

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 HATE SPEECH DAN ABUSIVE LANGUAGE
PADA TWITTER BAHASA INDONESIA DENGAN METODE
NAIVE BAYES CLASSIFIER**

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

TAMARA GHASSANI SASKIA

NIM. 11551200583



UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2021**

LEMBAR PERSETUJUAN

KLASIFIKASI HATE SPEECH DAN ABUSIVE LANGUAGE PADA TWITTER BAHASA INDONESIA DENGAN METODE NAIVE BAYES CLASSIFIER

TUGAS AKHIR

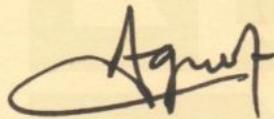
Oleh

TAMARA GHASSANI SASKIA

NIM. 11551200583

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 28 Oktober 2021

Pembimbing,



Surya Agustian, S.T, M.Kom

NIP. 19760830 201101 1 003

- 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.

LEMBAR PENGESAHAN

KLASIFIKASI HATE SPEECH DAN ABUSIVE LANGUAGE PADA TWITTER BAHASA INDONESIA DENGAN METODE NAIVE BAYES CLASSIFIER

TUGAS AKHIR

Oleh

TAMARA GHASSANI SASKIA

NIM. 11551200583

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 28 Oktober 2021

Mengesahkan,

Ketua Jurusan,

Iwan Iskandar, M.T

NIP. 19821216 201503 1 003



Dr. Hartono, M.Pd

NIP. 19640301 199203 1 003

DEWAN PENGUJI

Ketua	Iwan Iskandar, M.T
Pembimbing	Surya Agustian, S.T, M.Kom
Penguji I	Dr. Elin Haerani, S.T, M.Kom
Penguji II	Iis Afrianty, S.T, M.Sc



Lampiran Surat :
 Nomor : Nomor 25/2021
 Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Tamara Ghassani Saskia
 NIM : 11551200583
 Tempat/Tgl. Lahir : Medan / 12 Juni 1997
 Fakultas/Pascasarjana : Sains dan Teknologi
 Prodi : Teknik Informatika

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

KLASIFIKASI HATE SPEECH DAN ABUSIVE
LANGUANGE PADA TWITTER BAHASA INDONESIA
DENGAN METODE NAIVE BAYES CLASSIFIER

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.

30 November 2021

Pekanbaru,
 Yang membuat pernyataan



Tamara Ghassani
 Tamara Ghassani Saskia
 NIM : 11551200583

* pilih salah satu sesuai jenis karya tulis

Hak Cipta Dilindungi Undang-Undang
 © Hak Cipta milik 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.

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 Oktober 2021

Yang membuat pernyataan,



Tamara Ghassani Saskia

NIM. 11551200583

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 Rabbil ‘Alamin, puji syukur yang teramat dalam kepada Allah SWT, yang telah memberikan salah satu anugerah terindah yang tidak akan pernah terlupakan seumur hidup.

Kepada kedua Orang Tua tercinta, terima kasih selalu berjuang, memberikan kasih sayang dan do’a yang tak ternilai harganya, yang akan selalu dibutuhkan kapanpun dan dimanapun. Dari lubuk hati yang paling dalam, ku persembahkan gelar sarjan ini.

Kepada keluarga tersayang, terima kasih telah menjadi pendengar yang baik, memberi motivasi serta semangat yang sangat berarti bagi penulis hingga penulis bisa sampai ke titik ini.

Kepada teman-teman seperjuangan, terima kasih karena selalu ada dan berjuang bersama. Mari terus berjuang bersama-sama, karena bersama kita bisa.

Hak Cipta Diindungi 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.

ABSTRAK

Twitter merupakan salah satu media sosial yang sering digunakan oleh pengguna Indonesia untuk menyebarkan *hate speech* dan *abusive language*. Untuk mendeteksi label *tweetnya*, dapat dilakukan teknik klasifikasi teks dengan menggunakan metode *Naïve Bayes Classifier*. Klasifikasi dilakukan sebanyak tiga kali di dalam penelitian ini, yaitu klasifikasi *hate speech*, *abusive language*, dan *level hate speech*. Kemudian, *feature selection* dilakukan untuk mencari model dengan akurasi terbaik. Hasil penelitian ini menunjukkan bahwa akurasi yang tertinggi untuk klasifikasi *hate speech* sebesar 82,992%, klasifikasi *abusive language* sebesar 86,484%, dan klasifikasi *level hate speech* sebesar 71,875%. Penerapan klasifikasi teks dengan metode *Naïve Bayes Classifier* untuk klasifikasi *hate speech* dan *abusive language* pada *Twitter* bahasa Indonesia dapat disimpulkan berhasil karena hasil akurasi yang tinggi.

Kata kunci: akurasi, *abusive language*, *hate speech*, klasifikasi, *naïve bayes classifier*, *twitter*

ABSTRACT

Twitter is one of the social media that is often used by Indonesian users to spread hate speech and abusive language. To detect the tweet label, this study use a text classification technique called Naïve Bayes Classifier. The classification in this study is using in three times, namely the classification of hate speech, abusive language, and level hate speech. Then, feature selection is using to find the model classification with the best accuracy. The results of this study show that the highest accuracy for the classification of hate speech is 82.992%, the classification of abusive language is 86.484%, and the classification level hate speech is 71.875%. The application of text classification using Naïve Bayes Classifier method for the classification of hate speech and abusive language on Indonesian Twitter can be concluded to be successful because of the high accuracy results.

Keywords: accuracy, abusive language, classification, hate speech, naive bayes classifier, twitter

Hak Cipta Diindungi 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.

KATA PENGANTAR

Assalammu 'alaikum wa rohmatullohi wa barokatuh.

Alhamdulillah robbil' alamin, tak henti-hentinya penulis ucapkan kehadiran Allah *Subhanahu wa ta'ala*, yang dengan rahmat dan hidayah-Nya kami mampu menyelesaikan Tugas Akhir ini dengan baik. Tidak lupa bershalawat kepada Nabi dan Rasul-Nya, Nabi Muhammad *Sholallohu 'alaihi wa salam*, yang telah membimbing kita sebagai umatnya menuju jalan kebaikan.

Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar sarjana pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Banyak sekali pihak yang telah membantu kami dalam penyusunan laporan ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada kami. Semua itu tentu terlalu banyak bagi penulis untuk membalasnya, namun pada kesempatan ini kami hanya dapat mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Hairunas, M.Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Iwan Iskandar, M.T, selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Bapak Muhammad Affandes, M.T, selaku Pembimbing Akademis penulis selama menjalani perkuliahan di Jurusan Teknik Informatika, yang telah banyak memberikan arahan serta masukan mengenai perkuliahan bagi penulis.
5. Bapak Surya Agustian, S.T, M.Kom, selaku dosen pembimbing Tugas Akhir penulis, yang telah sangat banyak berbagi waktu, ilmu dan wawasan yang dimiliki kepada penulis sehingga penelitian dan Laporan Tugas Akhir ini selesai.

Hak Cipta Diindungi 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.

6. Ibu Dr. Elin Haerani, S.T, M.Kom, selaku dosen penguji I yang telah meluangkan waktunya dan banyak memberikan saran, dan masukan yang membangun bagi penulis.
7. Ibu Iis Afrianty, ST, M.Sc, selaku dosen penguji II yang telah meluangkan waktunya dan banyak memberikan saran, dan masukan yang membangun bagi penulis.
8. Seluruh Bapak/Ibu dosen Teknik Informatika yang telah sabar memberikan tunjuk ajar serta ilmu yang bermanfaat kepada penulis selama masa perkuliahan.
9. Keluarga penulis yang tidak pernah letih memberikan semangat, nasehat, dan doa yang tulus.
10. Teman-teman seperjuangan angkatan 2015 yang berjuang dan saling mendukung, serta semua pihak yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat kami harapkan untuk kesempurnaan laporan ini. Akhirnya penulis berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya.

Wassalamu 'alaikum wa rohmatullohi wa barokatuh.

Pekanbaru, 28 Oktober 2021



Penulis

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN.....	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK.....	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xvi
DAFTAR RUMUS	xvii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
BAB 2 KAJIAN PUSTAKA	5
2.1. Kajian Metode.....	5
2.1.1. <i>Twitter</i>	5
2.1.2. <i>Hate Speech</i>	5
2.1.3. <i>Level Hate Speech</i>	6
2.1.4. <i>Abusive Language</i>	7
2.1.5. <i>Natural Language Processing</i>	7
2.1.6. Klasifikasi Teks.....	8
2.1.7. <i>Bag of Word</i>	9
2.1.8. <i>Feature Selection</i>	9

- 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 Diindungi 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.

2.1.9. <i>Naïve Bayes Classifier</i>	11
2.1.10. <i>Confusion Matrix</i>	13
2.2. Penelitian Terkait	14
BAB 3 METODOLOGI PENELITIAN	19
3.1 Identifikasi Masalah.....	20
3.2 Pengumpulan Data	20
3.2.1 Dataset <i>Hate Speech</i> dan <i>Abusive Language</i>	20
3.2.2 Studi Pustaka	22
3.3 Analisa	22
3.3.1 Persiapan Data.....	23
3.3.2 <i>Feature Selection</i>	23
3.3.3 Pembobotan Kata	25
3.3.4 Klasifikasi.....	25
3.4 Perancangan	26
3.5 Implementasi.....	26
3.6 Pengujian.....	27
3.7 Kesimpulan dan Saran	27
BAB 4 PEMBAHASAN	28
4.1 Analisa	28
4.1.1 Kebutuhan Dataset	28
4.1.2 <i>Feature Selection</i>	32
4.1.3 Pembobotan Token/Kata	35
4.2 Perancangan	36
4.2.1 <i>Naïve Bayes Classifier</i>	36
4.3 Implementasi.....	43
4.3.1 Batasan Implementasi	43
4.3.2 Lingkungan Implementasi.....	43
4.3.3 Implementasi Model <i>Naïve Bayes Classifier</i>	43
4.4 Pengujian.....	57
4.4.1 Skenario Pengujian.....	57

4.4.2 Hasil Pengujian	63
BAB 5 PENUTUP	68
5.1. Kesimpulan	68
5.2. Saran	68
DAFTAR PUSTAKA	69
LAMPIRAN A.....	73
LAMPIRAN B.....	81
DAFTAR RIWAYAT HIDUP	89

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 1 <i>Confusion matrix</i>	13
Tabel 2 Penelitian Terkait	14
Tabel 3 <i>Case folding</i>	23
Tabel 4 <i>Removal punctuation</i>	24
Tabel 5 <i>Removal stopwords</i>	24
Tabel 6 Tokenisasi	25
Tabel 7 Contoh <i>tweet hate speech</i> dan <i>abusive language</i>	28
Tabel 8 Dataset awal <i>tweet hate speech</i> dan <i>abusive language</i>	29
Tabel 9 Kombinasi <i>feature selection</i>	34
Tabel 10 Contoh <i>feature selection</i>	34
Tabel 11 Contoh <i>tweet training</i>	35
Tabel 12 <i>Bag of word</i>	35
Tabel 13 Contoh <i>tweet testing</i>	42
Tabel 14 Perbandingan hasil prediksi label dengan label dataset pada klasifikasi <i>hate speech</i> dan <i>abusive language</i>	63
Tabel 15 Hasil akurasi klasifikasi <i>level hate speech</i>	65

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 RUMUS

(1) Persamaan Prior Probability.....	11
(2) Persamaan Probabilitas Tiap Kata.....	12
(3) Persamaan Probabilitas Kategori.....	12
(4) Persamaan <i>Maximum a Posterior</i>	12
(5) Persamaan Akurasi.....	13

Hak Cipta Diindungi 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 1

PENDAHULUAN

1.1. Latar Belakang

Media sosial adalah suatu alat digital yang digunakan pengguna dalam berkomunikasi, membuat, dan berbagi konten secara terbuka dan cepat [1]. Media sosial melingkupi berbagai situs web dan aplikasi. Siapapun yang memiliki akses internet dapat menggunakan media sosial. Media sosial tidak hanya memberi dampak positif seperti dapat berinteraksi dengan cepat, tetapi dapat memberi dampak negatif [2]. Salah satunya yaitu adanya kejahatan seperti *hate speech* (ujaran kebencian). Saat ini *hate speech* semakin besar yang dilakukan oleh individu atau kelompok baik lisan maupun tulisan yang memicu dampak sangat parah, antara lain pengucilan, permusuhan, kekerasan, bahkan pembantaian [3]. Di Indonesia, surat penanganan ujaran kebencian sudah ada dalam “Surat Edaran Kapolri Nomor: SE/06/X/2015” [4].

Pada penelitian ini media sosial yang digunakan adalah *Twitter* karena *Twitter* merupakan salah satu media sosial yang sering digunakan oleh pengguna Indonesia untuk menyebarkan *hate speech* dan *abusive language* [5]. *Twitter* merupakan alat berbagi konten dan pesan tertulis dalam format singkat. *Twitter* telah memainkan peran penting dalam memberikan informasi kepada pengguna serta mengubah cara orang menyampaikan informasi [6]. Dikutip dari situs [statista.com](https://www.statista.com) per April 2021, jumlah pengguna aktif di Indonesia yaitu 15,1 juta dari 2,79 milyar pengguna dunia. Karena jumlah ini termasuk banyak, maka ini dapat memberi pengaruh pada pengguna lain, termasuk hal negatif.

