

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.

**KLASIFIKASI JENIS *HATE SPEECH* PADA
TWEET DENGAN MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Teknik

Pada Jurusan Teknik Informatika

Oleh

FANRIAN RAYMONDRA

11351101176



FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM

RIAU

PEKANBARU

2020

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 PERSETUJUAN

**KLASIFIKASI JENIS *HATE SPEECH* PADA
TWEET DENGAN MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

TUGAS AKHIR

Oleh

FANRIAN RAYMONDRA
11351101176

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 28 Juli 2020

Pembimbing I,

Yusra, S.T., M.T.
NIP. 19840123 201503 2 001

Pembimbing II,

Muhammad Fikry, S.T., M.Sc
NIP. 19801018 200710 1 002



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 PENGESAHAN

**KLASIFIKASI JENIS *HATE SPEECH* PADA
TWEET DENGAN MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

TUGAS AKHIR

Oleh

FANRIAN RAYMONDRA
11351101176

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 28 Juli 2020

Pekanbaru, 28 Juli 2020

Mengesahkan,

Ketua Jurusan,

Dr. Elin Haerani, S.T., M.Kom.
NIP. 19810523 200710 2 003

Dekan,

Dr. Drs. Ahmad Darmawi, M.Ag.
NIP. 19660604 199203 1 004

DEWAN PENGUJI

- Ketua : Muhammad Irsyad, S.T., M.T.
 Sekretaris : Yusra, S.T., M.T.
 Pembimbing II : Muhammad Fikry, S.T., M.Sc.
 Penguji I : Dr. Alwis Nazir, M.Kom.
 Penguji II : Muhammad Affandes, M.T.



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 pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

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

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 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, 28 Juli 2020

Yang membuat pernyataan,

Fanrian Raymondra

11351101176

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.

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.

KLASIFIKASI JENIS *HATE SPEECH* PADA *TWEET* DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR

FANRIAN RAYMONDRA

113501176

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Sultan Syarif Kasim Riau

ABSTRAK

Hate speech (ungkapan kebencian) adalah perkataan, perilaku, tulisan, atau pertunjukan yang dilarang karena dapat memicu tindakan kekerasan dan sikap prasangka dari pihak pelaku atau korban dari tindakan tersebut. Penyebaran *Hate speech* ini ditemukan melalui sosial media seperti di *Twitter*. Proses klasifikasi diperlukan untuk menentukan apakah *tweet* tersebut termasuk kesalah satu jenis *Hate speech* atau tidak. Penelitian melakukan empat proses klasifikasi yaitu Suku, Agama, Ras dan Antargolongan. Proses klasifikasi pada penelitian ini diawali dengan *text preprocessing* dan pembobotan kata, kemudian dilanjutkan perhitungan *cosine similarity* untuk pengkategorian klasifikasi jenis *Hate speech*.

Aplikasi menggunakan algoritma *K-Nearest Neighbor* karena paling sering digunakan untuk proses klasifikasi dan dinilai cukup efektif dalam pengelompokan data. Jumlah dataset yang digunakan sebanyak 400 *tweet*. Pada proses pengujian didapatkan nilai *F-measure* 92% dengan *precision* 93% dan *recall* sebesar 92% pada nilai $k=7$, sehingga dapat disimpulkan efektivitas aplikasi berjalan dengan lancar.

Kata Kunci: *Hate speech*, *K-Nearest Neighbor*, Klasifikasi, *Twitter*, *Text Preprocessing*



HATE SPEECH CLASSIFICATION USING K-NEAREST NEIGHBOR METHOD

FANRIAN RAYMONDRA

11351101176

*Information Engineering Department
Faculty of Sciences and Technology
State Islamic University of Sultan Syarif Kasim Riau*

ABSTRACT

Hate speech is speech, behavior, writing, or performance that is prohibited because it can trigger acts of violence and prejudice on the part of the perpetrators or victims of such actions. The spread of hate speech was found through social media like on Twitter. The classification process is needed to determine whether the tweet is included in one of the hate speech types or not. The study conducted four classification processes namely Ethnic, Religious, Race and Intergroup. The classification process in this study begins with text preprocessing and word weighting, then proceed with the calculation of cosine similarity for categorizing the type of speech. The application uses the K-Nearest Neighbor algorithm because it is most often used for the classification process and is considered quite effective in grouping data. The amount of dataset used is 400 tweets. In the testing process, the F-measure value was 92% with a precision of 93% and a recall of 92% at a value of $k = 7$, so it can be concluded that the effectiveness of the application running well.

Keyword : Classification, Hate Speech, K-Nearest Neighbor, Twitter, Text Preprocessing

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

Assalammu'alaikum wa rahmatullahi wa barakatuh

Alhamdulillahirobbil'alamin, bersyukur kepada Allah, karena atas segala limpahan rahmat dan karunia-Nya yang diberikan sehingga penulis dapat menyelesaikan penelitian dan penulisan laporan Tugas Akhir yang berjudul “**Klasifikasi Jenis Hatespeech Pada Tweet Dengan Menggunakan Metode K-Nearest Neighbor**”. Tugas Akhir ini disusun sebagai salah satu syarat kelulusan dari jurusan Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Selama menyelesaikan Tugas Akhir ini, penulis banyak mendapat bantuan, bimbingan, arahan, petunjuk dari banyak pihak baik secara langsung maupun tidak langsung. Untuk itu dalam kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. H. Akhmad Mujahidin, S.Ag, M.Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Ahmad Darmawi, M.Ag selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Ibu Dr. Elin Haerani, M.T selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
Bapak Muhammad Fikry, S.T, M.Sc selaku Sekretaris Jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau sekaligus Pembimbing I yang telah banyak memberikan bimbingan, nasehat, serta kritik dan saran yang membangun dalam penyusunan Tugas Akhir.
5. Ibu Yusra, S.T, M.T selaku pembimbing II yang telah banyak memberi arahan, bimbingan, nasehat serta saran dalam penyusunan Tugas Akhir.
6. Bapak Dr. Alwis Nazir, M.Kom selaku penguji I dan Bapak Muhammad Affandes, M.T selaku penguji II yang telah memberikan banyak masukan agar Tugas Akhir ini dapat diselesaikan dengan baik.
7. Ibu Lestari Handayani, S.T, M.Kom selaku pembimbing Akademis penulis dari tahun 2013 sampai 2018 yang telah membimbing dan memberi support ketika penulis mengalami kesulitan selama dikampus.

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.

Ibu Dr. Okfalisa, S.T, M.Sc selaku pembimbing Akademis yang telah memberikan arahan dan motivasi kepada penulis.

Kepada Bapak dan Ibu Dosen Teknik Informatika yang telah banyak memberikan bimbingan dan ilmunya kepada penulis.

10. Papa dan Mama yang selalu memberikan semangat, dukungan, serta doa untuk kelancaran dalam menyelesaikan Tugas Akhir ini.

11. Fajar Maulana, Wirdatul Hasana, Ramadani, Ratna Sari dan Mursyida Fadhillah yang membantu proses dalam menyelesaikan Tugas Akhir ini.

12. Teman teman Classyfight F yang tidak bisa penulis sebutkan namanya yang telah banyak membantu, mendukung dan menyemangati dalam penulisan Laporan.

13. Semua pihak yang terlibat yang tidak dapat disebutkan satu per satu oleh penulis.

Semoga Tugas Akhir ini dapat bermanfaat bagi penulis khususnya maupun pembaca pada umumnya. Penulis berharap ada kritik dan saran dari pembaca atas Tugas Akhir ini yang dapat disampaikan ke alamat email penulis: fanrian.raymondra@students.uin-suska.ac.id. Akhir kata penulis ucapkan terima kasih dan selamat membaca.

