri Hulu, Provinsi Riau pada 13 Maret 2000, merupakan anak pertama dari Bapak Santiagus dan Ibu Painem. Penulis menempuh pendidikan dasar di SDN 022 Titian Tinggi tahun 2006-2012, pendidikan menengah pertama di SMPN 4 Rengat Barat Tahun 2012-2015, dan pendidikan menengah atas di SMAN 1 Rengat Barat tahun 2015-2018. Setelah menamatkan pendidikan dasar, pada tahun 2018 penulis melanjutkan pendidikan jenjang Strata 1 di Universitas Islam Negeri Sultan Syarif Kasim Riau pada Program Studi Sistem Informasi Fakultas Sains dan Teknologi.

Pengalaman organisasi penulis yaitu anggota Himpunan Mahasiswa Sistem Informasi (HIMASI) divisi kesekretariatan tahun 2018-2019. Anggota komunitas riset *Puzzle Research Data Technology* (PREDATECH) divisi Data Center tahun 2019-2020, divisi Publikasi tahun 2020-2021, dan Koordinator Mahasiswa tahun SK 2022. Penelitian Tugas Akhir penulis berjudul Pendekatan *Lexicon Based Features* untuk Identifikasi *Cyberbullying* Menggunakan Algoritma Klasifikasi.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Lampiran Surat :

Nomor : Nomor 25/2021

Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Rimelda Adha

NIM : 11850324476

Tempat/Tgl. Lahir : Titian Tinggi, 13 Maret 2000

Fakultas/Pascasarjana : Sains dan Teknologi

Prodi : Sistem Informasi

Judul Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya*:

Pendekatan Lexicon Based Features untuk Identifikasi Cyberbullying
Menggunakan Algoritma Klasifikasi

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan ~~Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya*~~ dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu ~~Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya*~~ saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan ~~Disertasi/Thesis/Skripsi/(Karya Ilmiah lainnya)*~~ saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 10 Juli 2023
Yang membuat pernyataan



Rimelda Adha
RIMELDA ADHA

NIM : 11850324476

*pilih salah satu sesuai jenis karya tulis