Hate Speech (Ujaran Kebencian) adalah ujaran langsung atau tidak langsung baik secara lisan maupun tulisan terhadap target yang memuat kebencian [3]. Penyebaran ujaran kebencian sering disertai dengan *abusive language*, walaupun tidak semua [7]. *Abusive Language* (Bahasa Kasar) adalah ujaran yang memuat frasa atau kata-kata kasar yang sensitif terhadap target. Munculnya kata-kata kasar biasanya berasal dari kondisi yang tidak menyenangkan seperti penyimpangan mental. *Hate speech* yang mengandung *abusive language* sering memicu emosi sehingga mempercepat terjadinya konflik [5].



Hak Cipta Diindungi 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.

Banyak terjadi kasus ujaran kebencian di media sosial *Twitter* Indonesia. Beberapa diantaranya yaitu dikutip dari situs jawapos.com pada bulan September 2019, sutradara film dokumenter *Sexy Killers* bernama Dhandy Laksono ditangkap polisi karena *tweet*-nya diindikasikan memuat ujaran kebencian terkait kerusuhan di Papua. Selanjutnya dikutip dari situs nasional.okezone.com pada bulan Januari 2019, salah satu musisi terkenal bernama Ahmad Dhani ditetapkan sebagai tersangka dalam menyebarkan ujaran kebencian terhadap mantan Gubernur Jakarta, Ahok. Selanjutnya dikutip msn.com pada bulan September 2019, seorang pegawai swasta ditangkap polisi usai *mentweet* ujaran kebencian yang mengandung rasis terhadap warga Papua. Selanjutnya dikutip dari kompas.com, seorang pengguna media sosial bernama Ibbas Kiswotomo telah ditangkap oleh Bareskrim Polri atas dugaan tindakan penyebaran ujaran kebencian di *Twitter* dan *Facebook*.

Oleh karena itu klasifikasi *hate speech* dan *abusive language* perlu dibuat dalam bentuk sistem supaya dapat mengurangi konflik antar warga dan anak-anak dibawah umur yang dikhawatirkan akan meniru pengucapan ujaran kebencian dan bahasa kasar di dunia nyata maupun media sosial [3] [8]. Selain *tweet hate speech* dan *abusive language*, tetapi lebih baik perlu juga mengetahui *level tweet hate speech* (tingkat) tersebut. Ini perlu dilakukan untuk membantu pihak wewenang dalam memberi hukuman yang sesuai kepada pelaku berdasarkan *level tweetnya* yang dilontarkannya serta memprioritaskan *tweet* yang harus diatasi terlebih dahulu. Ada tiga jenis *level tweet hate speech* berdasarkan penelitian Ibrohim & Budi (2018), yaitu *weak*, *moderate*, dan *strong*.

Penelitian ini juga bertujuan untuk mengikuti model penelitian Han dkk (2019) dan Ibrohim & Budi (2019). Han dkk (2019) melakukan identifikasi dan kategorisasi *tweet abusive language* dalam bahasa Inggris dengan menggunakan beberapa metode *machine learning* dan *deep learning*. Mereka membagi tugas menjadi tiga. Subtugas A mengklasifikasi *tweet* yang termasuk *abusive language* atau tidak, Subtugas B kategorisasi *tweet abusive language*, dan Subtugas C mengidentifikasi target *tweet abusive language* [9]. Ibrohim & Budi (2019) melakukan klasifikasi *tweet hate speech* dan *abusive language* dengan beberapa

Hak Cipta Diindungi 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.

metode *machine learning* dengan ekstraksi fitur. *Dataset tweet* terdiri dari *tweet* yang termasuk ujaran kebencian atau bukan dan bahasa kasar atau bukan, lalu *tweet* tersebut mengandung beberapa kelas target, kategori, dan tingkat. Penelitiannya melakukan dua eksperimen, yang pertama klasifikasi ujaran kebencian dan bahasa kasar termasuk target, kategori, dan tingkat lalu yang kedua klasifikasi ujaran kebencian dan bahasa kasar tanpa target, kategori, dan tingkat [5].

Metode klasifikasi teks yang digunakan adalah *Naïve Bayes Classifier* (NBC). NBC merupakan sebuah metode klasifikasi bersifat probabilitas. NBC termasuk metode yang sederhana dalam perhitungan namun memberikan hasil akurasi yang bagus. Dengan menggunakan teori probabilitas sebagai dasar, Han, J. dan Kamber, M. menyatakan di dalam bukunya bahwa “*Bayesian classifiers mempunyai tingkat kecepatan dan akurasi yang bagus ketika diimplementasikan dalam database yang besar*”. Beberapa penelitian yang terkait metode NBC. Penelitian Ibrohim & Budi (2018) melakukan klasifikasi *abusive language* dengan menggunakan metode NBC, SVM, dan RFDT. Jumlah dataset *tweet*-nya adalah 2.500 *tweet*. Hasilnya adalah NBC lebih baik daripada SVM dan RFDT di semua percobaan di penelitian tersebut, dengan kombinasi *word n-gram + bigram* memberikan hasil *F-score* terbaik yaitu 70,06% [10]. Penelitian Alfina dkk (2017) melakukan klasifikasi *tweet* yang mengandung *hate speech* atau bukan. Hasilnya adalah *F-measure* terbaik dicapai ketika menggunakan *word n-gram* dengan beberapa metode salah satunya NBC dengan persentasi tinggi yaitu 90,2% [11]. Fitri (2019) melakukan penerapan tingkat akurasi metode K-NN dan NBC dalam mengklasifikasi *tweet hate speech* terhadap agama Islam. Hasil akurasi metode NBC adalah 95% dengan pembagian data latih dan uji 90:10. Sedangkan hasil akurasi metode K-NN adalah 94% dengan memakai fitur *threshold* dengan pembagian data latih dan uji 80:20, sedangkan tanpa fitur *threshold* adalah 82% [12].

Berdasarkan uraian diatas, maka peneliti melakukan penelitian klasifikasi *hate speech* dan *abusive language* pada *Twitter* bahasa Indonesia dengan metode *Naïve Bayes Classifier*.



1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah pada penelitian ini yaitu:

1. Bagaimana cara klasifikasi *hate speech* dan *abusive language* pada *Twitter* bahasa Indonesia dengan metode *Naïve Bayes Classifier*?
2. Bagaimana perhitungan nilai akurasi pada klasifikasi *hate speech* dan *abusive language* pada *Twitter* bahasa Indonesia?

1.3. Batasan Masalah

Supaya permasalahan dalam penelitian ini tidak meluas dan menyimpang, maka dibutuhkan batasan masalah. Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Klasifikasi *tweet* yang diambil dari dataset penelitian Ibrahim & Budi (2019) yaitu klasifikasi *hate speech* (label *hate speech* dan *non hate speech*); klasifikasi *abusive* (label *abusive* dan *non abusive*); dan klasifikasi *level hate speech* (label *level weak*, *level moderate*, dan *level strong*).
2. *Feature Selection* yang digunakan adalah *Case Folding*, *Removal Punctuation*, *Removal Stopword*, dan Panjang Karakter Token.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini antara lain:

1. Mengklasifikasi *hate speech* dan *abusive language* pada *twitter* bahasa Indonesia dengan menggunakan metode *Naïve Bayes Classifier*.
2. Menghitung nilai akurasi pada klasifikasi *hate speech* dan *abusive language* pada *Twitter* bahasa Indonesia.

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian ini antara lain:

1. Melihat hasil klasifikasi *hate speech* dan *abusive language* pada *twitter* bahasa Indonesia dengan menggunakan metode *Naïve Bayes Classifier*.
2. Melihat nilai akurasi pada klasifikasi *hate speech* dan *abusive language* pada *Twitter* bahasa Indonesia.



BAB 2

KAJIAN PUSTAKA

2.1. Kajian Metode

Kajian metode berisi kerangka teori yang digunakan pada penelitian ini. Kajian metode diperoleh dari berbagai referensi. Berikut kajian metode pada penelitian ini:

2.1.1. *Twitter*

Twitter adalah sebuah situs *microblogging* yang mana pengguna dapat *memposting* singkat berupa *tweet* ke pengguna lain [13]. *Twitter* didirikan oleh Jack Dorsey pada tanggal 21 Maret 2006. Sejak saat itu, *Twitter* menjadi salah satu media sosial yang populer di dunia dan menarik sebagian besar pengguna untuk berbagi serta mengekspresikan ide dan perasaan seperti peristiwa dunia, kampanye politik, media dakwah, media komunikasi darurat, dan lain-lain secara *real time*. Awal mula karakter *tweet* berjumlah 140 karakter, lalu kini menjadi 280 karakter sejak tahun 2017 [14]. Seorang pengguna dapat mengikuti pengguna lain dan pengguna yang diikuti tidak perlu mengikuti kembali. Setiap pengguna dapat mengikuti pengguna lain untuk menerima *tweet* dari pengguna lain dengan menjadi *follower* pengguna yang diikuti [13]. *Tweet entity* atau konten *tweet* adalah isi dari *tweet* itu sendiri. *Tweet entity* terdapat teks, *retweet* (RT), *mention* pengguna (@), URL, *emoticons*, dan media.

2.1.2. *Hate Speech*

Hate Speech (Ujaran Kebencian) adalah ujaran langsung atau tidak langsung baik secara lisan maupun tulisan terhadap target yang memuat kebencian berdasarkan hal yang melekat pada target [3]. Pengertian lainnya adalah suatu komunikasi yang meremehkan karakteristik individu atau kelompok dalam hal ras, agama, kebangsaan, dan hal lainnya [11]. Saat ini *hate speech* semakin meluas yang dilakukan oleh individu atau kelompok baik lisan maupun tulisan. Dampak yang ditimbulkan sangat fatal antara lain pengucilan, permusuhan, kekerasan, bahkan pembantaian [3]. Di Indonesia, surat penanganan *hate speech*



Hak Cipta Diindungi 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.

telah diatur dalam Surat Edaran Kapolri Nomor: SE/06/X/2015 [4]. Salah satu yang isi surat edaran tersebut adalah terdapat Pasal 28 ayat 2 dan Pasal 45 UU Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik yang berbunyi [4]:

Pasal 28 ayat 2

“Setiap orang dengan sengaja dan tanpa hak menyebarkan informasi yang ditujukan untuk menimbulkan rasa kebencian dan permusuhan individu dan kelompok masyarakat tertentu berdasarkan atas suku, agama, ras, dan antar goloingan (SARA).”

Pasal 45

“Setiap orang yang memenuhi unsur sebagaimana yang dimaksud dalam Pasal 28 ayat 1 dan 2 dipidana dengan pidana penjara paling lama 6 tahun dan/atau denda paling banyak Rp,1.000.000.000.”

2.1.3. *Level Hate Speech*

Menurut Komnas HAM, *level hate speech* (tingkat ujaran kebencian) merupakan salah satu unsur yang terkandung dengan tujuan untuk membantu pihak wewenang dalam memprioritaskan ujaran *hate speech* yang harus diatasi terlebih dahulu dan memberi hukuman yang tepat sesuai dengan bentuk tingkat *hate speech* yang dilontarkan oleh pelaku [3]. *Level hate speech* terbagi menjadi tiga, yaitu [5]:

1. *Level Weak* (Tingkat Lemah). Kebencian tingkat ini dalam bentuk penistaan, sumpah, stereotip, atau pelabelan terhadap inividu tanpa provokasi,atau hasutan yang dapat memicu permusuhan. Di Indonesia, ini termasuk tingkat lemah karena masalah pribadi. Jadi jika target ujaran kebencian tidak melapor kepada pihak wewenang (seperti target merasa ini kebencian yang biasa saja), maka ujaran kebencian ini tidak terlalu diprioritaskan untuk diatasi.
2. *Level Moderate* (Tingkat Sedang). Kebencian tingkat ini dalam bentuk penistaan, sumpah, stereotip, atau pelabelan terhadap kelompok tanpa disertai provokasi,atau hasutan yang dapat memicu permusuhan. Walaupun



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. Case Folding

Merupakan salah satu bentuk *text preprocessing* yang paling sederhana dan efektif. *Case folding* berperan dalam mengkonversi huruf-huruf yang berada didalam dokumen dengan format huruf besar (*uppercase*) menjadi huruf-huruf kecil (*lowercase*). Hanya huruf 'a' sampai 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimiter*.

2. Removal Punctuation

Punctuation merupakan tanda baca atau simbol. Contohnya titik (.), koma (,), tanda tanya (?). Emoticon juga termasuk dalam punctuation karena sebuah emoticon terdiri dari gabungan beberapa karakter atau tanda baca. *Removal Punctuation* menghapus kata-kata tanda baca.

3. Removal Stopword

Stopword yaitu kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopword* dalam Bahasa Indonesia adalah 'yang', 'dan', 'di', 'dari', dan lain-lain. *Removal Stopword* menghapus kata-kata *stopword* yang memiliki informasi rendah dari sebuah teks agar kita dapat fokus pada kata-kata penting lainnya.