Pekanbaru, Juli 2020

Penulis

UIN SUSKA RIAU

DAFTAR ISI

	Halaman
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iv
LEMBAR PERNYATAAN	v
ABSTRAK	vi
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
DAFTAR SIMBOL	xv
BAB I PENDAHULUAN	I-1
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah.....	I-4
1.3 Batasan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Sistematika Penulisan	I-5
BAB II LANDASAN TEORI	II-1
2.1 <i>Hate speech</i>	II-1
2.2 Text Mining	II-2
2.3 Text Preprocessing.....	II-2
2.4 Fitur dan Pembobotan Kata	II-3
2.5 K-Nearest Neighbor	II-4
2.6 Cosine Similarity	II-5
2.7 Evaluasi.....	II-5
2.8 Penelitian Terkait	II-7
BAB III METODOLOGI PENELITIAN	III-1
3.1 Identifikasi Masalah.....	III-2
3.2 Perumusan Masalah	III-2

- 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.

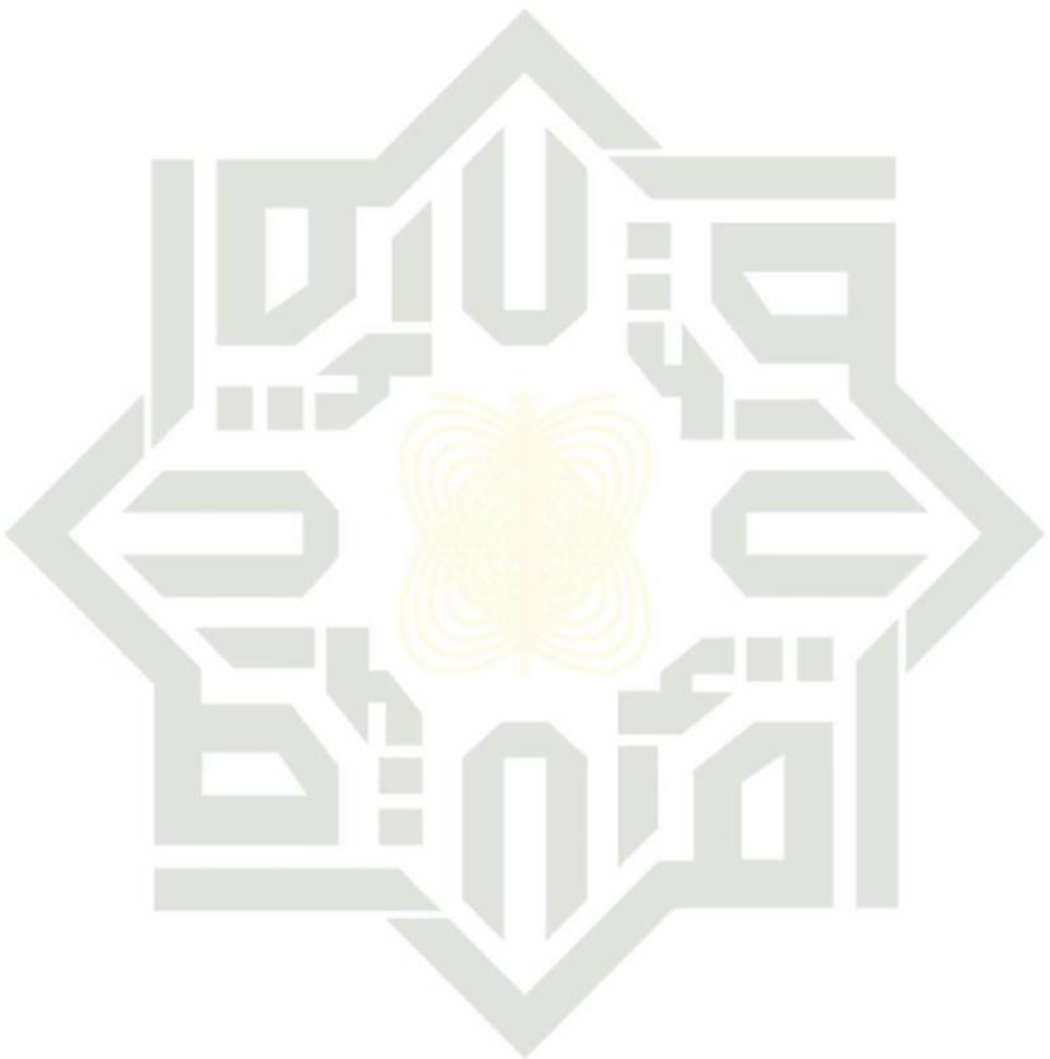
3.3	Pengumpulan Data	III-2
3.3.1	Studi Pustaka.....	III-2
3.3.2	Pengumpulan Data <i>Tweet</i>	III-2
3.4	Tahapan Analisa.....	III-3
3.4.1	Analisa Kebutuhan Data	III-3
3.4.2	Text Pre-processing	III-3
3.4.3	Fitur dan Pembobotan.....	III-3
3.4.4	Analisa K-Nearest Neighbor untuk Klasifikasi	III-4
3.5	Tahapan Perancangan	III-4
3.6	Implementasi.....	III-5
3.7	Pengujian.....	III-5
3.8	Kesimpulan dan Saran	III-6
BAB IV ANALISA DAN PERANCANGAN		IV-1
4.1	Tahapan Analisa.....	IV-1
4.1.1	Analisa Kebutuhan Data	IV-1
4.1.2	Text Preprocessing.....	IV-2
4.1.3	Fitur Dan Pembobotan Kata	IV-7
4.1.4	Analisa K-Nearest Neighbor.....	IV-21
4.2	Tahapan Perancangan	IV-22
4.2.1	Perancangan Database	IV-22
4.2.2	Perancangan <i>Interface</i>	IV-24
BAB V IMPLEMENTASI DAN PENGUJIAN.....		V-1
5.1	Implementasi.....	V-1
5.1.1	Batasan Implementasi	V-1
5.1.2	Lingkungan Implementasi	V-1
5.1.3	Hasil Implementasi Antar Muka.....	V-2
5.2	Pengujian.....	V-5
5.2.1	Analisa Hasil Pengujian.....	V-13
BAB VI PENUTUP		VI-1
6.1	Kesimpulan	VI-1
6.2	Saran	VI-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 PUSTAKA xvi



UIN SUSKA RIAU



DAFTAR GAMBAR

Gambar	Halaman
Gambar 3. 1 Tahapan Metodologi Penelitian	III-1
Gambar 4. 1 Perancangan Halaman Utama	IV-24
Gambar 4. 2 Perancangan Halaman Kelas	IV-25
Gambar 4. 3 Perancangan Halaman Stopword	IV-25
Gambar 4. 4 Perancangan Halaman Normalisasi	IV-26
Gambar 4. 5 Perancangan Halaman Klasifikasi	IV-27
Gambar 4. 6 Perancangan Klasifikasi <i>Tweet</i>	IV-27
Gambar 5. 1 Tampilan Halaman Utama	V-2
Gambar 5. 2 Tampilan Halaman Kelas	V-3
Gambar 5. 3 Tampilan Halaman Stopword	V-3
Gambar 5. 4 Tampilan Halaman Normalisasi	V-4
Gambar 5. 5 Tampilan Halaman Klasifikasi	V-5
Gambar 5. 6 Tampilan Halaman Klasifikasi <i>Tweet</i>	V-5
Gambar 5. 7 Perhitungan Manual pada Sistem	V-6
Gambar 5. 8 Proses Perhitungan Manual	V-6
Gambar 5. 9 Proses Perhitungan Manual	V-7
Gambar 5. 10 Proses Perhitungan Manual	V-7

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.