4. Tokenisasi

Tokenisasi bertugas untuk pemotongan kalimat menjadi bagian-bagian yang disebut token [18]. Token bisa berupa karakter atau kata, tergantung jenis tokenisasi yang mau diproses. Contoh token kata: "Kamu makan nasi yang enak", setelah dilakukan tokenisasi maka kalimat tersebut akan menjadi "Kamu", "makan", "nasi", "yang", "enak".

2.1.6. Klasifikasi Teks

Klasifikasi teks merupakan suatu tugas untuk menetapkan kategori yang telah ditentukan sebelumnya ke dokumen teks berdasarkan konten teksnya [19]. Klasifikasi bisa disebut kategorisasi. Seluruh atau bagian dari dokumen teks seperti paragraf dapat dibagi dalam banyak kelas. Tujuannya adalah menguji klasifikasi teks yang belum diketahui labelnya.

Terdapat dua tahap dalam melakukan klasifikasi, yaitu:



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. Tahap *training*, berupa pelatihan *tweet* yang akan menjadi representasi data tersebut. Tujuannya untuk membangun model klasifikasi yang akan diimplementasikan ke tahap pengujian [12].
2. Tahap *testing*, berupa pengujian data yang belum diketahui labelnya berdasarkan *term* yang muncul [12]. Pengujian ini dilakukan dengan menggunakan model klasifikasi yang telah dibentuk di tahap *training*. Jadi jika ada teks yang baru, dapat diklasifikasikan pada suatu label berdasarkan teks-teks yang sudah ada sebelumnya [20].

2.1.7. *Bag of Word*

Bag of Word (BoW) adalah metode pembobotan teks yang mempresentasikan kumpulan kata yang tidak teratur, mengabaikan kata, dan urutan kata [19]. Dalam klasifikasi teks, sebuah kata dalam dokumen diberi bobot sesuai dengan frekuensinya dan frekuensi diantara dokumen yang berbeda. Kumpulan kata tersebut membentuk *Bag of Word*.

Berikut contoh model *Bag of Word*, adalah sebagai berikut:

D1: Ayah adalah seorang guru. Ayah hebat.

D2: Ibu adalah seorang penjual makanan.

Berdasarkan dua dokumen teks diatas, *vocab* dan frekuensi yang terbentuk adalah:

{ Ayah: 1 - adalah: 2 - seorang: 3 - hebat: 4 – Ibu: 5 – penjual: 6 – makanan: 7 }

Terdapat 7 kata yang berbeda. Setiap dokumen direpresentasi sebagai vektor 7 elemen. Hitung ada/tidaknya kata setiap dokumen dengan nilai 1 jika ada dan 0 jika tidak ada. Setiap entri vektor mengacu pada jumlah entri yang sesuai dengan *vocab* [19].

[1, 1, 1, 1, 0, 0, 0,]

[0, 1, 1, 0, 1, 1, 1]

2.1.8. *Feature Selection*

Feature Selection juga dikenal sebagai pemilihan variabel, pemilihan atribut atau pemilihan subset variabel, merupakan sebuah proses dalam *machine learning* yang menggunakan sekumpulan fitur yang dimiliki oleh data untuk digunakan dalam proses algoritma. *Feature selection* telah menjadi bidang penelitian aktif

yang digunakan dalam pengenalan pola, statistik, dan data *mining* [21]. Teknik *feature selection* digunakan karena beberapa alasan:

1. Penyederhanaan model agar lebih mudah diinterpretasikan oleh pengguna
2. Waktu pelatihan yang lebih singkat
3. Meningkatkan kompatibilitas data dengan kelas model pembelajaran,

Premis utama saat menggunakan teknik pemilihan fitur adalah bahwa data berisi beberapa fitur yang berlebihan atau tidak relevan, dan dengan demikian dapat dihapus tanpa menimbulkan banyak kehilangan informasi. Redundan dan tidak relevan adalah dua pengertian yang berbeda, karena satu fitur yang relevan mungkin berlebihan dengan adanya fitur lain yang relevan yang berkorelasi kuat.

Feature Selection merupakan salah satu faktor penting yang dapat mempengaruhi tingkat akurasi dari klasifikasi. Apabila dalam *dataset* terdapat sejumlah fitur, dimensi *dataset* akan menjadi besar dan dapat menyebabkan rendahnya akurasi yang didapat. Permasalahan utama dalam *feature selection* adalah pengurangan dimensi, dimana awalnya semua atribut diperlukan untuk mendapatkan hasil yang maksimal.

Terdapat beberapa alasan mengapa perlu dilakukan pengurangan dimensi menurut [21]:

1. Penurunan biaya pelatihan algoritma (*decreasing the training cost*),
2. Meningkatkan kinerja pelatihan algoritma (*increasing the training performance*),
3. Mengurangi dimensi yang tidak relevan (*reducing irrelevant dimensions*), dan
4. Mengurangi dimensi yang berlebihan (*reducing redundant dimensions*).

Ide utama dari *feature selection* yaitu memilih subset dari fitur yang ada tanpa transformasi, karena tidak semua fitur relevan dengan masalah. Bahkan ada fitur atau atribut yang mengganggu dan dapat mengurangi akurasi. *Noisy feature* atau fitur yang tidak dipakai tersebut dapat dihapus agar dapat meningkatkan akurasi dan mempercepat proses komputasi.

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.

2.1.9. Naïve Bayes Classifier

Metode *Naive Bayes Classifier* (NBC) merupakan salah satu metode klasifikasi *machine learning* jenis *supervised learning*. *Supervised learning* merupakan teknik yang menggunakan *training* data sudah berlabel dengan tujuan untuk melakukan prediksi maupun klasifikasi [22]. Tokoh yang pertama kali kemukakan metode ini adalah seorang ilmuwan dari Inggris bernama Thomas Bayes. Metode ini bermodel probabilitas dan statistik yang dikenal dengan penerapan Teorema Bayes [23]. Ini didasarkan pada asumsi penyederhanaan bahwa jika diberikan nilai output, nilai atribut secara kondisional saling bebas [24]. Metode ini akan menghitung probabilitas setiap data kemudian diklasifikasikan nilai probabilitas tertinggi. Dengan struktur *Naive Bayes Classifier* yang sederhana, metode ini memiliki waktu yang singkat untuk memproses dan tingkat akurasi yang tinggi algoritma ini mudah untuk diimplementasikan [25]. Proses klasifikasi dalam penelitian ini akan memakai metode *Multinomial Naive Bayes*. *Multinomial* NBC merupakan sebuah model pengembangan dari algoritma *bayes* yg cocok dalam pengklasifikasian teks atau dokumen [26]. Pada formula *Multinomial Naive Bayes Classifier*, kelas dokumen tidak hanya ditentukan dengan kata yang muncul tetapi juga jumlah kemunculannya [27]. *Bayes Naive Multinomial* digunakan dalam penelitian ini dikarenakan metode ini bagus digunakan dalam proses klasifikasi dan cocok digunakan terhadap dokumen teks [28].

Persamaan untuk menghitung *prior probability* atau probabilitas pada masing-masing kelas ditunjukkan oleh persamaan:

$$P(C_i) = \frac{|doc C_i|}{|D|} \quad (1)$$

Keterangan:

- $P(C_i)$: Probabilitas suatu kategori dokumen.
 $Doc C_i$: Jumlah dari dokumen untuk tiap kategori C_i .
 $|D|$: Jumlah dari seluruh dokumen latih.

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.

Kemudian perhitungan probabilitas tiap *term* terhadap setiap kelas dari semua dokumen. Banyak *term* tergantung dari hasil preproses data. Berikut perhitungan probabilitas tiap kata:

$$P(w_{kj}|C_i) = \frac{f(w_{kj}|C_i)+1}{f(C_i)+|W|} \quad (2)$$

Keterangan:

$P(w_{kj}|C_i)$: Probabilitas muncul kata W_{kj} pada suatu dokumen dgn kategori kelas C_i

$f(w_{kj}|C_i)$: Hasil BoW yang sudah dihitung sebelumnya pada kata W_{kj} pada suatu dokumen dgn kategori kelas C_i .

W_{kj} : Frekuensi kata ke - k pada tiap kategori.

W : Jumlah kata pada dokumen test.

$f(C_i)$: Jumlah semua kata dari suatu kategori.

$|W|$: Jumlah semua kata dari semua kategori.

Kemudian untuk melakukan perhitungan probabilitas kategori :

$$P(C_i) = P(C_i) * \prod P(w_{kj}|C_i) \quad (3)$$

Keterangan:

$P(C_i)$: Peluang munculnya suatu dokumen dengan kategori C_i

$P(w_{kj}|C_i)$: Peluang munculnya w_{kj} pada kategori C_i

MAP / *Maximum a Posterior* digunakan saat menentukan kelas dari suatu dokumen testing dgn mengambil nilai maksimum probabilitas dari setiap dokumen. Berikut rumus dalam MAP:

$$C_{MAP} = \underset{C_i \in C}{arg\ max} P(C_i) \prod P(w_{kj}|C_i) \quad (4)$$

Keterangan:

$P(C_i)$: Probabilitas dari suatu kategori dokumen

$P(w_{kj}|C_i)$: Probabilitas munculnya kata W_{kj} dalam suatu dokumen dgn kategori kelas C_i .

2.1.10. Confusion Matrix

Confusion Matrix adalah suatu tabel yang digunakan untuk menggambarkan kinerja akurasi dalam klasifikasi. Ada empat jenis representasi hasil proses klasifikasi, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TP mempresentasikan jumlah data positif yang terdeteksi benar, TN mempresentasikan jumlah data negatif yang terdeteksi benar, FP mempresentasikan jumlah data positif yang terdeteksi salah, dan FN mempresentasikan jumlah data negatif yang terdeteksi salah. TP dan TN menunjukkan hasil klasifikasi yang benar, sedangkan FP dan FN menunjukkan hasil klasifikasi yang salah.

Confusion Matrix pada penelitian ini melakukan perhitungan akurasi saja karena salah satu tujuannya adalah membandingkan hasil penelitian ini dengan penelitian [5]. Berikut ini bentuk gambaran *Confusion Matrix* di Tabel 1.

Tabel 1 *Confusion matrix*

Kelas	Terklasifikasi	
	+	-
+	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
-	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Setelah mendapatkan empat nilai di tabel diatas, dapat diperoleh nilai akurasi.

Rumus akurasi adalah:

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (5)$$

Keterangan:

TP: *True positive*, merupakan jumlah data dengan kelas positif yang diklasifikasi positif

TN: *True Negative*, merupakan jumlah data dengan kelas negatif yang diklasifikasikan

FN: *False Negative*, merupakan jumlah data dengan kelas positif diklasifikasikan negatif.

FN: *False Positive*, merupakan jumlah data dengan kelas negatif diklasifikasikan positif.

2.2. Penelitian Terkait

Berikut referensi yang terkait dengan penelitian ini:

Tabel 2 Penelitian Terkait

No	Peneliti	Judul	Metode	Kesimpulan
1	Alfina, Mulia, & Fanany (2017)	<i>Hate Speech Detection in the Indonesian Language: A Dataset and Preliminary Study</i>	<i>Naïve Bayes Classifier, Support Vector Machine, Bayesian Logistic, dan Random Forest Decision Tree</i>	Jumlah <i>dataset tweet</i> adalah 520 yang terdiri dari label ujaran kebencian atau tidak. Hasilnya adalah <i>F-measure</i> dapat dicapai ketika menggunakan <i>word n-gram</i> dengan RFDT (93,5%), BLR (91,5%) dan NBC (90,2%). Lalu urutan ekstraksi fitur terbaik yaitu <i>word n-gram</i> , <i>character n-gram</i> , dan <i>negative sentiment</i> .
2	Del Vigna, Cimino, Dell'Orletta, Petrocchi, & Teschoni (2017)	<i>Hate Me, Hate Me Not: Hate Speech Detection on Facebook</i>	<i>Support Vector Machine dan Long Short Term Memory</i>	Jenis label di penelitian tersebut yaitu bukan ujaran kebencian, ujaran kebencian yang lemah, dan ujaran kebencian yang tinggi. Hasilnya menunjukkan keefektifan dari dua metode ini yang diuji pada teks pertama yang dianalisis oleh corpus teks ujaran kebencian bahasa Italia. Klasifikasi mencapai hasil yang sebanding dengan yang diperoleh oleh sebagian analisis sentimen yang diteliti.
3	Ibrohim & Budi (2018)	<i>A Dataset and Preliminary Study for Abusive Language Detection in Indonesian Social Media</i>	<i>Naïve Bayes Classifier, Support Vector Machine, dan Decision Tree</i>	Jumlah <i>dataset tweet</i> adalah 520 yang terdiri dari label ujaran kebencian atau tidak. Selain menggunakan beberapa metode <i>machine learning</i> , penelitian ini juga menggunakan fitur <i>n-gram</i> dan <i>char n-gram</i> . Ada tiga label yang digunakan dalam klasifikasi, yaitu bukan bahasa kasar, bahasa kasar tapi tidak ofensif, dan bahasa kasar. Hasil percobaan ini menunjukkan