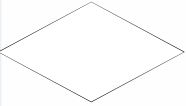
DAFTAR TABEL

Tabel	Halaman
Tabel 2. 1 Tabel <i>Confusion matrix</i>	II-6
Tabel 2. 2 Penelitian Terkait	II-7
Tabel 4. 1 <i>Tweet</i> yang Sudah Diberikan Label	IV-2
Tabel 4. 2 Proses <i>Cleaning Data Tweet</i>	IV-3
Tabel 4. 3 Proses <i>Case Folding Data Tweet</i>	IV-3
Tabel 4. 4 Proses <i>Tokenizing Data Tweet</i>	IV-4
Tabel 4. 5 Proses Normalisasi Data <i>Tweet</i>	IV-5
Tabel 4. 6 Proses <i>Filtering Data Tweet</i>	IV-6
Tabel 4. 7 Proses <i>Stemming Data Tweet</i>	IV-7
Tabel 4. 8 Pembobotan Kata menggunakan TF-IDF	IV-8
Tabel 4. 9 Perhitungan Wdt ($Wdt = tf.idf$).....	IV-11
Tabel 4. 10 Perhitungan Pembobotan Kata.....	IV-14
Tabel 4. 11 Perhitungan Panjang Vektor	IV-18
Tabel 4. 12 <i>Cosine Similarity</i>	IV-22
Tabel 4. 13 Struktur Tabel Document.....	IV-22
Tabel 4. 14 Struktur Tabel Kelas	IV-23
Tabel 4. 15 Struktur Tabel Normalisasi	IV-23
Tabel 4. 16 Struktur Tabel Stemming.....	IV-23
Tabel 4. 17 Struktur Tabel Stopword.....	IV-24
Tabel 5. 1 Pengujian Skenario 1	V-11
Tabel 5. 2 Pengujian Skenario 2	V-11
Tabel 5. 3 Pengujian Skenario 3	V-12
Tabel 5. 4 Rata-rata Precision, Recall, dan F-measure	V-12
Tabel 5. 5 <i>Tweet</i> yang Tidak Sesuai	V-13

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 SIMBOL

<i>Chart</i>	Keterangan
	Terminal : simbol yang digunakan untuk mulai dan akhir/selesai
	Input-output : simbol yang digunakan untuk menyatakan input dan output
	Proses : simbol yang digunakan untuk melakukan pemrosesan oleh user ataupun komputer
	Decision : simbol yang digunakan untuk memutuskan proses selanjutnya yang akan dilakukan.

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.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hate speech (Ujaran Kebencian) menjadi salah satu masalah yang sering ditemukan pada media sosial, menurut Bapak Wiranto selaku Menteri Politik, Hukum, dan Keamanan Republik Indonesia menyatakan bahwa sepanjang tahun 2018 telah terjadi sebanyak 324 kasus *Hate speech* (Ujaran Kebencian) (Permanasari, 2018). Dalam arti hukum, *Hate speech* adalah perkataan, perilaku, tulisan, atau pertunjukan yang dilarang karena dapat memicu terjadinya tindakan kekerasan dan sikap prasangka entah dari pihak pelaku pernyataan tersebut atau korban dari tindakan tersebut yang berujung pada Suku, Agama, Ras dan Antargolongan (SARA) (Ismunarno, 2016).

SARA adalah berbagai pandangan dan tindakan yang didasarkan pada sentimen identitas yang menyangkut keturunan, agama, kebangsaan, atau kesukuan dan golongan. Dalam pengertian lain SARA dapat disebut diskriminasi yang merujuk kepada pelayanan yang tidak adil terhadap individu tertentu. Diskriminasi merupakan suatu kejadian yang biasa dijumpai dalam masyarakat, hal ini dikarenakan kebiasaan masyarakat untuk membeda-bedakan yang lain. Ketika seseorang diperlakukan secara tidak adil karena karakteristik SARA atau karakteristik lain yang diduga merupakan dasar dari tindakan diskriminasi (Ismunarno, 2016).

Di Indonesia khususnya beberapa kasus *Hate speech* yang cukup menyita perhatian masyarakat Indonesia adalah kasus Florence yang menghina kota Yogyakarta, Basuki Tjahya Purnama atas penistaan agama, ujaran kebencian yang dilakukan Jonru Ginting di akun sosial medianya, unggahan Jonru di media sosial dinilai sangat berbahaya dan jika dibiarkan dapat memecah belah bangsa, kemudian kasus ujaran kebencian yang dilakukan oleh Ahmad Dani dimana dalam unggahan akun *Twitter* nya dianggap menghasut dan penuh kebencian terhadap kelompok pendukung Ahok, lalu kasus pembakaran mesjid di Tolikara Papua yang menyebabkan konflik agama di Papua sehingga berujung dengan

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.

ujaran kebencian antar agama di media sosial. Dari kasus-kasus diatas pelaku penyebar *Hate speech* ini banyak memanfaatkan media online untuk menebar kebencian (Movanita, 2017).

Menurut Gagliardone (dikutip dalam Juditha, 2017) sebuah studi yang dilakukan oleh Unesco pada tahun 2015 yang berjudul “*Countering Online Hate Speech*” menyebutkan bahwa fenomena *Hate speech* secara *online* semakin berkembang dan menimbulkan beragam masalah baik di dalam maupun di luar Eropa. *Hate speech* secara *online* merupakan salah satu tren dari tahun sebelumnya. Laporan ini juga menekankan bahwa *Hate speech* melalui media *online* sudah semakin pesat dan memiliki potensi untuk mencapai audiens yang lebih besar. Juga dalam buku tersebut menjelaskan *Hate speech online* cukup sulit untuk pemerintah dalam menerapkan undang-undang nasional di dunia maya, hal itu dikarenakan kecepatan dan jangkauan internet yang sangat luas. Dan masalah seputar *Hate speech* juga menimbulkan adanya semacam ruang pribadi untuk berekspresi yang bertujuan untuk berinteraksi di media sosial seperti *Twitter*.

Saat ini media online yang populer di Indonesia adalah *Twitter*. Terlihat sejak di *release Twitter* versi Lite, hal itu diutarakan oleh Dwi Adriansyah sebagai *Country Industry Head Twitter Indonesia and Malaysia*, pada saat menghadiri acara bertajuk *#RameDiTwitter* pada tahun 2018. *Twitter* adalah media sosial yang penggunanya bisa membaca dan mengirim text maximal 280 karakter. *Twitter* juga sebagai wadah tempat pembicaraan topik yang sedang hangat dan sebagai tempat untuk saling bertukar informasi.

Tidak semua informasi-informasi yang terdapat di *Twitter* itu *Good speech*, jika lebih di cermati terdapat beberapa informasi yang mengandung *Hate speech*. Oleh karena itu diperlukan klasifikasi *Hate speech* dan jenis *Hate speech* yang terdapat pada *tweet*. Klasifikasi *hate speech* ini diperlukan agar bisa memahami suatu *tweet* apakah di dalam *tweet* tersebut mengandung unsur hatespeech atau tidak. Selain itu tujuan dalam pengklasifikasian jenis hatespeech ini adalah untuk melihat jenis ujaran kebencian apa yang sering terjadi dari beberapa kasus ujaran kebencian, terkhususnya di indonesia, dengan adanya aplikasi untuk

pengklasifikasian *Hate speech* ini diharapkan bisa mempermudah lembaga-lembaga yang menangani kasus-kasus ujaran kebencian.

Penelitian terkait tentang *Hate speech* yaitu penelitian Ghulam Asrofi Guntoro (2016) yang berjudul “*Analisis Sentimen pada Twitter dengan Metode Naïve Bayes Classifier dan Support Vector Machine*”. Pada penelitian ini data yang diambil hanya berdasarkan *Tweet* yang menggunakan tagar *Hate speech* (#Hatespeech) dari penelitiannya di dapat nilai akurasi sebesar 66,6%. Penelitian yang dilakukan oleh Nakamura (2017) yang berjudul “*Detecting Hate, Offensive, and Regular Speech in short Comments*”. Pada penelitian ini Nakamura menggunakan kombinasi TF-IDF dan unigram untuk klasifikasi teks, dan mendapatkan hasil *F1-score* 86%, 84% dan 96% for classifying *hate, offensive, and regular speech*.

Selanjutnya penelitian yang dilakukan oleh Ulfadli Rahman (2015) yang berjudul “*Klasifikasi Iklan Pada Twitter Menggunakan K-Nearest Neighbor*”. Hasil pengujian nilai akurasi terbaik pada komposisi tiga dengan hasil 94,1% dan nilai k terbaik pada $k=15$. Penelitian oleh Sandi Eko Yulianto (2016) yang berjudul “*Klasifikasi Topik Berita Menggunakan Metode K-Nearest Neighbor Classifier*”. Hasil pengujian tingkat akurasi sebesar 86.7% dari pengklasifikasian tiga topik berita. Selanjutnya penelitian yang dilakukan oleh Zuhdiyah Ulfah Siregar (2019) yang berjudul “*Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor*”. Pada penelitian tersebut data komentar dibagi menjadi empat kategori yaitu instruktur, materi, sarana dan prasarana, dan hasil dari pengujian di dapatkan akurasi sebesar 94,23%.

K-Nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, dimana hasil dari sampel uji yang baru di klasifikasikan berdasarkan mayoritas dari kategori pada *K-Nearest Neighbor* (Krisandi dkk, 2013). Selain itu metode *K-Nearest Neighbor (KNN)* adalah metode yang paling sering digunakan untuk proses klasifikasi, karena dinilai cukup efektif dalam pengelompokan data (Tan & Steinbach, 2006).

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.

Tujuan penelitian ini ialah karena masih banyak netizen yang salah mengartikan maksud dari beberapa *tweet* yang mengandung unsur kalimat *Hate speech*, contohnya saja yang penulis kutip dari artikel, bahwa *Hate speech* selalu dikaitkan dengan *Freespeech* padahal nyatanya tidak seperti itu, keduanya memang terkait kebebasan berekspresi, namun jelas bahwa kebebasan berekspresi itu terkait hal yang umum dalam kepentingan sosial serta batasan-batasan alami. Ujaran kebencian (*Hate speech*) walau merupakan pernyataan yang intoleran, namun perlu diketahui bahwa tidak semua pernyataan intoleran itu mengandung *Hate speech*, selama pernyataan tersebut tidak mengandung advokasi kebencian yang membentuk hasutan untuk melakukan diskriminasi, permusuhan, atau kekerasan. Berdasarkan penjelasan di atas, penulis tertarik untuk melakukan penelitian yaitu klasifikasi *Hate speech* pada *Tweet* dengan menggunakan metode *K-Nearest Neighbor*. Output dari penelitian ini adalah mengelompokkan *tweet* yang bersifat *Hate speech* berdasarkan jenisnya.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang terdapat pada latar belakang, maka didapatkan rumusan masalah yaitu: “Bagaimana menerapkan metode *K-Nearest Neighbor* pada Klasifikasi jenis *Hate speech* pada *tweet*”.

1.3 Batasan Masalah

Agar permasalahan yang dibahas lebih terfokus pada topik yang diangkat, maka penelitian ini memiliki batasan masalah diantaranya:

1. Objek pada penelitian ini adalah *tweet* berbahasa Indonesia yang diambil dari *Twitter*.
2. Jenis *Hate speech* yaitu *Hate speech* Suku, *Hate speech* Agama, *Hate speech* Ras dan *Hate speech* Antargolongan.
3. Menggunakan 400 data *tweet*, yang dibagi menjadi 100 *tweet* perkelasnya.
4. Data *tweet* yang digunakan adalah kata murni, bukan *tweet* foto dan video.

1.4 Tujuan Penelitian

Tujuan dari penelitian tugas akhir ini adalah:

1. Mengklasifikasi jenis kalimat *Hate speech* dengan *K-Nearest Neighbor*.



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.

2. Mengetahui hasil akurasi dari metode *K-Nearest Neighbor*.

1.5 Sistematika Penulisan

Laporan tugas akhir ini terdiri dari enam bab, dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang dari permasalahan pada topik penelitian hingga penelitian dapat dirumuskan dan didapatkan tujuan dari penelitian. Batasan masalah diberikan agar lebih terfokus pada apa yang diteliti dan tidak keluar pada topik penelitian. Terakhir adalah sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas teori-teori literatur yang mendukung untuk pengerjaan penelitian tugas akhir. Pada bab ini menjelaskan *Hate speech*, kasus-kasus *Hate speech* di Indonesia, *Twitter*, *K-Nearest Neighbor* dan penelitian-penelitian yang terkait.

BAB III METODOLOGI PENELITIAN

Bab ini berisi rangkaian tahapan penelitian, identifikasi masalah, pengumpulan data, analisa kebutuhan data, implementasi serta waktu penelitian.

BAB IV ANALISA DAN PERANCANGAN

Bab ini berisi analisa dan perancangan dalam proses pengklasifikasian *tweet* dengan menggunakan *K-Nearest Neighbor*.

BAB V IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi penjelasan penerapan dari algoritma yang digunakan dalam klasifikasi.

BAB VI PENUTUP

Bab ini berisi tentang kesimpulan dan saran yang berkaitan dengan penelitian ini.

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 II

LANDASAN TEORI

2.1 *Hate speech*

Hate speech atau dalam bahasa Indonesia Ungkapan kebencian adalah tindakan komunikasi yang dilakukan oleh suatu individu atau kelompok dalam bentuk provokasi, hasutan ataupun hinaan kepada individu atau kelompok yang lain dalam hal berbagai jenis seperti ras, warna kulit, etnis, gender, cacat fisik, orientasi seksual, kewarganeraan, agama dan lain-lain. Dalam arti hukum, *hate speech* adalah perkataan, perilaku, tulisan ataupun pertunjukan yang dilarang karena dapat menimbulkan terjadinya kekerasan dan sikap prasangka entah dari pihak pelaku pernyataan tersebut atau dari pihak korban tindakan tersebut (Ismunarno, 2016).

Dalam Surat Edaran Kapolri No: SE/06/X/2015 tentang ujaran kebencian (*Hate speech*) dijelaskan dapat berupa tindak pidana yang diatur dalam KHUP (Kitab Undang-undang Hukum Pidana) dan ketentuan pidana lainnya di luar KHUP yang berbentuk antara lain :

1. Penghinaan
2. Pencemaran nama baik
3. Penistaan.
4. Perbuatan yang tidak menyenangkan.
5. Memprovokasi.
6. Menghasut.
7. Menyebarkan berita bohong.

Semua tindakan diatas memiliki tujuan atau bisa berdampak pada tindak diskriminasi, kekerasan, pengilangan nyawa dan konflik sosial (Mangantibe, 2016).

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.

2.2 Text Mining

Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. Penambangan teks memiliki tujuan dan menggunakan proses yang sama dengan penambangan data, namun memiliki masukan yang berbeda. Masukan untuk penambangan teks adalah data yang tidak terstruktur, seperti dokumen Word, PDF, Kutipan teka dan lain-lain, sedangkan penambangan data adalah data yang terstruktur (Feldman, 2007).

Menurut Hotho (dikutip dalam Yulianto, 2016) Penelitian dibidang *text mining* digunakan untuk menangani masalah dibidang representasi teks, klasifikasi, *clustering*, ekstraksi informasi atau pencarian dan pemodelan pola. Dalam penelitian *text mining* diperlukan tahapan *text preprocessing* pada koleksi dokumen dan menyimpan informasi tersebut dalam struktur data. Pendekatan *text mining* didasarkan pada pemikiran bahwa dokumen teks dapat diwakili oleh satu set kata-kata, yaitu dokumen teks digambarkan berdasarkan pada set kata-kata yang terkandung didalamnya.