© Hak cipta m				bahwa <i>F-score</i> NBC dengan kombinasi <i>word unigram</i> + <i>bigram</i> menghasilkan 70,06%. Jika hanya mengklasifikasi dua label (bukan bahasa kasar dan bahasa kasar), semua metode memberikan nilai <i>F-score</i> lebih dari 83%.
4	Gaydhani, Doma, Kendre, & Bhagwat, (2018)	<i>Detecting Hate Speech and Offensive Language on Twitter using Machine Learning: An N-gram and TFIDF based Approach</i>	<i>Naïve Bayes Classifier, Support Vector Machine, dan Logistic Regression</i>	Ada tiga label yang digunakan dalam klasifikasi, yaitu ujaran kebencian, bahasa kasar, dan bersih. Peneliti menggunakan <i>n-gram</i> sebagai fitur dan TFIDF. Peneliti melakukan analisis komparatif model dengan mempertimbangkan beberapa nilai <i>n</i> dalam <i>n-gram</i> dan metode normalisasi TFIDF. Lalu peneliti melatih model pada pelatihan data dengan melakukan pencarian fitur-fitur untuk semua kombinasi dan menggunakan <i>10-fold cross validation</i> . Hasil akurasi terbaik mencapai 95,6%.
5	Kiilu, Okeyo, Rimiru, & Ogada (2018)	<i>Using Naïve Bayes Algorithm in Detection of Hate Tweets</i>	<i>Naïve Bayes Classifier</i>	Tujuan penelitian ini adalah untuk mengembangkan <i>tools</i> pendeteksi ujaran kebencian. Penelitian ini mendeteksi ujaran kebencian dengan menggunakan konten yang diproduksi oleh komunitas kebencian yang mengidentifikasi diri dari <i>Twitter</i> . Hasil penelitian ini adalah NBC dapat mencapai kinerja yang baik dengan nilai presisi, <i>recall</i> , dan akurasi masing-masing 58%, 62%, dan 67,47%.
6	Rozi, Hamdana, Alfahmi, & Balya (2018)	Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan NBC	<i>Naïve Bayes Classifier</i>	Penelitian ini menerapkan aplikasi analisis sentimen tweet dalam bahasa Indonesia dengan menggunakan metode NBC. Proses klasifikasi dilakukan oleh preprocessing terlebih dahulu. Jenis kategorinya yaitu tweet positif, negatif, dan netral.

Hak Cipta Diindungi 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 milik UIN Suska Riau				Hasil akurasi dengan jumlah data 200 tweet positif, 200 tweet negative, dan 200 tweet netral masing-masing sebesar 82%, 92%, 80%.
7	Purnamasari, Fauzi, Indriarti, & Dewi (2018)	Identifikasi <i>Tweet Cyberbullying</i> pada Aplikasi Twitter menggunakan <i>Support Vector Machine</i> dan <i>Information Gain</i>	<i>Support Vector Machine</i> dan <i>Information Gain</i>	Hasil akurasi pada penelitian ini adalah 75% untuk SVM dan 76% untuk <i>Information Gain</i> .
8	Ratnawati (2018)	Implementasi Algoritma Naïve Bayes terhadap Analisa Sentimen Film pada <i>Twitter</i>	<i>Naïve Bayes Classifier</i>	Berdasarkan hasil percobaan, hasil akurasi didapat mencapai 90% dengan rincian presisi 92%, akurasi 90% , dan <i>f-score</i> 90%.
9	Han, Wu, & Liu (2019)	<i>Identifying and Categorizing Offensive Language in Social Media</i>	Berkisar pada <i>machine learning</i> , misal: SVM, Logistic Regression; <i>deep learning</i> , misal: CNN, RNN, BiLSTM), termasuk mekanisme perhatian; hingga <i>deep learning</i> yang canggih	Penelitian ini didasarkan dengan dataset yang berisi 14.000 <i>tweet</i> bahasa Inggris. Ada tiga kelas dalam klasifikasi ini sebagai sub tugas. Sub tugas A membedakan label bahasa kasar atau tidak, sub tugas B mendeteksi label jenis bahasa kasar, dan sub tugas C mendeteksi label target bahasa kasar.

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 milik UIN Suska Riau			seperti ELMo dan BERT	
10	Ibrohim & Budi (2019)	<i>Multi-label Hate Speech and Abusive Language Detection in Indonesia Twitter.</i>	<i>Naïve Bayes Classifier, Support Vector Machine, dan Random Forest Decision Tree</i>	Jumlah <i>dataset tweet</i> -nya adalah 13.169 tweet yang terdiri dari ujaran kebencian, bahasa kasar, dan dua-duanya. Lalu <i>tweet</i> tersebut mengandung target, kategori, dan tingkat. Penelitian ini juga menggunakan beberapa metode transformasi dan ekstraksi fitur. Ada dua eksperimen. (1) Mendeteksi ujaran kebencian dan bahasa kasar termasuk target, kategori, dan tingkat. (2) mendeteksi ujaran kebencian dan bahasa kasar tanpa target, kategori, dan tingkat. Hasilnya adalah metode RFDT dengan kombinasi Lower Powerset dan <i>word n-gram</i> merupakan yang terbaik. Lalu hasil akurasi eksperimen pertama dan kedua masing-masing adalah 77,36% dan 66,12%.
11	Chalida (2019)	Analisis Sentimen Ujaran Kebencian Pemilihan Presiden 2019 Menggunakan Algoritma Naïve Bayes	<i>Naïve Bayes Classifier</i>	Hasil percobaan analisis sentimen ujaran kebencian pada data uji sejumlah 5055 data tweet dengan hastag pilpres2019 di kota Jakarta, Bandung, Semarang, Surabaya, dan Yogyakarta dengan memanfaatkan model klasifikasi dari data latih menggunakan NBC dan pembobotan TF-IDF didapatkan hasil klasifikasi sentiment irrelevant sebanyak 11,3% dengan 573 data, sentiment negatif sebanyak 35,4% dengan 1786 data, sentiment netral sebanyak 26,7% sebanyak 1343 data, dengan kecenderungan pada sentiment negatif dengan nilai terbesar yaitu 35,4%.

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.

12	Fitri (2019)	Tingkat Akurasi Metode K-NN dengan NBC oada Klasifikasi Ujaran Kebencian terhadap Agama Islam	<i>K-Nearest Neighbor</i> dan <i>Naïve Bayes Classifier</i>	Hasil akurasi metode NBC adalah 95% dengan pembagian data latih dan uji 90:10. Sedangkan hasil akurasi metode K-NN adalah 94% dengan memakai fitur threshold dengan pembagian data latih dan uji 80:20, sedangkan tanpa fitur <i>threshold</i> adalah 82%.
13	Wahyono (2018)	Fundamenta l of Python for Machine Learning: Dasar-dasar Pemrogram an Python untuk Machine Leaning dan Kecerdasan Buatan		Pengertian <i>machine learning</i> , kecerdasan buatan, jenis metode klasifikasi, serta contoh implementasi dengan menggunakan bahasa <i>python</i> .
14	Polri (2015)	Surat Edaran Kapolri Mengenai Penanganan Ujaran Kebencian		Tentang aturan Kapolri dalam menagani ujaran kebencian di Indonesia, seperti dasar ujaran kebencian, bentuk ujarankebencian, efek yang ditimbul, dan lain-lain.

Hak Cipta Diindungi 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 3

METODOLOGI PENELITIAN

Metodologi penelitian adalah langkah-langkah atau tahapan yang akan dilakukan dalam melakukan suatu penelitian. Berikut merupakan langkah-langkah yang dilakukan bisa dilihat pada Gambar 1 berikut:



Gambar 1 Metodologi penelitian

emas. *Dataset* yang digunakan untuk standar emas berasal dari penelitian sebelumnya Alfina dkk. (2017), Ibrohim & Budi (2018), dan buku pedoman kebencian Komnas HAM.

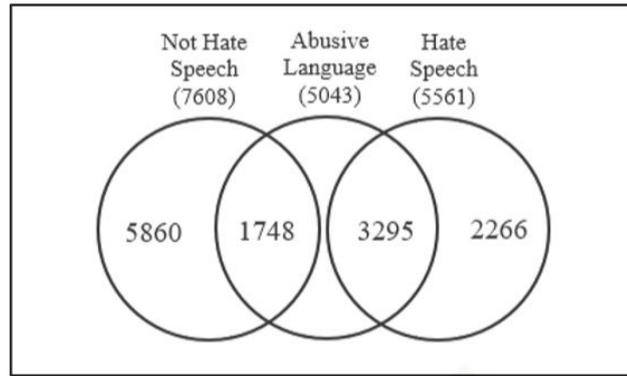
Setelah sistem anotasi dibuat dan diuji, proses selanjutnya adalah proses rekrutmen 30 anotator yang berasal dari berbagai jenis latar belakang dengan tujuan untuk mengurangi bias karena anotasi ucapan kebencian cukup subyektif [11]. Kriteria anotator adalah sebagai berikut: (a) berusia 20-30 tahun (karena sebagian besar pengguna *Twitter* di Indonesia berasal dari usia tersebut) (B) asli dalam bahasa Indonesia; (c) berpengalaman menggunakan *Twitter*; (d) bukan anggota partai / organisasi politik mana pun.

Pada fase anotasi pertama, mereka mengumpulkan 16.500 *tweet* dari proses *crawling* dan penelitian sebelumnya (Alfina dkk., 2017, 2018) dan T. T. A. Putri (2017) untuk dianotasi oleh 30 anotator tersebut. Setiap *tweet* dianotasi oleh 3 anotator dan label akhir diputuskan menggunakan teknik kesepakatan 100%. Dari fase ini, mereka mendapatkan 11.292 yang terdiri dari 6.187 *tweet hate speech* dan 5.105 *tweet hate speech* yang memiliki 100% perjanjian (*dataset* yang dapat diandalkan).

Pada fase anotasi kedua, mereka memberi anotasi 5.700 *tweet hate speech*. Mereka menggunakan 3 anotator terbaik dari fase anotasi pertama untuk menjelaskan terhadap *tweet target*, *category*, dan *level hate speech*. Label terakhir pada fase ini diputuskan menggunakan suara terbanyak. Karena mereka menggunakan 3 anotator, setiap label *tweet* harus memiliki perjanjian minimum dari 2 anotator. Jika tidak ada kesepakatan di antara para anotator dalam memberikan label, maka *tweet* tersebut dihapus. Dari hasil anotasi tahap kedua, ada 139 *tweet* yang dihapus karena tidak ada kesepakatan dalam *category hate speech* atau *level hate speech*. Oleh karena itu, mereka mendapatkan 5.561 data yang andal yang dapat digunakan untuk percobaan penelitian. Distribusi *hate speech* dan *abusive* dari proses anotasi dua fase ini dilihat pada Gambar 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.



Gambar 2 Distribusi dataset *tweet hate speech, non hate speech, dan abusive language*

Dari Gambar 2, dataset menghasilkan irisan sebanyak 13.169 *tweet* yang terdiri dari: 7.608 *tweet non hate speech* dan 5.561 *tweet hate speech*; 8.125 *tweet abusive* dan 5.043 *tweet abusive*; 3.295 *tweet hate speech* dan *abusive*; 5.860 *tweet non hate speech* dan *non abusive* . Jadi sebuah *tweet* dapat mengandung *hate speech* dan *abusive* atau tidak sama sekali atau salah satu dari dua tersebut.

Lalu dalam 5.561 *tweet hate speech*, 3.575 *tweet* target ke individu dan 1.986 *tweet* target ke grup. Beberapa *category hate speech* yang terdiri dari 793 *tweet* agama / kepercayaan, 566 *tweet* ras / etnis, 323 *tweet* fisik / kemampuan, 306 *tweet* gender / orientasi seksual, dan 3,740 *tweet* lainnya makian / fitnah. Sementara, untuk kelas *tweet level hate speech*, dataset *tweet hate speech* mereka terdiri dari 3.383 *tweet level weak speech*, 1.705 *tweet level moderate*, dan 473 *tweet level strong* [5].

3.2.2 Studi Pustaka

Tujuan tahap studi pustaka adalah untuk mengumpulkan pustaka yang berkaitan dengan penelitian ini untuk dijadikan sebagai referensi. Pustaka dapat berupa jurnal, buku, *e-book*, dan media *online* lainnya.

3.3 Analisa

Tahap ini merupakan tahap pengumpulan data yang akan dibutuhkan untuk melakukan analisa, perancangan, dan pengujian sistem pada penelitian ini. Berikut tahapan analisa dalam 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.

3.3.1 Persiapan Data

Setelah mendapatkan datasetnya, selanjutnya masuk tahap persiapan. Klasifikasi yang diambil dari dataset tersebut yaitu klasifikasi *hate speech* dan klasifikasi *abusive language*. Kemudian sub klasifikasi *hate speech* yang diambil yaitu klasifikasi *level* dengan kondisi jika *tweet* tersebut termasuk *hate speech*. Untuk memproses *dataset* awal menjadi dataset yang produktif, ini dilakukan dengan *cleaning* dan *splitting* data.