2.3 Text Preprocessing

Menurut (Winarko, 2014) *text preprocessing* merupakan langkah penting dalam klasifikasi. Tujuan dilakukannya *text preprocessing* yaitu untuk menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi volume kata.

Menurut (Mooney, 2006) Tahapan pada *text processing* terdiri dari beberapa tahapan yaitu :

1. Case Folding

Case Folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter (pembatas).

2. Tokenizing

Tokenizing atau *parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.

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. Filtering

Filtering adalah tahap mengambil kata-kata penting dari hasil langkah sebelumnya. Dalam prosesnya bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-word*. Contoh *stopword* adalah “yang”, “dan”, “di”, “dari”.

4. Stemming

Tahap *stemming* adalah tahap mencari root kata dari tiap kata hasil *filtering*. Pada tahap ini dilakukan proses pengambilan berbagai bentuk kata kedalam suatu representasi yang sama. Algoritma yang digunakan pada tahapan ini adalah algoritma *Enhanced Confix Stripping* (ECS).

Enhanced Confix Stripping adalah metode stemming pada bahasa Indonesia yang diperkenalkan oleh Jelita Asian yang merupakan pengembangan dari metode stemming yang dibuat oleh Nazief dan Andriani (1996) yang dikutip oleh (Herdiawan, 2016). Kata-kata yang muncul didalam sebuah dokumen sering kali terdapat imbuhan, oleh karena itu setiap kata yang kata yang mengandung imbuhan setelah melalui proses *filtering* dibentuk kedalam kata dasar untuk menghilangkan imbuhan pada kata tersebut.

2.4 Fitur dan Pembobotan Kata

Fitur dan pembobotan kata berfungsi untuk memilih kata (*term*) yang digunakan dalam proses melatih data set. Peningkatan ukuran *term* pada proses data latih dapat dikurangi dengan menentukan fitur yang akan digunakan, selain itu proses ini juga berfungsi untuk meningkatkan tingkat akurasi dengan menghapus fitur yang tidak relevan. Sebuah dokumen dapat diwakili dengan sebuah vektor dari *keyword* yang diektrak, dari vektor tersebut dilakukan pembobotan yang mewakili *keyword* dokumen dan dalam keseluruhan dokumen. Perhitungan fitur dan pembobotan yang biasa digunakan adalah:

1. Document Frequency (DF)

Document frequency disebut juga sebagai *word frequency*. Konsep *document frequency* secara umum adalah menghitung jumlah dokumen



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.

dimana t_m muncul dan disajikan dalam matrik dimana baris menunjukkan dokumen dan kolom menunjukkan *term* (kata). *Document frequency* adalah metode seleksi fitur yang paling sederhana dengan komputasi yang rendah (Rahman, 2015). Sebagai contohnya, diketahui terdapat data berjumlah 100 dokumen, dan jumlah dokumen yang mengandung *term* atau kata “skripsi” adalah 45 dokumen, maka nilai *document frequency* untuk kata skripsi adalah 45.

2. Inversi Document Frequency (IDF)

Pada *information retrieval*, umumnya menggunakan *inversi document frequency*. *Inversi document frequency* fokus pada kemuncula *term* (kata) pada seluruh dokumen. Pada *Inversi document frequency*, *term* yang jarang muncul pada keseluruhan dinilai lebih berharga (Rahman, 2015).

$$IDF = \log \frac{d}{df} \tag{2.1}$$

Keterangan : D = jumlah dokumen dalam data set.

DF = jumlah dokumen yang mengandung data.

2.5 K-Nearest Neighbor

Menurut Gorunescu (dikutip dalam Rohman, 2012) *K-Nearest Neighbor* merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. *K-Nearest Neighbor* termasuk algoritma *supervised learning*, dimana hasil dari query yang baru, diklarifikasikan berdasarkan mayoritas dari kategori pada *K-Nearest Neighbor*. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi. Berikut langkah urutan dalam menggunakan algoritma *K-Nearest Neighbor* (Kurniawan, 2012):

1. Hitung jarak antara data uji dengan data latih yang telah dibangun.
2. Menentukan parameter nilai k = jumlah tetangga terdekat.
3. Mengurutkan jarak terkecil dari data sample.
4. Pasangkan kategori sesuai dengan kesesuaian.



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.

5. Jumlah terbanyak dari tetangga terdekat dengan persamaan kemudian tetapkan kategori.

Urak yang digunakan pada penelitian ini adalah *Cosine Similarity*

2.6 Cosine Similarity

Metode *Cosine Similarity* adalah metode untuk menghitung kesamaan dari dua dokumen. Untuk menyamakan jangka frekuensi setiap kata pada kalimat yang digunakan pada pembobotan, proses pembobotan mengekstrak dokumen menjadi proses yang terdiri dari kumpulan kata perkalimat. Tujuannya adalah untuk menyamakan kedua kalimat pada suatu dokumen yang nantinya akan dibandingkan, sehingga kita dapat melangkah ketahap selanjutnya, yaitu tahapan *similarity* (Putri, 2013).

Penentuan kesesuaian dokumen dengan query atau biasa disebut juga dengan pengukuran (*similarity measure*) antara dokumen vector (D) dengan vector query (Q). Semakin sama suatu vector dokumen dengan vector query maka dokumen bisa dikatakan sesuai dengan query. Rumus yang digunakan untuk menghitung *cosine similarity* adalah sebagai berikut:

$$\cos(i, k) = \frac{\sum_k(d_i d_k)}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \tag{2.2}$$

Keterangan:

$\sum_k(d_i d_k)$: *vector dot* produk dari i, dan k.

$\sqrt{\sum_k d_{ik}^2}$: panjang *vector* i.

$\sqrt{\sum_k d_{jk}^2}$: panjang *vector* k.

2.7 Evaluasi

Evaluasi adalah suatu kegiatan untuk membandingkan hasil implementasi yang telah dilakukan dengan kriteria atau kategori yang telah ditetapkan sebelumnya. Kegiatan ini berfungsi untuk melihat keberhasilan. Hasil evaluasi ini



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.

akan melihat informasi mengenai sejauh mana kegiatan tersebut tercapai, sehingga bisa diketahui bila terdapat selisih dengan kriteria atau kategori yang telah ditetapkan. Tingkat akurasi metode yang dilakukan dapat diuji tingkat akurasinya dengan menggunakan *confusion matrix*.

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat empat istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN).

Sistem temu kembali informasi mengembalikan kumpulan data dokumen yang merupakan jawaban dari *query* pengguna. Ada dua kategori dokumen yang dihasilkan dari sistem temu kembali yaitu *relevant document* (dokumen yang relevan dengan *query*) dan *retrieved document* (dokumen yang diterima pengguna). Biasanya ukuran yang digunakan adalah kombinasi dari *precision* dan *recall*. *Precision* adalah jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif, *precision* dapat diperoleh dengan melakukan persamaan (2.3). Nilai *Recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem, nilai *recall* dapat diperoleh dengan persamaan (2.4).

Tabel 2. 1 Tabel *Confusion matrix*

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (True Positif)	FN (False Negative)
Negatif	FP (False Positive)	TN (True Negative)

$$Precision = \frac{TP}{FP+TP} * 100\% \tag{2.3}$$



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.