3.3.2 Feature Selection

Feature Selection bertujuan memeriksa kerja fitur dengan model yang dibuat dan meningkatkan fitur jika diperlukan. Contoh fiturnya adalah seberapa banyak tanda seru, seberapa banyak penggunaan huruf besar, dan lain-lain. Jenis-jenis *feature selection* yang digunakan dalam penelitian ini yaitu menggunakan kombinasi beberapa macam *text preprocessing* karena satu *text preprocessing* belum tentu cocok dengan segala kondisi dataset. Penggunaan *feature selection* di penelitian ini bertujuan untuk mengolah kata-kata dalam suatu *tweet* sebelum melakukan tahap klasifikasi.

Ada empat fitur yang digunakan pada penelitian ini, yaitu *case folding*, *removal punctuation*, *removal stopword*, dan *tokenizing*. Untuk proses tokenisasi, setiap token kata dibuat panjang minimal 2 dan 3 karakter. Suatu kombinasi *feature selection* bisa terdiri dari semua fitur dan bisa terdiri dari beberapa saja.

Berikut penjelasan empat fitur yang digunakan pada penelitian ini:

1. Case Folding

Case folding adalah proses mengubah huruf besar menjadi kecil. Berikut contoh *case folding* yang terdapat pada Tabel 2:

Tabel 3 Case folding

Sebelum	Sesudah
Oleh rezim skrg atas kendali Negara Cina Komunis yg mjd patron tunggal rezim anti islam (Jokowi)	oleh rezim skrg atas kendali negara cina komunis yg mjd patron tunggal rezim anti islam (jokowi)

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. *Removal Punctuation*

Removal punctuation adalah proses menghapus tanda baca atau simbol, seperti (;), (:), dan (?). Berikut contoh *removal punctuation* yang terdapat pada Tabel 3:

Tabel 4 *Removal punctuation*

Sebelum	Sesudah
Cina perusak bangsa!! Usir !! Stuju??	Cina perusak bangsa Usir Stuju

3. *Removal Stopword*

Removal Stopword menghapus kata-kata *stopword* yang memiliki informasi rendah dari sebuah teks, seperti ‘yang’, ‘dan’, ‘di’, ‘dari’. Berikut contoh *removal stopwords* yang terdapat pada Tabel 4:

Tabel 5 *Removal stopwords*

Sebelum	Sesudah
USER USER Kita semua tidak pernah akan mempercayai yg negara Kita disuatu hari nanti akan dijajah oleh kerana terdapat kes2pembubaran sistem2 beraja dinegara Kita kini untuk menjadikan sebagai negara republik komunis padahal negara2 penjaja	USER USER mempercayai yg negara disuatu dijajah kerana kes2pembubaran sistem2 beraja dinegara menjadikan negara republik komunis negara2 penjaja

4. Tokenisasi

Tokenisasi adalah memisahkan setiap kata yang ada pada sebuah dokumen. Setiap token dibuat panjang minimal 2 dan 3 karakter. Berikut contoh *removal punctuation* yang terdapat pada Tabel 5:

Hak Cipta Diindungi 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.

Tabel 6 Tokenisasi

Sebelum	Sesudah (>= 2 karakter)	Sesudah (>= 3 karakter)
USER USER Etnis cina memang ada yg ikut berjuang tapi lebih sangat banyak yg jadi penghianat jd antek belanda..	USER USER Etnis Cina Memang Ada Ikut Berjuang tapi Lebih Sangat Banyak Jadi Pengkhianat Antek belanda	USER USER Etnis Cina Memang Ikut Berjuang tapi Lebih Sangat Banyak Jadi Pengkhiant Antek belanda

3.3.3 Pembobotan Kata

Jenis perhitungan pembobotan kata yang digunakan adalah *Bag of Word*. Sebuah dokumen direpresentasikan sebagai token atau kata, mengabaikan tata bahasa dan bahkan urutan kata tetapi tetap menghitung frekuensi atau ada tidaknya kemunculan kata dalam suatu dokumen [38]. Sistem pembobotan kata *Bag of Word* adalah bernilai 1 jika token X ada di *tweet* Y dan bernilai 0 jika token X tidak ada di *tweet* Y. Jadi setiap token akan dilihat token tersebut ada atau tidak di suatu *tweet*.

3.3.4 Klasifikasi

Tahap selanjutnya dilakukan salah satu metode klasifikasi, yaitu *Naïve Bayes Classifier* (NBC). Hal yang pertama dilakukan dalam proses klasifikasi ini adalah *training* data sampel untuk menghasilkan suatu model klasifikasi dalam bentuk probabilitas. Setelah itu dilakukan *testing* data dengan menggunakan model klasifikasi yang sudah dilakukan di tahap *training* data.

Tahap klasifikasi metode *Naïve Bayes Classifier* adalah sebagai berikut:

Hak Cipta Diindungi 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. Pelatihan
 - a. Membentuk vocabulary data latih.
 - b. Menghitung bobot dari masing-masing kata dengan menggunakan Bag of Word.
 - c. Menghitung nilai probabilitas dari masing-masing kelas menggunakan persamaan.
 - d. Menghitung nilai probabilitas dari setiap kata terhadap masing-masing kelas dengan persamaan.
2. Pengklasifikasian
 - a. Menghitung probabilitas dokumen terhadap masing-masing kelas menggunakan persamaan.
 - b. Menentukan kelas dengan nilai probabilitas tertinggi menggunakan persamaan.

Klasifikasi dengan metode *Naïve Bayes Classifier* di penelitian ini dilakukan sebanyak beberapa kali. Jenis klasifikasi tersebut yaitu klasifikasi *hate speech*, klasifikasi *abusive language*, dan klasifikasi *level hate speech* jika *tweet* tersebut termasuk *hate speech*.

3.4 Perancangan

Tahap ini merancang model sesuai dengan hasil analisa yang sudah dilakukan pada proses sebelumnya agar dapat membuat model yang terstruktur dengan baik. Dalam tahap ini, sistem klasifikasi akan dijalankan dengan menggunakan *Google Colab* sebagai wadah untuk menerapkan kode *python*.

3.5 Implementasi

Implementasi adalah tahap pengoperasian aplikasi yang sudah dibangun sesuai dengan analisis dan perancangan yang telah dilakukan. Ruang lingkup implementasi pengembangan sistem ini akan dibangun dengan spesifikasi perangkat keras dan lunak sebagai berikut:

1. Perangkat Keras:
 - a. Prosesor : AMD *Ryzen 5* ~ 2.10 GHz
 - b. Memori : 6 GB


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.

- c. SSD : 500 GB
2. Perangkat Lunak:
 - a. Sistem Operasi : *Windows 10 Home* 64-bit
 - b. Bahasa Pemrograman : *Python*
 - c. *Text Editor* : *Google Colab*

3.6 Pengujian

Setelah dilakukan semua tahap, selanjutnya dilakukan tahap terakhir. Pengujian adalah tahap untuk menguji kinerja dari suatu mesin yang dibuat. Pada tahap ini, dilakukan perhitungan akurasi dengan menggunakan teknik *confusion matrix*. *Confusion Matrix* adalah suatu tabel yang menggambarkan kinerja akurasi dalam klasifikasi. *Confusion Matrix* pada penelitian ini melakukan perhitungan akurasi saja supaya bisa membandingkan hasil penelitian ini dengan penelitian [5] serta mengukur seberapa besar akurasi metode *Naïve Bayes Classifier*.

3.7 Kesimpulan dan Saran

Bagian kesimpulan merupakan tahap penentuan kesimpulan terhadap hasil pengujian yang telah dilakukan. Hal tersebut bertujuan untuk mengetahui apakah penerapan algoritma yang telah dilakukan menggunakan metode NBC berhasil dan mengetahui tingkat akurasinya. Pada bagian saran berisi kemungkinan pengembangan yang dapat dilakukan terhadap penelitian ini.

BAB 5

PENUTUP

5.1. Kesimpulan

Berdasarkan implementasi dan pengujian yang telah dilakukan, maka kesimpulan penelitian ini bahwa:

1. Metode *Naïve Bayes Classifier* pada klasifikasi *hate speech* dan *abusive language* pada *Twitter* bahasa Indonesia dapat diterapkan karena hasil akurasi penelitian ini termasuk tinggi.
2. Akurasi tertinggi pada klasifikasi *hate speech* yaitu sebesar 82,992% dengan kombinasi *feature selection* yang terdiri dari *case folding* dan *panjang* token minimal 2 karakter.
3. Akurasi tertinggi pada klasifikasi *abusive language* yaitu 86,484% dengan kombinasi *feature selection* yang terdiri dari *case folding*, *removal stopword*, dan *panjang* token minimal 3 karakter.
4. Akurasi tertinggi pada klasifikasi *level hate speech* yaitu 71,875% dengan kombinasi *feature selection* yang terdiri dari *case folding* dan *panjang* token minimal 2 karakter.

5.2. Saran

Berikut saran dari peneliti untuk penelitian selanjutnya:

1. Penelitian selanjutnya dapat menggunakan beberapa jenis metode klasifikasi teks lainnya dengan tujuan melihat perbandingan nilai akurasi. Contoh metodenya antara lain *Decision Tree*, *Random Forest Decision Tree*, dan *Support Vector Machine*.
2. Penelitian selanjutnya dapat mengkreasikan fitur pada *feature selection* berbeda dari peneliti.

DAFTAR PUSTAKA

- [1] M. Hudson, "What Is Social Media?," *the balance small business*, 2020. [Online]. Available: thebalancesmb.com/what-is-social-media-2890301. [Accessed: 09-Sep-2021].
- [2] A. S. Cahyono, "Pengaruh media sosial terhadap perubahan sosial masyarakat di Indonesia," *J. ilmu Sos. ilmu Polit. diterbitkan oleh Fak. Ilmu Sos. Polit. Univ. Tulungagung*, vol. 9, no. 1, pp. 140–157, 2016.
- [3] Komnas HAM, *Buku Saku Penanganan Ujaran Kebencian (Hate Speech)*. Jakarta: Komisi Nasional Hak Asasi Manusia, 2015.
- [4] Polri, "Surat Edaran Kapolri Mengenai Penanganan Ujaran Kebencian," *Kontras Website*. Jakarta, 2015.
- [5] M. O. Ibrohim and I. Budi, "Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter," pp. 46–57, 2019.
- [6] A. Rachmadany *et al.*, "Classification of Indonesian quote on Twitter using Naïve Bayes," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 288, no. 1, 2018.
- [7] Z. Waseem, T. Davidson, D. Warmesley, I. Weber, and U. Kingdom, "Understanding Abuse: A Typology of Abusive Language Detection Subtasks," 2017.
- [8] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting Offensive Language in Social Media to Protect Adolescents," *2012 Int. Conf. Privacy, Secur. Risk Trust 2012 Int. Confernece Soc. Comput.*, pp. 71–80, 2012.
- [9] J. Han, S. Wu, and X. Liu, "Identifying and Categorizing Offensive Language in Social Media," *SemEval-2019*, pp. 652–656, 2019.
- [10] M. O. Ibrohim and I. Budi, "A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media," *Procedia Comput. Sci.*, vol. 135, pp. 222–229, 2018.
- [11] I. Alfina, R. Mulia, M. I. Fanany, and Y. Ekananta, "Hate Speech Detection in the Indonesian Language : A Dataset and Preliminary Study," *2017 Int. Conf. Adv. Comput. Sci. Inf. Syst.*, no. October, pp. 233–238, 2017.
- [12] A. Fitri, "Tingkat Akurasi Metode K-NN dan NBC Pada Klasifikasi Ujaran



Hak Cipta Diindungi Undang-Undang

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

Kebencian Terhadap Agama Islam,” UIN Suska Riau, 2019.

- [13] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a Social Network or a News Media?,” *19th Int. Conf. World Wide Web*, pp. 1–7, 2013.
- [14] K. Gligorić, A. Anderson, and R. West, “How constraints affect content: The case of Twitter’s switch from 140 to 280 characters,” *12th Int. AAAI Conf. Web Soc. Media, ICWSM 2018*, no. Icws, pp. 596–599, 2018.
- [15] H. Yenala, A. Jhanwar, M. K. Chinnakotla, and J. Goyal, “Deep learning for detecting inappropriate content in text,” *Int. J. Data Sci. Anal.*, 2017.
- [16] J. Pustejovsky and A. Stubbs, *Natural Language Annotation For Machine Learning*, I. Sebastopol, CA: O’Reilly Media, 2012.
- [17] S. Mujilawati, “Pre-Processing Text Mining Pada Data Twitter,” *Semin. Nas. Teknol. Inf. Dan Komun.*, 2016.
- [18] W. Budiharto, *Pemrograman Python untuk Ilmu Komputer & Teknik*. Yogyakarta: Andi, 2018.
- [19] S. George K and S. Joseph, “Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature,” *IOSR J. Comput. Eng.*, vol. 16, no. 1, pp. 34–38, 2014.
- [20] S. V. Gaikwad and A. Chaugule, “Text Mining Methods and Techniques,” *Int. J. Comput. Appl.*, vol. 85, 2014.
- [21] L. Rokach, B. Chizi, and O. Maimon, “Feature selection by combining multiple methods,” *Adv. Web Intell. Data Min.*, pp. 295–304, 2006.
- [22] T. Wahyono, *Fundamental Of Python For Machine Learning: Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan*, I. Yogyakarta: Gaya Media, 2018.
- [23] N. Busiarli, L. A. Aditya, and A. Y. Andika, “Penerapan Algoritma Naïve Bayes & Natural Language Processing Untuk Mengklasifikasi Jenis Berita,” *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 6–7, 2016.
- [24] B. Santosa, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, 1st ed. Yogyakarta: Graha Ilmu, 2007.
- [25] N. M. S. Hadna, P. I. Santosa, and W. W. Winarno, “Studi Literatur

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.

Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di Twitter,” 2016.

- [26] D. H. Kalokasari, D. I. M. Shofi, and A. H. Setyaningrum, “IMPLEMENTASI ALGORITMA MULTINOMIAL NAIVE BAYES CLASSIFIER PADA SISTEM KLASIFIKASI SURAT KELUAR,” vol. Vol.10 No., pp. 109–118, 2017.
- [27] I. H. Witten, E. Frank, and M. A. Hall, *Practical Machine Learning Tools and Techniques*, Third Edit. USA: Elsevier, 2011.
- [28] J. Wisudawati, Adiwijaya, and S. Al Faraby, “Klasifikasi Sentimen pada Movie Review dengan Metode Multinomial Naïve Bayes,” vol. .4, No.2, p. 2978, 2017.
- [29] F. Del Vigna, A. Cimino, F. Dell’Orletta, M. Petrocchi, and M. Tesconi, “Hate Me, Hate Me Not: Hate Speech Detection on Facebook,” *CEUR Workshop Proc.*, vol. 1816, pp. 86–95, 2017.
- [30] A. Gaydhani, V. Doma, S. Kendre, and L. Bhagwat, “Detecting Hate Speech and Offensive Language on Twitter using Machine Learning : An N-gram and TFIDF based Approach,” *IEEE Int. Adv. Comput. Conf. 2018*, 2018.
- [31] K. K. Kiilu, G. Okeyo, R. Rimiru, and K. Ogada, “Using Naïve Bayes Algorithm in Detection of Hate Tweets,” *Int. J. Sci. Res. Publ.*, vol. 8, no. 3, pp. 99–107, 2018.
- [32] I. F. Rozi, E. N. Hamdana, I. Alfahmi, and M. Balya, “Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus Samsat Kota Malang),” *J. Inform. Polinema*, vol. 4, no. 2, p. 149, 2018.
- [33] N. M. G. D. Purnamasari, M. A. Fauzi, Indriarti, and L. S. Dewi, “Identifikasi Tweet Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine (SVM) dan Information Gain (IG) sebagai Seleksi Fitur,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5326–5332, 2018.
- [34] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis

Hak Cipta Diindungi 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.

Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018.

- [35] M. Chalida, “Analisis Sentimen Ujaran Kebencian Pemilihan Presiden 2019 Menggunakan Algoritma Naïve Bayes (Studi Kasus: Tweet #Pilpres2019 Di Kota Jakarta, Bandung, Semarang, Surabaya Dan Yogyakarta),” *Jnanaloka (Jurnal Open Access Yayasan Lentera Dua Indones.*, no. 2001, pp. 5–10, 2019.
- [36] I. Alfina, S. H. Pratiwi, I. Budi, R. Mulia, and Y. Ekanata, “Detecting hate speech against religion in the indonesian language,” *J. Telecommun. Electron. Comput. Eng.*, 2018.
- [37] T. T. A. Putri, “Analisis dan deteksi hate speech pada sosial twitter berbahasa indonesia,” *Master’s thesis, Fac. Comput. Sci. Univ. Indones.*, 2018.
- [38] M. McTear, Z. Callejas, and D. Griol, *The Conversational Interface - Talking to Smart Devices*. 2016.

LAMPIRAN A

Berikut ini merupakan dataset beserta label yang digunakan pada penelitian ini:

Keterangan:

HS=*Hate Speech* AB=*Abusive Language* W=*Weak* M=*Moderate* S=*Strong*

Indeks	Tweet	HS	AB	Level HS (HS = 1)		
				W	M	S
0	disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego !!!'	1	1	1	0	0
1	RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga'	0	1	0	0	0
2	41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu meninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakakku lebih memilih jadi Kristen. Ketika aku anak ter	0	0	0	0	0
3	USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIT TAPI DILIAT DARI MANA ITU AKU'	0	0	0	0	0
4	USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah'	1	1	0	1	0
5	USER Ya bani taplak dkk \xf0\x9f\x98\x84\xf0\x9f\x98\x84\xf0\x9f\x98\x84'	1	1	0	1	0
6	deklarasi pilkada 2018 aman dan anti hoax warga dukuh sari jabon	0	0	0	0	0
7	Gue baru aja kelar re-watch Aldnoah Zero!!! paling kampret emang endingnya! 2 karakter utama cowonya kena friendzone bray! XD URL	0	1	0	0	0
8	Nah admin belanja satu lagi port terbaik nak makan Ais Kepala Milo, Ais Kepala Horlicks atau Cendol Topping kaw kaw. ð??; ; Docket mano tu ? Gerai Rojak	0	0	0	0	0

© Hak cipta milik UIN Suska Riau		Mertuaku - Taipan 2 (depan TWINS BABY & ROMANTIKA / Bank Islam Senawang) ð???						
9		USER Enak lg klo smbil ngewe'	0	1	0	0	0	
10		Setidaknya gw punya jari tengah buat lu, sebelum gw ukur nyali sama bacot lu \xf0\x9f\x98\x8f'	1	1	1	0	0	
11		USER USER USER USER BANCI KALENG MALU GA BISA JAWAB PERTANYAAN KAMI DARI 2 HARI LALU.... NYUNGSEP KOE USER URL	1	1	1	0	0	
12		Kalo belajar ekonomi mestinya jago memprivatisasi hati orang. Duh.. ironi USER	0	0	0	0	0	
13		Aktor huruhara 98 Prabowo S ingin lengserkan pemerintahan Jokowi.... Nyata	1	0	1	0	0	
14		USER Bu guru enakan jadi jablay atau guru esde sih.\nKayaknya menikmati jadi pecun ini guru.'	1	1	1	0	0	
15		USER USER USER USER USER USER Lawan bicara gw gak intelek kyk loe, yg otak gak punya. Ttg kencing onta, gw mengakui hadis nabi dan itu sahih. Kafir kyk loe pasti menolak, makanya loe ahlunnar.'	1	1	1	0	0	
16		Belakangan ini kok fikiran ampas banget ya'	0	0	0	0	0	
17		Ari sarua beki mah reph monyet\xf0\x9f\x98\x86\x0\x9f\x98\x86'	0	1	0	0	0	
18		Jadi cowo itu harus Gantle kalo ga Gantle itu namanya BANCI !'	1	1	0	1	0	
19		USER Slga mnr bom \xf0\x9f\x98\x82'	0	0	0	0	0	
20		Asw ya tapi gua jarang ngambek, tacut wkkwkwkwkw gua kan bucin.'	0	1	0	0	0	
21		USER kalo kamu noob pasti peluang disakitin nya lebih gede si'	0	0	0	0	0	
22		USER Joko Widodo dinilai sebagai presiden terlemah dalam sejarah Indonesia. Hal ini terjadi bukan saja karena Jokowi tidak memiliki modal dukungan politik yang cukup, melainkan juga karena ketidakmampuan pemerintahannya menghadapi situasi ekonomi	0	0	0	0	0	

Hak Cipta Diindungi 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 milik UIN Suska Riau	23	PELAJAR SMA KEC BILAH HILIR DEKLARASI ANTI HOAX & PILKADA 2018 DAMAI	0	0	0	0	0
	24	Bandara Udara Internasional Kertajati dibangun oleh Gubernur Ahmad Heryawan tanpa gunakan APBN pemerintah pusat. Insya Allah siap beroperasi bulan depan;	0	0	0	0	0
	25	Siapapun gubernur dan presidennya, rakyatnya, ya kita ² juga...	0	0	0	0	0
	26	Ini si USER kerjanya delay mulu! Setan!	1	1	0	1	0
	27	menurutku pintu sorga ada yaitu pintu sorga yang asli dan pintu hatimu modusbanget	0	0	0	0	0
	28	RT USER USER USER PKI hanya muncul jika jelang pemilihan kepala daerah dan pemilihan presiden krn digoreng sampai gosong.	1	0	0	1	
	29	USER USER Itu mah sdh nenek-nenek sy heran sama cebong biasay bohong terus'	1	1	1	0	0
	30	USER USER USER USER USER USER USER USER USER Islam Nusantara produk JIL, dipasarkan dg gencar oleh antek JIL.\n\nSementara JIL anti Arab, tapi pro Aseng!\n\nKata (alm) KH. Hasyim Muzadi,	1	0	0	1	0
State Islamic University of Sultan Syarif Kasim Riau	31	Dari habis sahur sampe jam 10. Sibayik udah nete 4x. Skg rasanya lemas tak terkira \xf0\x9f\xa4\xa2'	0	0	0	0	0
	32	USER Gak kak emak mah gak demen ama yang sipit wkwk'	0	0	0	0	0
	33	RT USER: Kelakuan homok jaman now, ngentot aja sambil live di Blued #gayvid #gvid #lokalhangat URL	0	1	0	0	0
	34	USER Ga kak, gua bukan orang jawa maaf yak :(; ; gua orang palembang soalnya, kakak aja deh yg belajar aksara jawa. Atau kalak mau belajar budaya palembang?	0	0	0	0	0
	35	Pegawai Guardian training lagi kek Mentang2 gue kek gembel cuman kaos sama Flat shoes nanya dijutekin bangsad emang males belinya juga kabur gua ke matahari'	1	1	0	1	0

Hak Cipta Diindungi 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.

36	\xe2\x80\x9dPara Bajingan yang telah melukai Lady harus menerima ganjaran seratus kali lipat!!\xe2\x80\x9d(sanji) ##HBD17thPelautJKT48'	0	1	0	0	0
37	ajakan menolak berita hoax dan sukseskan pilkada di wilayah kota keidri:	0	0	0	0	0
38	USER USER Presiden RI Jokowi luarBiasa kerja nya tulus IHLAs untuk #RakyatB3rsamaJKW	0	0	0	0	0
39	Edan emang goollnya fabiano!!'	0	1	0	0	0
40	GAPERNAH MENDALAMI AL-QURAN YA BANG??? PANTESAN MULUTNYA KAYA ORANG KAFIR BEJAT HINA farhatabbaslaw'	1	1	1	0	0
41	Sangat terlihat beda sekali Kewibawaannya ; yg sangat berwibawa cuma gubernur Indonesia USER ; Massya Allah...	0	0	0	0	0
42	Anjir baru kali ini lihat orang bego mau debat pilkada. Kerjanya nyerang karakter. Shame on you mpok silvy and agus. #DebatFinalPilkadaJKT	1	1	0	1	0
43	#GubernurZamanNow #GusIpulPuti2 #GanjarYasin1 #DjarotSihar2 #HasanAnton2 #KosterAce1 #NurdinSudirman3 #KarolinGidot2 RT USER: "Keanekaragaman budaya, suku, agama, itulah sesungguhnya kekayaan kita,â?? tutur Djarot di hadapan peserta Rakercabsuâ?'	0	0	0	0	0
44	USER USER Apa iya Islam itu ga boleh dilokalisasi? Lha Kristen aja mas ada Nusantaranya mas, liat aja di Gereja di bandung ada yg pake gamelan dan ada yg khotbah pake bhs Sunda/Jawa, ada HKBP juga kurang Nusantara apa?'	0	0	0	0	0
45	Ciptakan Situasi Kamtibmas Menjelang Pilkada 2018, Kapolsek Konang Silaturahmi ke Tokoh Masyarakat	0	0	0	0	0
46	USER USER positif kok.\nse-positif respon masyarakat terhadap\n-CANGKUL impor\n- GARAM impor\n-BERAS impor\n- JAGUNG impor\n-TENAGA KERJA impor\n- ESEMKA impor\n\nURL	0	0	0	0	0