Nama Peneliti	Tahun	Judul Penelitian	Kesimpulan
Ghulam Asrofi Guntoro	2016	Analisis Sentimen Hatespeech pada <i>Twitter</i> dengan Metode Naïve Bayes Classifier dan Support Vector Machine	Analisis sentiment <i>tweet</i> bahasa Indonesia dengan tagar hatespeech dapat membantu menentukan sentiment <i>tweet</i> opini yang ada di <i>Twitter</i> , setelah di lakukan analisis banyak <i>tweet</i> yang tidak mengarah ke hatespeech namun tetap menggunakan tagar hatespeech. Nilai akurasi dari metode SVM didapat 66,6%, presisi 67,1%, nilai recall 66,7%, TP rate 66,7% dan TN rate 75,8%. Dalam penelitian ini juga di dapat tingkat akurasi metode SVM lebih tinggi dibandingkan dengan metode Naïve Bayes.
Pratama Dwi Nagraha, Said Al Faraby, Adiwijaya	2108	Klasifikasi Dokumen Menggunakan Metode k-Nearest Neighbor dengan <i>Information Gain</i>	Berdasarkan hasil pengujian metode KNN tanpa <i>Information Gain</i> memiliki nilai rata-rata akurasi yaitu 93,94438% pada seluruh dokumen data training dengan parameter-parameter. Dengan menggunakan kombinasi KNN dan <i>Information Gain</i> di dapat nilai akurasi yaitu 93,4999% pada seluruh data dokumen training. Pada dokumen testing di dapat akurasi yaitu 92% tanpa <i>information Gain</i> dan 90,5%



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.

Nama Peneliti	Tahun	Judul Penelitian	Kesimpulan
			dengan menggunakan <i>information Gain</i> .
Sandi Eko Yulianto	2016	Klasifikasi Topik Berita Menggunakan Metode K-Nearest Neighbor	Pada penelitian ini menggunakan data sebanyak 150 data, sistem klasifikasi topic berita menggunakan metode <i>K-nearest Neighbor</i> berhasil mengklasifikasikan 3 jenis berita dengan tingkat akurasi sebesar 86,7%, hasil akurasi terbaik dengan nilai parameter $k = 1$ dan presentase pembagian data latih dan data uji 60:40.
Thais G. Almaeida, Fabiola G. Nakamura, Bruno A. Souza, Eduardo F. Nakamura	2017	Detecting Hate, Offensive, and Regular Speech in Short Comment	Penelitian ini menggunakan TF-IDF dan unigram untuk proses klasifikasi teks. Hasil pengujian ini didapatkan skor dari F1 86% untuk <i>Hate speech</i> , 84% untuk <i>offensive</i> dan 96% untuk <i>regular speech</i> .



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.

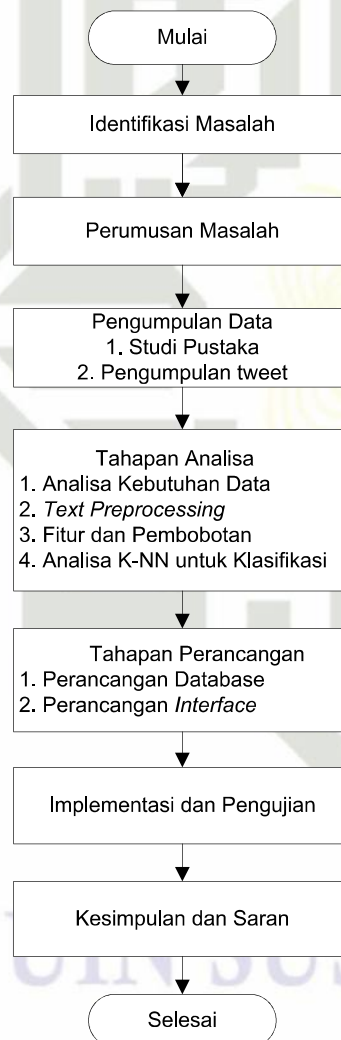
Nama Peneliti	Tahun	Judul Penelitian	Kesimpulan
Fadli Rahman	2015	Klasifikasi Iklan Pada <i>Twitter</i> Menggunakan K-Nearest Neighbor	Penelitian ini menggunakan 1000 <i>tweet</i> , aplikasi ini mengklasifikasi iklan kedalam tiga kelas yaitu jasa, gadget dan pakaian, pada pengujian data latih dan data uji di bagi menjadi tiga komposisi, dari hasil pengujian di dapatkan nilai akurasi terbaik pada komposisi tiga yaitu 94,1% pada nilai k=15.
Zuhdiyah Ulfah Siregar, Riki Ruli A. Siregar, Rakhmat Arianto	2109	Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor	Penelitian ini menggunakan data komentar dari peserta yang mengikuti diklat terhadap udiklat jakarta , data komentar diklasifikasikan kedalam empat kategori yaitu, instruktur, materi, sarana dan prasarana, hasil penelitian mendapatkan akurasi sebesar 94,23%.

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 III**METODOLOGI PENELITIAN**

Penelitian agar bisa terstruktur dalam pengerjaannya dibutuhkan metodologi penelitian. Metodologi penelitian berisikan rencana kerja yang terstruktur agar hasil yang diperoleh sesuai dengan yang diharapkan. Berikut ini rancangan metodologi yang digunakan dalam penelitian tugas akhir dengan judul "Klasifikasi Jenis *Hate speech* Pada *Tweet* Dengan Menggunakan Metode *K-Nearest Neighbor*."



Gambar 3. 1 Tahapan 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.1 Identifikasi Masalah

Tahap ini merupakan tahap awal dari metodologi penelitian. Pada langkah awal ini yang dilakukan adalah mengidentifikasi masalah pada *Hate speech* (ungkapan kebencian) di media *online* seperti *Twitter* dan penelitian-penelitian terkait.

3.2 Perumusan Masalah

Tahap selanjutnya adalah perumusan masalah, setelah mengidentifikasi masalah pada tahap awal, maka didapatkan suatu ide topik dari permasalahan yang diidentifikasi yakni mengklasifikasikan jenis *Hate speech* (Ungkapan kebencian).

3.3 Pengumpulan Data

Setelah melewati tahap awal yakni identifikasi masalah, tahap selanjutnya adalah pengumpulan data. Tahap pengumpulan data berfungsi untuk mengumpulkan informasi dari *tweet-tweet* yang ada di *Twitter*, selain data dari *Twitter* penulis juga mengumpulkan data dari studi pustaka. Setelah data terkumpul selanjutnya akan dilakukan pelabelan data secara manual terlebih dahulu oleh ahli dibidangnya.

3.3.1 Studi Pustaka

Studi pustaka dilakukan dengan cara membaca literatur atau buku-buku yang menjadi acuan dalam mengerjakan penelitian yang dilakukan. Literatur yang digunakan adalah jurnal-jurnal penelitian yang terkait dengan penelitian yang telah dilakukan oleh para peneliti sebelumnya. Selain dari jurnal-jurnal penelitian, literatur juga didapatkan dari buku-buku yang membahas teori yang terkait.

3.3.2 Pengumpulan Data *Tweet*

Pengambilan data pada penelitian ini dilakukan secara manual. *Keyword* yang digunakan dalam penelitian ini meliputi Suku, Agama, Ras, Antarkelompok, Kadal Gurun, Islam Munafik, Batak Bau, FPI, Papua, Kulit Hitam, Cina sipit, Bau Bawang, Minang, Jawa Miskin, Suku Binatang, Kafir. Setelah data *tweet* terkumpul masing-masing data per-*keyword* akan diambil sebanyak 100 *tweet*.



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.

3.4 Tahapan Analisa

Tahapan analisa merupakan gambaran dari penelitian sudah sejauh mana pengerjaan penelitian yang dilakukan. Analisa data yang dilakukan berdasarkan data yang sudah didapat dari studi literatur dan pengumpulan data, kemudian diproses dengan menggunakan metode *K-Nearest Neighbor*. Dalam tahapan analisa terdapat beberapa tahapan didalamnya berikut dijelaskan di bawah ini:

3.4.1 Analisa Kebutuhan Data

Analisa kebutuhan data adalah proses untuk mendapatkan data yang digunakan pada penelitian. Data didapatkan melalui proses unduh dan studi pustaka.