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 milik UIN Suska Riau	47	USER USER USER USER Kata Sembahyang mengadopsi tradisi lokal pada masa kerjaan Hindu.\n\nIstilah\\ sembahyang\\ pun lahir dari tradisi	0	0	0	0	0
	48	KAU PENDIKAN TINGGI TAPI BODOH...GK ADA LARANG BAWA AGAMA..BIAR ORANG ISLAM YG DI BAYAR AHOK SADAR GK JADI MILIH AHOK GLA	1	1	1	0	0
State Islamic University of Sultan Syarif Kasim Riau	49	USER USER BTP the best ‘,,abaikan kaum kampret otak micin ‘,,’	1	1	0	1	0
	50	USER USER , si otak tempurung mah ga usah d ladenin. Smkn d lawan semakin Bloon dan Dunggu. Dan kalo ga gitu dia ga makan. Hehe'	1	1	1	0	0
State Islamic University of Sultan Syarif Kasim Riau	51	USER Awaasss... Jgn sampe beritanya sampe ke akhi gerung! Ntar susah nyari kosakatany. klu bcr aja micingin mata, bibir monyong kesamping, suara keluar ky lg nhn kentut.. "politik kotor skl jmn ini!". Klu tahu brt ini.. Plng Sambil ky nhn ngeden.."dungu!". \ð\Ÿ\˜\‹	1	1	1	0	0
	52	pinokio umur 3 taonan tuh. #2019GantiPresiden	1	0	1	0	0
State Islamic University of Sultan Syarif Kasim Riau	53	USER Cerita silat lae \ð\Ÿ\˜\€ semacam Kho Ping Hoo yang dari Indonesia, tapi Liang Ie Shen penulis dari China'	0	0	0	0	0
	54	RT USER benci sekali dgn Umat Islam	1	0	0	1	0
State Islamic University of Sultan Syarif Kasim Riau	55	USER USER USER "Provokasi mayat, politisasi agama" penyebab kekalahan pilkada DKI, beginilah cara cebong mendeskripsikan kekalahan junjungannya.; FYI ahog blm pernah ikut pemilihan apapun (kec jd wakil) dan dia bukanlah etnis mayoritas	1	1	1		
	56	USER dari pada gua nete'	0	1	0	0	0
State Islamic University of Sultan Syarif Kasim Riau	57	USER Lol bani lowo'	1	1		1	
	58	Takabur bisa lahir dari siapa saja. Penguasa, rakyat, ulama maupun cendikia. Bahkan iblis mahluq yg alim & telah beribadah beratus2 th dilaknat krn hasud takabburnya. Apalagi	0	0	0	0	0

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

		hanya seorang ahli ilmu filsafat yg tdk failosof; ; Mari kita demo diri sendi						
© Hak cipta milik UIN Suska Riau	59	A orang yang suka ke konser-konser, sedikit boros.	0	0	0	0	0	0
	60	[SAMPAI KAPAN MAKAN RACUN ?]; _; Banyak ulama yang mengatakan hukum rokok mubah dan ada juga?!	0	0	0	0	0	0
	61	Susilo Bambang Yudhoyono Presiden RI ke 6 meminta Khofifah Emil untuk dekat dengan masyarakat . No 1 Bisa ! BaGusS #KhofifahEmilNomer1	0	0	0	0	0	0
	62	USER udah biasa kali. Gue kpoper tapi kalo ngomov sama temen2 gue, idol gua banci homo'	1	1	1	0	0	0
	63	Saya suka video USER URL Awal Kisah Perseteruan Haji Bolot & Malih Tongtong'	0	0	0	0	0	0
	64	Budaya kafir bersungguh, klau hal2 islam tidak diendahkan...	0	0	0	0	0	0
	65	USER USER USER USER \xf0\x9f\x8e\xb6 la la la...hm hmm \xf0\x9f\x8e\xa7 "Semua diam ,semua bisu"\n"Kita coba tanya sama rumput yg bergoyang" \n\xe2\x99\xab\xe2\x99\xab\xe2\x99\xab\xe2\x99\xaa\xe2\x99\xaa\xe2\x99\xaa'	0	0	0	0	0	0
State Islamic University of Sultan Syarif Kasim Riau	66	Apabila setan membisikkan was-was kepadamu, mintalah perlindungan kepada Allah. (QS. al-A\xe2\x80\x99raf: 200 dan Fushilat: 36).'	0	0	0	0	0	0
	67	USER USER USER USER USER USER USER USER Klw ente atheis..ngapain sibuk dgn agama orang...	0	0	0	0	0	0
	68	Lengserkan Jokowi Bangsat!!	1	1	0	0	0	1
	69	USER Selamat hari kartini ; ; #2019GANTIPRESIDEN; #2019PRESIDENBARU	0	0	0	0	0	0
	70	RT USER: Bajingan Homo!!!\nHati2 terhadap anak2 kita terutama di kamar mandi laki2, jgn biarkan sendiri! URL	1	1	1	0	0	0
	71	USER Iya cocot'	1	1	1	0	0	0
	72	USER karena budaya indonesia.	0	0	0	0	0	0
	73	USER Yang ngajarin solat'a ajah itu orang aseng dia hanya ingin memecah	0	0	0	0	0	0

- Hak Cipta Dilindungi Undang-Undang
1. Diarangi 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. Diarangi mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		belah islam"					
74	USER x Lo ngomong aja kaga bener bangke'	1	1	1	0	0	
75	USER USER Kita semua tidak pernah akan mempercayai yg negara Kita disuatu hari nanti akan dijajah oleh kerana terdapat kes2pembubaran sistem2 beraja dinegara Kita kini untuk menjadikan sebagai negara republik komunis padahal negara2 penjaja	0	0	0	0	0	
76	tapi masih menikmati kebijakan yang dulu diusulkan PKI.. lha yo opo...	1	0	0	1	0	
77	Kalo seumpama orang cina di usir paksa dari indonesia gimana	1	0	0	1	0	
78	USER bunda ngentot yuk'	0	1	0	0	0	
79	RT USER: china babi... china maling... \n#UsirChina https://t.co/9y4HIWHJp2	1	1	0	0	1	
80	Bung Karno anti asing dan beliau dari Jawa, coba deh adain konsensus lebih banyak mana Jawa yang anti asing atau pro asing karena kebo*ohannya. \n\nKalau kami keliru kami tutup akun'	0	0	0	0	0	
81	PROYEK KORUPSI REZIM SBY URL	1	0	1	0	0	
82	USER Wajar saja, lihat saja tingkah bbrp petinggi negeri ini yg suka mencampur bhs indonesia dgn bhs asing setiap kali bicara'	0	0	0	0	0	
83	AB: pribadi yg senang berkompetisi dalam hal yang baik dan positif	0	0	0	0	0	
84	USER USER Terus menurut lo kitab suci agAma apa yg di bilang om USER ; ; #Mendadakkalap	1	0	1	0	0	
85	USER Suka nutup mata. Kaga ditutup pun udah terlihat tertutup gara2 sipit /?'	0	0	0	0	0	
86	USER Sy sbg muslim tdk merasa pak USER menistakan kitab suci Alqur'an, krn dlm Alqur'an sndri ada bbrp suku kata yg smpai dtik ini blm diketahui mkna faktualnya olh umat islam, tp kami yakini kata tsb ada makna yg tersirat (fiksi) yg blm	0	0	0	0	0	
87	USER eh kasar kau kampung'	1	1	1	0	0	

Hak Cipta Diindungi 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.

88	BENTAR LAGI GUA RILIS ALBUM THE ONTA'S WKWKWKWKWK"	0	0	0	0	0
89	USER Ada cebong bang dia merasa paling benar'	1	1	1	0	0
90	Najis banget gw ngetwit apa barusan wey'	0	1	0	0	0
91	Kamu transgender atau gmn anjing :(URL	1	1	1		
92	USER awas bintitan lu ngintipin titit kucing'	0	1	0	0	0
93	USER USER Bawel, baper, bolot, baik, bodor , gaje wkwk apa ya udah itu aja deh'	0	1	0	0	0
94	USER USER USER USER Gue saranin gk perlu bnyk bacot maling ayat, langsung to the point aja lu jadi orang... Yg tdk sepaham lu teriakin aja kafir sesat, liberal, pemuja dukun, bla blaaa.. pasti hatimu puas deh ..'	1	1	1	0	0
95	Jgn salah target bkn ahok... agenda utama lengserkan Jokowi ..	1	0	0	0	1
96	USER USER masa baru kena hestek aja udah kepanasan ??? Kepo banget ; ; #2019GantiPresiden	1	0	1	0	0
97	Maju lu sini duit tapi kau goyang kaki tak buat pape , mana la duit nak maju kat kau bodo URL	0	1	0	0	0
98	USER HKBP berdiri itu sejak pecahnya kristen khatolik pada zaman penjajahan belanda,dl tdk ada yg namanya kristen protestan yg ada cm kristen khatolik.\nTerjadilah perpecahan dan berdirilah HKBP yg prtama kali di tarutung.\n\nNah skrg anda sy k	0	0	0	0	0
99	Ha .. ha .. Banyak baca biar pintar ... JEPANG, AMERIKA, CHINA... Utang sama siapa ... Anda tau enggak ..?	0	0	0	0	0
...
13.168	USER Mana situ ngasih("): itu cuma foto ya kutil onta'	1	1	1	0	0

Hak Cipta Diindungi 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 B

Berikut ini merupakan hasil prediksi label *tweet* serta perbandingan sama atau tidak dengan label *tweet* yang ada di dataset pada penelitian ini. Urutan indeks telah teracak:

Keterangan:

HS Ds=*Hate Speech* di Dataset HS Pr=*Hate Speech* di Prediksi

AB Ds=*Abusive Language* di Dataset AB Pr=*Abusive Language* di Prediksi

Indeks	<i>Tweet</i>	HS Ds	HS Pr	AB Ds	AB Pr
0	USER USER USER USER USER Nggak sama lah. Bank asing gak bisa atur2 perusahaan. Pemerintah bisa atur perusahaan sbg pemilik saham. Utang bisa dibayar dr keuntungan freeport.'	0	0	0	0
1	Cara paling rasional dan realistis dalam menilai kepemimpinan Presiden Jokowi adalah: sejauh mana dia telah membawa perbaikan dalam berbagai aspek pembangunan dan kesejahteraan masyarakat dibandingkan dengan sebelumnya. Jika cukup banyak maka sangat layak	0	0	0	0
2	Mungkin kamu yg smakin bodoh? Pancasila, UUD'45, TAP MPR no.25 Sudah Final PKI terlarang! kecebong gak usah rayu-rayu Presiden utk buat Perpu	1	1	1	0
3	dan gus dur bersikap sebagai seorang pemimpin muslim yang ramah idealis namun mencintai perbedaan pluralisme keren	0	0	0	0
4	USER USER ih masak di samain smaa sampah \xf0\x9f\x98\xa3\xf0\x9f\x98\xa3\xf0\x9f\x98\xa3'	0	1	0	1
5	Selamat siang masyarakat Banten~°??; Sambut Pilkada Serentak 2018 di Banten dengan penuh kebahagiaan serta aman dan kondusif~°??	0	0	0	0
6	USER USER Pemandangan yg gak asing lagi \nApa lagi yg jualan'	0	0	0	0
7	RT USER: Ironinya banyak bisnis dg mega triliunan profit namun dulu pendirinya sama sekali nggak punya mimpi kesana. \n\nContohny\xe2\xe6'	0	0	0	0



© Hak cipta milik UIN Suska Riau	8	USER USER USER USER USER USER USER USER Ulah cebong kafir keparat itu'	1	1	1	1
	9	RT USER: #RETWEET \nDigoyang sampe basah.. \xf0\x9f\x98\x9d \n\n#horny #memek #colmek #colokmemek #jembut #memekbecek #ngentot \n#videongentot #Vi\xe2\x80\xa6'	0	0	1	1
	10	USER kirain banci:('	0	0	1	1
State Islami: University of Sultan Syarif Kasim Riau	11	USER USER Ini Kehebatan Dlm Merusak Bangsa USER. URL	1	1		
	12	Ustadz PKS Yudi Widiana nilep duit pk bahasa agama Juz & Liqo Felix Kwetiauw & alumni 212 pd mingkem & pura2 budek..giliran Bu Sukma bc puisi pd teriak2 kesurupan'	1	1	1	1
	13	USER keluarga halilintar hmm masa km gatau. udik'	1	0	1	1
Hak Cipta Diindungi Undang-Undang	14	Makanya, rakyat juga sudah gak sabar pengen lengserkan, tapi para tokoh negri pada takut sama santet jokowi. Dicari tokoh nasional yg berani	1	1	0	0
	15	USER janji prabowo apa ? apa janji2nya nanti bisa terealisasi semua selama 1 periode seandainya dia menang pilpres ? janji gombal ? gimana dengan UU MD3 ? siapa yg berangus kebebasan berpendapat ? melihat cermin orang lain tapi tidak mau beli cer	1	0	0	0
	16	Bisa bertitel Kaji dri dulu sampai sekarang memang tdk Murah(Muahal). Blm lagi ngantri keberangkatanya. Bsa 5-8an thn lebih, bru bisa brangkt. Kaji+, pun, yg katanya bs cpt, juga hrs nugu. Alternatif cept kaji adlh jadi TKI di middle east. Salut dg mreka.	0	0	0	0
Hak Cipta Diindungi Undang-Undang	17	USER Siklus couple-an\n\n- beda rasa\n- rasa asing\n- break\n- break up\n\nDah gitu aja terus...\n\nKalau merasa tak sama seperti dulu coba saling terbuka. Itu biasa nya krn ada nya kurang komunikasi atau ketidak pekaan di satu pihak \xf0\x9f\x98\x85	0	0	0	0
	18	USER USER Bapak saya kok gembel bengad yaa'	0	0	1	1
	19	Kalian perankan Tanpa ADA Cacat'	0	1	0	1
Hak Cipta Diindungi Undang-Undang	20	USER Pamor RK jadi seakan tak ada artinya jika parpol pengusungnya pendukung penista agama...	1	1	0	0