3.4.2 Text Pre-processing

Tahapan *text Pre-processing* adalah tahapan menganalisa masukan bagi suatu sistem klasifikasi yang didapat dari kumpulan-kumpulan tahapan yang ada di dalam *text Pre-processing*, yang meliputi *cleaning*, *case folding*, *tokenizing*, *normalisasi* *filtering* dan *stemming*. *Cleaning* adalah proses pembersihan *tweet* dari kata-kata yang tidak dibutuhkan seperti URL, dan *emoji*. *Case folding* ialah proses untuk menyamaratakan semua huruf menjadi huruf kecil. Proses *Tokenizing* adalah proses untuk memisahkan kalimat menjadi kata per-kata. *Normalisasi* adalah tahapan untuk mengkoreksi ejaan kata pada *tweet* yang tidak sesuai dengan kata baku menjadi kata baku. Selanjutnya proses *Filtering* adalah tahap mengambil kata-kata penting dari hasil langkah sebelumnya. Dalam prosesnya bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). Kemudian tahap *stemming* adalah tahap mencari root kata dari tiap kata hasil *filtering*. Pada tahap ini dilakukan proses pengambilan berbagai bentukan kata kedalam suatu representasi yang sama.

3.4.3 Fitur dan Pembobotan

Pada penelitian ini menggunakan TF-IDF untuk proses pembobotan kata. Untuk mendapatkan nilai TF-IDF proses yang dilakukan terlebih dahulu adalah mencari nilai TF, DF dan IDF. Berikut di bawah ini tahapan proses fitur dan pembobotan kata:

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.

1. Hitung muncul kata dari *tweet* yang ada.
2. Setelah proses pertama dilakukan, hitung nilai IDF dengan rumus $log=jumlah\ tweet/jumlah\ perulangan\ kata$.
3. Selanjutnya lakukan proses perhitungan TF.IDF dengan memasukkan nilai IDF kedalam kolom yang bernilai = 1 pada *tweet*.
4. Hitung bobot dari setiap kata pada *tweet*.
5. Setelah hasil didapatkan selanjutnya dilakukan perhitungan panjang vektor untuk mencari nilai *Cosine similiarity*.

3.4.4 Analisa K-Nearest Neighbor untuk Klasifikasi

Pada tahapan analisa ini dilakukan dengan *K-Nearest Neighbor* untuk melakukan proses Klasifikasi. Hasil tahapan yang telah dilakukan sebelumnya akan menjadi acuan dalam penelitian ini untuk mencari nilai terdekat dari nilai ketetanggaanya. Proses dapat dilakukan dengan sebagai berikut:

1. Melakukan tahapan *preprocessing* sehingga didapatkan representasi dari data uji dan data latih.
2. Hitung nilai *cosine similiarity* dan data uji dengan semua data latih.
3. Selanjutnya mengurutkan hasil dari perhitungan dari menggunakan *cosine similiarity* secara menurun. Nilai yang lebih tinggi berarti terdapat kemiripan dari data uji dan data latih.
4. Mengelompokkan hasil perhitungan dari nilai *cosine similiarity* berdasarkan kategorinya.
5. Dari hasil yang telah didapatkan, kemudian ditentukan nilai *k* pada algoritma K-NN dan diketahui *query* yang memiliki nilai terdekat dengan data uji.

3.5 Tahapan Perancangan

Tahapan perancangan adalah tahapan yang dilakukan untuk merencanakan bagaimana struktur aplikasi yang akan dibuat berdasarkan analisa-analisa yang telah dilakukan sebelumnya. Dalam tahapan perancangan ini terdapat proses yang dilakukan:



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.

1. Perancangan Database

Perancangan database ialah merancang penyimpanan data untuk kebutuhan aplikasi, tahapan ini sangat penting karena apabila database yang dirancang tidak benar maka aplikasi tidak dapat berjalan dengan benar. Proses ini penting agar dalam proses pengerjaan aplikasi dapat berjalan dengan lancar.

2. Perancangan Interface (Antarmuka)

Perancangan Interface ialah proses mendesain dari tampilan aplikasi yang dirancang.

3.6 Implementasi

Tahap implementasi adalah tahap pembuatan aplikasi dari hasil analisa dan perancangan yang telah dilakukan pada tahapan sebelumnya, sehingga aplikasi bisa digunakan secara nyata. Pada tahapan implementasi ini dibutuhkan perangkat pendukung, yaitu:

1. Perangkat Keras (Hardware)

- Processor : Intel Celeron (Minimum)
- Memory (RAM) : 4 GB
- Hardisk : 300 GB

2. Perangkat Lunak (Software)

- Sistem Operasi : Windows 7 Ultimate
- Web Server : Apache
- Browser : Mozilla Firefox, Google Chrome
- Tools : Notepad++
- DBMS : MySQL
- Bahasa Pemrograman : PHP

3.7 Pengujian

Setelah tahapan implementasi selesai, tahapan selanjutnya adalah tahapan pengujian. Tahapan pengujian dilakukan terhadap perancangan yang telah dirancang sebelumnya. Dalam tahapan ini dilakukan testing terhadap aplikasi hal ini berfungsi untuk melihat apakah aplikasi yang rancang sudah sesuai dengan



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.

yang diinginkan. Pada tahapan ini pengujian dilakukan dengan cara sebagai berikut:

1. White Box

Pengujian *white box* adalah pengujian yang didasarkan pada pengecekan terhadap detil perancangan, menggunakan struktur kontrol dari desain program secara prosedural untuk membagi pengujian ke dalam beberapa kasus pengujian.

2. Confusion Matrix

Pada pengujian pada tingkat akurasi metode yang digunakan bisa digunakan *confusion matrix*. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh aplikasi dengan hasil klasifikasi yang seharusnya. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat empat istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Berdasarkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN) didapatkan nilai akurasi, *precision* dan *recall*.

3.8 Kesimpulan dan Saran

Kesimpulan merupakan tahap menyajikan hasil pengujian yang telah dilakukan, tahapan ini berfungsi untuk mengetahui apakah penelitian yang dilakukan telah sesuai dengan kriteria yang telah ditetapkan dan apakah penelitian ini sesuai dengan kebutuhan dan dapat dioperasikan dengan baik. Pada bagian saran berisi kemungkinan pengembangan yang dapat dilakukan dari penelitian tersebut.



BAB VI PENUTUP

6.1 Kesimpulan

Dari hasil analisa dan implementasi data *tweet* pada Klasifikasi Jenis Hatespeech Pada *Tweet* Dengan Menggunakan K-Nearest Neighbor, maka kesimpulan yang dapat diambil adalah :

1. Aplikasi yang dibuat dengan menggunakan metode K-Nearest Neighbor dapat mengklasifikasikan jenis *Hate speech* kedalam empat kelas yaitu Suku, Agama, Ras dan Antarkelompok.
2. Akurasi tertinggi pada saat pengujian dengan menggunakan metode K-Nearest Neighbor terdapat pada skenario tiga yaitu *precision* 93%, *recall* 92% dan *F-measure* 92% pada nilai $k=7$. Hasil rata-rata pengujian dari ketiga skenario didapatkan *precision* (akurasi) sebesar 0,846, *recall* (sensivitas atau kemampuan sistem dalam memilih hasil yang sesuai) sebesar 0,845 dan nilai *F-measure* nya sebesar 0,837.

6.2 Saran

Berdasarkan penelitian yang telah dilakukan, penulis memberi saran untuk pengembangan penelitian lebih lanjut.

Mengembangkan aplikasi klasifikasi yang dapat memproses *tweet* yang berisi gambar dan video, karena trend pada saat ini terdapat beberapa *tweet* yang berupa gambar meme dan video yang mengandung *Hate speech*.

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 PUSTAKA

- Feldman, R. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*.
- Herdiawan. (2016). Analisis Sentimen Terhadap Telkom Indihome Berdasarkan Opini Publik Menggunakan Metode Improved K-Nearest Neighbor Teknik Informatika – Universitas Komputer Indonesia.
- Imunarno, M. A. G. dan. (2016). Pertanggungjawaban Pidana Terhadap Penyebaran Informasi Suku Agama Ras Dan Antargolongan Yang Menimbulkan Permusuhan Dan Kebencian, *8*(2), 187–194.
- Juditha, C. (2017). Hatespeech Di Media Online : Kasus Pilkada DKI Jakarta 2017 hatespeech in online media: jakarta on election 2017.
- Krisandi, N., Prihandono, B., & Bayes, N. (2013). Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT.Minamas, *2*(1), 33–38.
- Kurniawan, H. (2012). Sistem Penentuan Kualitas Air Pada Depot Air Minum Menggunakan Metode K-Nearest Neighbor.
- Mangantibe, V. (2016). Ujaran Kebencian Dalam Surat Edaran Kapolri Nomor: SE/6/X/2015 Tentang Penanganan Kebencian (Hate Speech), *V*(1), 159–162.
- Mooney, U. Y. N. and R. J. (2006). *Text Mining with Information Extraction*.
- Movanita, A. N. K. (2017). 11 Kasus Ujaran Kebencian dan Hoaks yang Menonjol Selama 2017 Halaman all - Kompas Retrieved from <https://nasional.kompas.com/read/2017/12/24/23245851/11-kasus-ujaran-kebencian-dan-hoaks-yang-menonjol-selama-2017?page=all>
- Permanasari, N. (2018). Wiranto: Ada 53 Kasus Hoax dan 324 Hate Speech Sepanjang 2018. Retrieved from <https://news.detik.com/berita/d-4272642/wiranto-ada-53-kasus-hoax-dan-324-hate-speech-sepanjang-2018>
- Putri, P. A. (2013). Implementasi Metode Improved K-Nearest Neighbor Pada Analisis Sentimen *Twitter* Berbahasa Indonesia.
- Rahman, U. (2015). *Klasifikasi Iklan Pada Twitter Menggunakan k-nearest*

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.



neighbor. Universitas Islam Negeri Sultan Syarif Kasim RIAU.

Rohman, A. (2012). Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa.

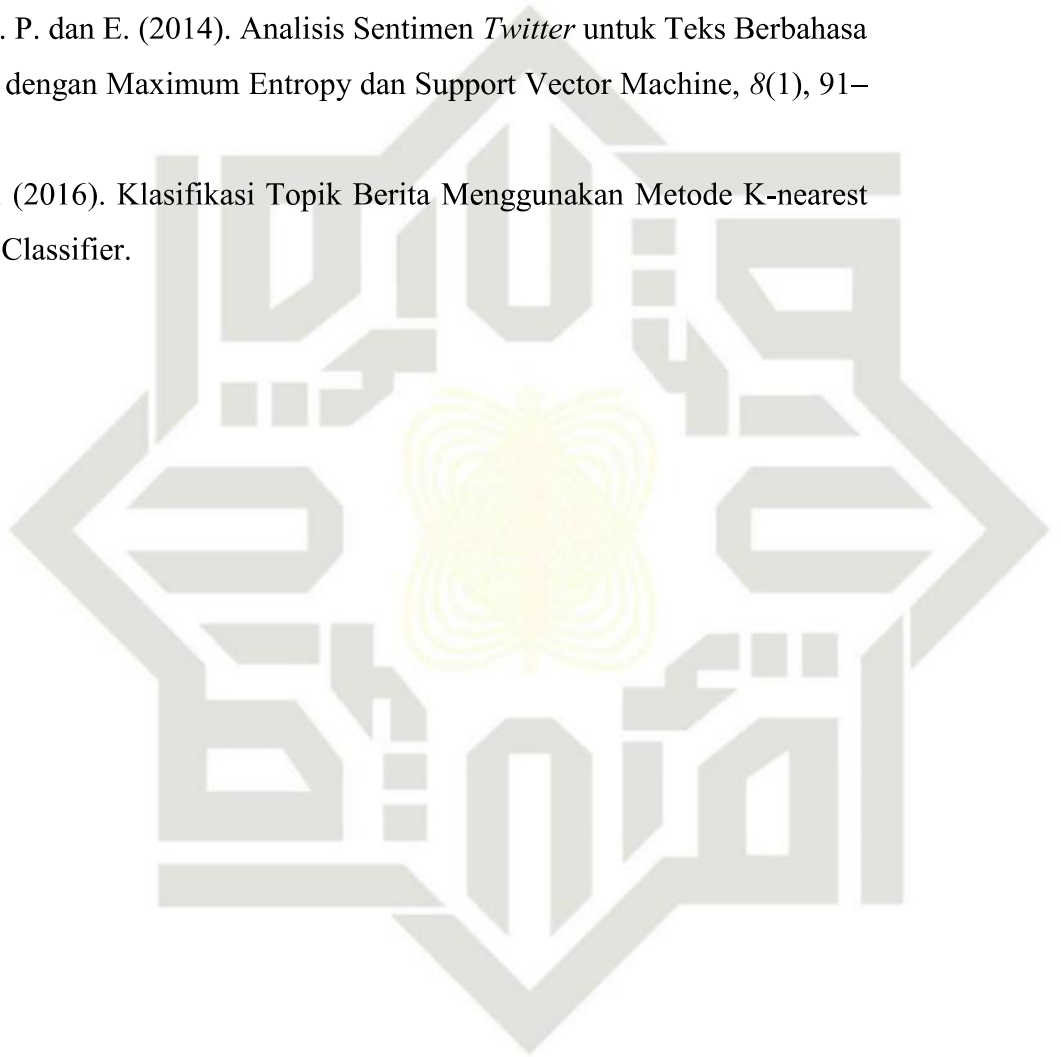
Fan, P., & Steinbach, M. (2006). Introduction to Data Mining Instructor ' s Solution Manual.

Winarko, N. D. P. dan E. (2014). Analisis Sentimen *Twitter* untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine, 8(1), 91–100.

Yulianto, S. E. (2016). Klasifikasi Topik Berita Menggunakan Metode K-nearest Neighbor Classifier.

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.



UIN SUSKA RIAU

DAFTAR RIWAYAT HIDUP

Informasi Personal



Nama : Fanrian Raymondra
 Tempat/Tgl Lahir : Duri, 08 Oktober 1994
 Jenis Kelamin : Laki-laki
 Agama : Islam
 Tinggi Badan : 171 cm
 Kewarganegaraan : Indonesia

Alamat

Sekarang : Jalan Guna Karya, No 13, Tampan, Pekanbaru
 Asal : Duri, Riau
 No. Hp : 082382209798
 Email : fanrian.raymondra@students.uin-suska.ac.id

Informasi Pendidikan

Tahun 2000 – 2001 : TK Aisyiyah 1 Cabang Duri, Kabupaten Bengkalis, Riau
 Tahun 2001 – 2007 : SDN 49 Mandau, Kabupaten Bengkalis, Riau
 Tahun 2007 – 2010 : SMPN 1 Mandau, Kabupaten Bengkalis, Riau
 Tahun 2010 – 2013 : SMAN 4 Mandau, Kabupaten Bengkalis, Riau
 Tahun 2013 – 2020 : S1 Teknik Informatika, Fakultas Sains dan Teknologi,
 Universitas Islam Negeri 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.