1. Diarangi 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 milik UIN Suska Riau	21	Presiden Minta Proyek Strategis Nasional 2017 Segera Dirampungkan -	0	0	0	0
	22	USER Mata krna sipit dibilang kayak Korea bahkan pas SMA dibilang mirip sama salah satu member SNSD \xf0\x9f\xa4\xa3'	0	0	0	0
	23	USER duuuuhhhh,, Jaya Suprana yang kristen saja mau berguru dengan beliau, lah yg ngaku muslim kok ora	0	0	0	0
	24	tempat paporit kerjain tugas adlh bux, krn kita sering dpt kupon gratisan dr temen \xf0\x9f\x98\x85 so lucky lah.\nngerjain tugas plus curhat, pokonya baru pulang pas uda ditelpon \xf0\x9f\x98\x82\nkalo nda makan mi ayam rica biasanya kita jajan mi goreng	0	0	0	0
	25	USER USER BTP the best ,,,'abaikan kaum kampret otak micin ,,'	1	1	1	1
	26	USER USER Ya..ttp aj waspada..!!Antek komunis itu pintr kamuflose..	1	1	0	0
	27	USER Bajingan la bukan banjangan'	0	1	1	1
	28	RT USER: Jilat kontol sampe tegang baru masukin memek'	0	0	1	1
	29	USER ASU ASU ASUUUU SI KUYAAAAAAAAAAAAAAAAA'	1	1	1	1
	30	USER Dalam agama Hindu, karma itu artinya perbuatan dalam bahasa sansekerta. Kamu berbuat apapun itu, itu adalah karma yang kamu lakukan, dan harus kamu jalani.'	0	0	0	0
State Islamic University of Sultan Syarif Kasim Riau	31	USER Curang apah???? Titit Sele masi sakit nih :('	0	0	1	1
	32	Yai Ma'ruf Amin salah seorang Ulama yg hadir maulid Majelis Rasulullah sejak pertama kali diadakan di Monas...; .; .; AlHabib Munzir AlMusawa menjuluki Yai Ma'ruf sbg Gerahamnya Ahlusunnah Wal Jama'ah	0	0	0	0
Syarif Kasim Riau	33	USER lost respect untuk kesekian kalinya buat agency yg lo banggain, buat artis yang lo banggain. \nkelakuan tai kucing di kasih nyawa sih kea gini. \n\nlo di tampol duit sama lee sooman mau? di tampol lamborgini sama GD mau? di tampol sama material	1	0	1	1
	34	RT USER: USER Ahok ini sosok yg komplit, selain hatinya busuk, ucapannya menjijikkan, perbuatannya gak kalah brengsek,	1	1	1	1

Hak Cipta Diindungi Undang-Undang

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

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

		menista\xe2\x80\xa6'				
Hak cipta milik UIN Suska Riau	35	tobat yuk sebelum terlambat. biar gerombolan JIL aja yg kita doakan konsisten bela kekafiran hingga akhir hayat URL	1	1	0	0
	36	Kalau kerasnya nasionalis religius dan Islam melawan politik kawan penista agama, kawan simpatisan PKI, kawan yg dikutuk Palestina, kawan pengubur agama kalau itu BETUL, bahkan HARUS dgn spirit agama. Apa jgn2 politik yg dikafirkan itu benar2 mrk	0	1	0	0
	37	USER Yang ada di otaknya tentang kesetaraan gender cuma telanjang. Dasar mesum cacad.'	1	1	1	1
	38	Tapi si \xf0\x9f\x91\xa9\xf0\x9f\x8f\xbb pas dperhatikan warna badannya kontras sekali sama warna mukanya, muka putih badan coklat. \nI was like.... \xf0\x9f\x99\x82 lu kata mata gue picek apa, tidak bisa bedakan yang pake krim pemutih dan yang tidak pa	0	0	0	1
	39	Dinamika politik hari ini adalah refleksi upaya koalisi SBY - Jenderal Merah dan TNI merebut Jokowi	1	1	0	0
	40	USER USER HTI jg organisasi dgn paham terlarang di Bumi Pancasila jd hrs Segera di Bubarkan	1	1	0	0
	41	USER Pak USER saya tambahkan lagi bukti Muhammad Nabi Gadungan, masyarakat menunggu pengumuman Anda	1	1	1	
State Islamic University of Sulhan Syarif Kasim Riau	42	..Puri Tri Agung adalah tempat ibadah/sembah bagi penganut Agama Budha, Khongfutze dan Laotze.Ä-Ä;Ä½	0	0	0	0
	43	USER "Namaku Baik Yoon\xe2\x80\x94gi ..."\n\nUcapan terputus seiring dia mengerjap mata yang terasa mengantuk sedikit demi sedikit. Gigit bibirnya ragu karena dia ingin menggapai gelas sekali lagi untuk meminum sebagaimana wine yang ditinggal di m	0	0	0	0
	44	Saya suka video USER URL A Way Out - Kolaborasi 2 Idiot Kabur Dari Penjara'	0	0	0	1
	45	USER Kok ganti namanya jd hatstag Ä°??Ä°??Ä°?? pengen bgt ganti presiden ganteng yaÄ°??	0	0		1
	46	Paling nanti komentarnya "jangan transaksi pake dollar".; #2019GantiPresiden ; #2019PresidenBaru	1	1	0	0
	47	Tahu goreng & kopi, Prabowo &	0	1	0	0

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 milik UIN Suska Riau		Jokowi di Pilpres 2019. ; Siapa paling pas jadi Presiden?...pas kopinya...pas Jokowi. ; ; #T3tapJokowi ; #T3tapJokowi2019 ; #JokowiPresidenkuKerjaNyata ; #MenujuIndonesiaMaju ; #JokowiMembangunIndonesia; #UangKita #J				
48	USER USER USER USER USER	Maaf nih....emang ada katolik nusantara?'	0	0	0	0
49	USER USER	Betul.... beban hidup meningkat dirasakan oleh bani sumbu pendek yg biasa treak2 kopar kapir... mrk kekurangan job demo krn udh ga laku....'	1	1	1	1
50	USER spt ahok dulu .. nasdem disusul dg cs nya .. yg mendukung paslon penista agama dan tukang caci maki tai tai ..		1	1	1	1
51	RT USER: USER USER	Dia lupa ini twitter, bukan instagram yang kalo foto pasti ada yg komen 'ih cantik' 'cantikan kamu lah\xe2\x80\xa6"	0	0	0	1
52	USER USER	Waakaakaak.... Ni orang pasti pendukung Yahudi... Jangan jangan dipikiran dia yg ada negara Israel... waakaakaak'	1	1	0	0
53	Rezim anti Islam pro komunis.. Insyallah kita akan saksikan kehancurannya.		1	1	0	0
54	RT USER	Tdk cocok jd calon presiden nya jkw .cocok untuk prabowo	1	1	0	0
55	Saya itu gak mikir copras capres...skrng ngurus hutang, impor, TDL Naik, BBM naik, dollar tinggi, pertumbuhan ekonomi cuma 5,2%, premium langka, 10 jt lpngan pekerjaan gak nyampe2, daya beli msyrkt melorot....itu aja dah pusing!; #2019GantiPresiden #2019Ga		1	1	0	0
56	USER USER USER	Wajah anak lo Kliatan Bgt ya!! Penuh dendam dan kebencian! Pasti anak anak lo makan dari hasil uang haram. Bapaknya kayak lo!! Banci dan rumpi!!! Lo ngatain gue makan tai?? Nah lo dan istri serta anak anak dan ortu lo makan bangke uang haram	1	1	1	1
57	Menyamakan Allah SWT yang dungu & biadab dengan Tuhan di balik gerhana adalah PENYESATAN		1	1	1	1
58	USER	Jangan jangan otak makelar penjual asset negara apa maling asset negara ? \xf0\x9f\x91\x89USER pertamina di jual mau	1	1	0	1

Hak Cipta Diindungi Undang-Undang

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

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- 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 milik UIN Suska Riau		ingin meniadakan KOREM,KODIM,KORAMIL.? Zona merah masih tercatat rapi di Korem.kodim.koramil. pak rymizard ryacudu				
73		bagai rumah bawah ranah yang indah, seorang pendiam dan culun walau dengan bakar besar #keren #bijak #semangat'	0	0	0	0
74		hantu kak limah 2 & munafik 2 hmm'	0	0	0	0
75		Bila sesuai kepentingan jokowi didukung bila mengganggu kepentingan akan disikat	1	0	0	0
76		USER emang cuma babi yg punya buntut hihihi mungkin kamu unicorn'	0	1	0	1
77		Selamat Hari Koperasi Nasional, Koperasi sebagai landasan ekonomi bangsa harus dipelihara dan ditumbuh kembangkan dari ancaman praktek kartel asing. \xf0\x9f\x87\xae\xf0\x9f\x87\xa9'	0	0	0	0
78		Mereka tidak tertarik melihat KORUPTOR BLBI yg pelaku2nya Taipan binaan KOMUNIS CHINA dan mitra nya MEGAWATI SP	1	1	0	0
79		USER Anjing2 kelaparan yg rakus duit skrg pesta pora.... curut2 maling duit rakyat makan kenyang skrg... tereak2 kopar kapir aseng asing cuma kedok buat goblokin pemujanya'	1	1	1	1
80		USER Alah bangkai lah kau dh balik sendirii padan muka kau jejauh sana hahahaa'	1	1	1	1
81	State Islamic University of Sultan Syarif Kasim Riau	RT USER: ABG Toge Mendesah Ngentot Diatas\n\n#VIDEOBOKEP Full ==> URL URL	0	0	1	1
82		USER USER Coba klu bukan petugas partai...sy rasa bisa menjalankan pemerintahan dgn benar dan tdk perlu byk infrastrutur klu sektor yg lain malah nggak jalan kok skrg malah byk hutang dan byk tka yg masuk.pikirkan itu jgn byk pencitraan.'	1	1	0	0
83		RT USER Infrastruktur yg mana era SBY bangun setahu saya gk ada	1	1	0	0
84		Memakai Cara Licik Membuat Akun Palsu dan Menyebarkannya, Kemudian Membully Akun Palsu Itu utk Membunuh Karakter Orang, adalah Salah Satu Cara PKI dan Orang Atheis Anti Allah dan Agama.; Sdh Diperingatkan Jendral Nasution Dahulu Kala.; Waspada! KOMUNIS dan	1	1	0	0

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.

85	Ini screenshot udh agak lama tapi lupa mau kuposting \xf0\x9f\x98\xa5ada yg ketawa ga? Sirinya agak budek dengernya bungkus \xf0\x9f\x98\xb0 URL	0	0	1	1
86	Ga kuat ngadepin bacot ibu ibu. Sumpah gakuat.'	0	1	1	1
87	USER Gue cuma makan snack doang anjir belum makan nasi wkwkwk'	0	0	1	1
88	Tawuran itu bukan budaya bangsa INDONESIA	0	0	0	0
89	Ini ngapa yak USER sama USER sama sama kacrut jaringannya. Taekkkkkk.'	1	1	1	1
90	ini kok ngenes ya bacanya.	0	1	0	1
91	RT USER USER USER Heleh istana di kuasai cina hancur sudah rakyat indonesia	1	1	0	0
92	dia enggak hebat. pengikutnya aja yang dungu bin idiot. URL	1	1	1	1
93	Depan cermin bukan main elok buat posing bagai,cuba dlm gambar nohhai CACAT HAHHAHAHAHA'	0	0	1	0
94	USER kamu ga budek ko, cuma tuli\n'nnga'	0	1	1	1
95	Hanya butuh laki setia bukan laki bajingan \xf0\x9f\x91\xbf\xf0\x9f\x91\xbf'	0	1	1	1
96	(apalagi selain mereka) sehingga mereka tidak mampu melepaskan diri darinya".*_ (HR: Tirmidzi dengan sanad hasan menurut syaikh Al Bani).\nUlama suu\xe2\x80\x99 (sesat/jahat) akan menyuguhkan keburukan dalam bentuk kebaikan, membungkus kesesatan dengan ke	0	0	0	0
97	USER PLH JKW EKONOMI SMKIN MEROKET	0	1	0	0
98	awas ini Proyek PKI CINA KOMUNIS	1	1	0	0
99	USER Biasa.. bani coli, ngga beda jauh sm junjungan dan kawan2nya.. haha'	1	1	1	1
...
	Kita akan belajar giat, dengan penuh dedikasi, tapi bukan hanya untuk lulus ujian. (3 idiot). #TweetDSAP'	0	0	0	0

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 RIWAYAT HIDUP

INFORMASI PERSONAL



Nama	Tamara Ghassani Saskia
NIM	11551200583
Jurusan	Teknik Informatika
Agama	Islam
E-mail	tamara.ghassani.saskia@students. uin-suska.ac.id

INFORMASI PENDIDIKAN

Tahun 2002-2009	SDN 58 Pekanbaru
Tahun 2009-2012	SMPN 13 Pekanbaru
Tahun 2012-2015	SMAN 5 Pekanbaru
Tahun 2015-2021	UIN Sultan Syarif Kasim Riau S1 Teknik Informatika

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 milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau