



USKA RIAU

© Hak Cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

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

IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK MENDETEKSI *HATE SPEECH* DAN *ABUSIVE LANGUAGE* PADA TWITTER BAHASA INDONESIA

TUGAS AKHIR

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

Oleh :

AULIL AMRI
11651103437



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2020



Hak Cipta Dilindungi Undang-Undang

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

LEMBAR PERSETUJUAN

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK
MENDETEKSI *HATE SPEECH* DAN *ABUSIVE LANGUAGE*
PADA TWITTER BAHASA INDONESIA**

TUGAS AKHIR

Oleh :

AULIL AMRI
11651103437

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 06 Agustus 2020

Pembimbing,

Surya Agustian, S.T, M.Kom
NIP. 197908302011011003

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

**IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK
MENDETEKSI *HATE SPEECH* DAN *ABUSIVE LANGUAGE* PADA
TWITTER BAHASA INDONESIA**

TUGAS AKHIR

Oleh :

AULIL AMRI
11651103437

Telah dipertahankan di depan sidang dewan penguji
Sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau di
Pekanbaru, pada tanggal 06 Agustus 2020 / 16 Dzulhijjah 1441

Pekanbaru, 06 Agustus 2020
Mengesahkan,

Ketua Jurusan,

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



Wakil Dekan I,

Dr. Harris Simaremare, ST, MT.
NIP. 19830625 200801 1 008

DEWAN PENGUJI

- Ketua : Dr. Alwis Nazir, M.Kom
 Sekretaris : Surya Agustian, S.T, M.Kom
 Penguji I : Fadhilah Syafria, S.T, M.Kom, CIBIA
 Penguji II : Muhammad Affandes, MT



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 didalam daftar pustaka.

Pekanbaru, 06 Agustus 2020

Yang Membuat Pernyataan,

AULIL AMRI
11651103437

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

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

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

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 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.

IMPLEMENTASI ALGORITMA *RANDOM FOREST* UNTUK MENDETEKSI *HATE SPEECH* DAN *ABUSIVE LANGUAGE* PADA TWITTER BAHASA INDONESIA

AULIL AMRI
11651103437

Tanggal Sidang : 06 Agustus 2020
Periode Wisuda :

Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Ujaran kebencian (*hatespeech*) dan Bahasa kasar (*abusive language*) adalah suatu tindakan negatif yang sering kali terjadi di lingkungan kita. Terlebih lagi dengan bantuan teknologi yang semakin maju, siapa saja bisa melakukan penyebaran ujaran kebencian ataupun bahasa kasar ke siapa saja yang ia sukai. Umumnya peristiwa ini berlaku pada masa pemilihan presiden, bupati, dan sebagainya. Sering terjadi pertikaian antara masing-masing pihak yang berkepentingan. Salah satunya melalui media sosial *twitter*. Dengan melakukan sebuah *tweet* ujaran kebencian dan di *re-tweet* oleh sekelompok lainnya. Namun hampir tak dapat dibedakan sebuah *tweet* itu apakah termasuk ke dalam *hatespeech* ataupun *abusive*. Dengan adanya penelitian ini diharapkan dapat membantu kita semua untuk dapat membedakan keduanya dan lebih bijak lagi dalam ber-media sosial. Dengan menggunakan data *twitter* sebanyak 13.167 *tweet* serta metode *Random Forest* sebagai pengklasifikasi. Berdasarkan hasil pengujian metode *Random Forest* dengan menggunakan fitur terbaik dan parameter terbaik mendapatkan hasil akurasi sebesar **75,96%**.

Kata Kunci : *hatespeech*, *abusive*, *tweet*, *twitter*, Klasifikasi, *Random Forest*.

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.

**IMPLEMENTATION OF RANDOM FOREST ALGORITHM TO
DETECT HATE SPEECH AND ABUSIVE LANGUAGE IN
INDONESIAN TWITTER**

**AULIL AMRI
11651103437**

Date of Final Exam : 06th August 2020

Date of Graduation Ceremony :

Informatic Engineering Department

Faculty of Science and Technology

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRACT

The words of hate (hatespeech) and abusive language are negative actions that often occur in our environment. Moreover, with the help of increasingly advanced technology, anyone can disseminate hate speech or abusive language to anyone he likes. Generally this event applies to the election of presidents, regents, and so on. Disputes often occur between each of the parties concerned. One of them is through Twitter social media. By doing a hate speech tweet and re-tweeted by another group. But it is almost indistinguishable whether a tweet is included in hatespeech or abusive. The existence of this research is expected to help us all to be able to distinguish between them and be wiser in social media. By using Twitter data of 13,127 tweets and the Random Forest method as a classification. Based on the results of testing the Random Forest method using the best features and the best parameters get an accuracy of 75.96%.

Keywords : *hatespeech, abusive, tweet, twitter, classification, random forest.*

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



Assalamua'alaikum Warahmatullahi Wabarakatuh

Alhamdulillahirobbil'alamiin. Puji syukur kepada Allah SWT atas seluruh rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian dan penulisan Laporan Tugas Akhir di Jurusan Teknik Informatika dengan Judul "Implementasi Algoritma *Random Forest* untuk mendeteksi *Hate Speech* dan *Abusive Language* pada Twitter Bahasa Indonesia". Laporan Tugas Akhir ini merupakan prasyarat untuk memperoleh gelar sarjana strata satu di Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau (UIN SUSKA RIAU).

Selama dilaksanakannya penelitian dan penulisan Laporan Tugas Akhir ini, penulis mendapatkan banyak pengetahuan, pengalaman, bimbingan, dukungan dan juga arahan dari semua pihak yang telah membantu hingga Laporan Tugas Akhir ini dapat diselesaikan. Untuk itu pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Allah SWT dengan segala rahmat-Nya memberikan semua yang terbaik dan dengan hidayah-Nya memberikan petunjuk sehingga dalam penyusunan laporan ini berjalan lancar.

Bapak Prof. Dr. H. Ahmad Mujahidin, M.Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.

Bapak Drs. Ahmad Darmawi, M.Ag selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.

Ibu Elin Haerani, S.T., M.Kom., selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau.

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 Dilindungi Undang-Undang

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

Ibu Fadhilah Syafria, S.T, M.Kom, CIBIA, selaku dosen penguji I yang telah meluangkan waktunya dan banyak memberikan saran, dan masukan yang membangun bagi penulis.

Bapak Muhammad Affandes, MT, selaku dosen penguji II yang telah meluangkan waktunya dan banyak memberikan saran, dan masukan yang membangun bagi penulis.

Bapak Novriyanto, S.T, M.Sc, selaku Pembimbing Akademis penulis selama menjalani perkuliahan di Jurusan Teknik Informatika, yang telah banyak memberikan arahan serta masukan mengenai perkuliahan bagi penulis.

Seluruh Bapak/Ibu dosen Teknik Informatika yang telah sabar memberikan tunjuk ajar serta ilmu yang bermanfaat kepada penulis selama masa perkuliahan.

10. Kedua orang tua penulis, Ayah dan Ibu yang tidak pernah letih memberikan semangat, nasehat, dan doa yang tulus untuk anaknya. Semoga Allah SWT senantiasa memberikan kesehatan kepada Ayah dan Ibu.
 11. Teman-teman seperjuangan kelas B angkatan 2016 yang selalu berjuang bersama dan selalu saling mendukung.
 12. Kos Awak, yang telah banyak memberikan saran, semangat, dan motivasi kepada penulis, semoga kita semua dapat mencapai apa yang diinginkan.
- Semua pihak yang tidak dapat penulis sebutkan satu persatu. Terimakasih atas dukungan moril maupun materil dalam pengerjaan Laporan Tugas Akhir ini.

Semoga laporan ini dapat bermanfaat bagi penulis khususnya maupun pembaca pada umumnya. Serta dapat menjadi referensi dan rujukan bagi hal-hal yang bermanfaat. Penulis berharap adanya kritik dan saran guna memperbaiki atau sebagai pengembangan kedepannya. Akhir kata penulis ucapkan terima kasih dan selamat membaca.

Wassalamualaikum Wr. Wb

Pekanbaru, Agustus 2020

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	xv
DAFTAR TABEL.....	xvii
BAB I PENDAHULUAN.....	I-1
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-4
1.3 Batasan Masalah.....	I-5
1.4 Tujuan Penelitian	I-5
1.5 Sistematika Penulisan	I-5
BAB II LANDASAN TEORI.....	II-1
2.1 Twitter	II-1
2.2 <i>Hate speech</i> (Ujaran Kebencian)	II-1
2.3 <i>Abusive language</i> (Bahasa Kasar).....	II-3
2.4 Python	II-4
2.5 <i>Natural Language Processing (NLP)</i>	II-5

Hak Cipta Dilindungi Undang-Undang

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



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.

2.5.1 <i>Tokenizing</i>	II-6
2.5.2 <i>Case folding</i>	II-6
2.5.3 <i>Cleaning</i>	II-6
2.5.4 <i>Filtering (Stopword Removal)</i>	II-6
2.6 <i>Machine Learning</i>	II-7
2.6.1 <i>Machine Learning Workflow</i>	II-8
2.7 <i>Word Embedding</i>	II-9
2.8 <i>Random Forest Classifier</i>	II-10
2.8.1 <i>Keuntungan dan Kerugian Random Forest</i>	II-12
2.8.2 <i>Parameter Random Forest</i>	II-13
2.8.3 <i>Algoritma Random Forest</i>	II-14
2.8.4 <i>Kerja Algoritma Random Forest</i>	II-15
2.9 <i>Feature Selection</i>	II-16
2.10 <i>Confusion Matrix</i>	II-17
2.11 <i>Penelitian Terkait</i>	II-18
BAB III METODOLOGI PENELITIAN	III-1
3.1 <i>Identifikasi Masalah</i>	III-2
3.2 <i>Pengumpulan Data</i>	III-2
3.2.1 <i>Pengumpulan Data tweet Hate Speech dan Abusive</i>	III-2
3.2.2 <i>Studi Literatur</i>	III-4
3.3 <i>Analisa</i>	III-5
3.3.1 <i>Analisa Dataset Preparation</i>	III-5
3.3.2 <i>Analisa Text Preprocessing</i>	III-5
3.3.3 <i>Analisa Language Model Training</i>	III-6



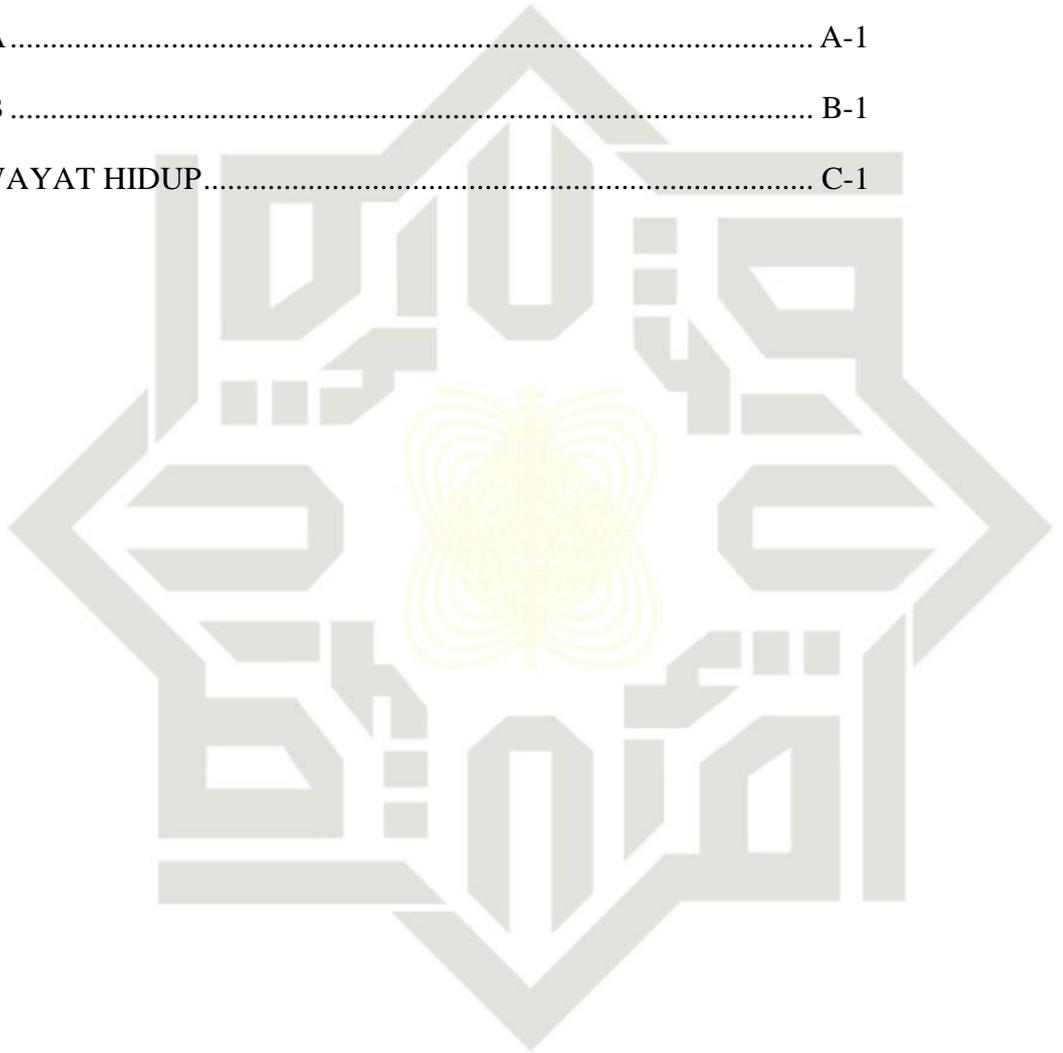
Hak Cipta Dilindungi Undang-Undang

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.

3.3.4 Analisa Parameter Terbaik.....	III-6
3.3.5 Analisa <i>Feature Selection</i>	III-7
3.3.6 Analisa Klasifikasi Random Forest.....	III-7
3.4 Perancangan	III-8
3.5 Implementasi	III-9
3.6 Pengujian.....	III-9
3.7 Kesimpulan dan Saran.....	III-10
BAB IV ANALISA DAN PERANCANGAN.....	IV-1
4.1 Analisa	IV-1
4.1.1 <i>Dataset Preparation</i>	IV-1
4.1.2 <i>Text Preprocessing</i>	IV-13
4.1.3 <i>Language Model Training</i>	IV-14
4.1.4 Mencari Parameter Terbaik.....	IV-18
4.1.5 <i>Feature Selection</i>	IV-20
4.2 Perancangan	IV-21
4.2.1 <i>Random Forest Classifier</i>	IV-21
BAB V IMPLEMENTASI DAN PENGUJIAN	V-1
5.1 Implementasi	V-1
5.1.1 Batasan Implementasi	V-1
5.1.2 Lingkungan Implementasi.....	V-1
5.1.3 Implementasi Model Random Forest.....	V-2
5.2 Pengujian.....	V-8
5.2.1 <i>Confusion Matrix</i>	V-8
5.2.2 Hasil Pengujian	V-10



BAB VI PENUTUP	VI-1
6.1 Kesimpulan	VI-1
6.2 Saran.....	VI-1
DAFTAR PUSTAKA	xix
LAMPIRAN A	A-1
LAMPIRAN B	B-1
DAFTAR RIWAYAT HIDUP.....	C-1



UIN SUSKA RIAU

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

DAFTAR GAMBAR

Gambar	Halaman
Gambar 2. 1 Twitter <i>User Interface</i>	II-1
Gambar 2. 2 Ilustrasi <i>Word Embedding</i>	II-10
Gambar 2. 3 Ilustrasi <i>Random Forest</i>	II-11
Gambar 3. 1 Tahapan Metodologi Penelitian	III-1
Gambar 3. 2 Distribusi <i>tweet Abusive</i> ke <i>tweet HS dan Not HS</i>	III-4
Gambar 4. 1 Planning Penelitian.....	IV-1
Gambar 4. 2 <i>Import Library</i>	IV-3
Gambar 4. 3 <i>Import Dataset</i>	IV-3
Gambar 4. 4 Proses Pengecekan Label	IV-4
Gambar 4. 5 Proses Pemisahan Label	IV-5
Gambar 4. 6 Hapus Label <i>Abusive</i>	IV-5
Gambar 4. 7 Hapus Label HS dan Level	IV-6
Gambar 4. 8 <i>Dataset Splitting</i>	IV-8
Gambar 4. 9 Bagi Data Train dan <i>Test</i>	IV-11
Gambar 4. 10 Tahap Preprocessing	IV-13
Gambar 4. 11 <i>Import Library FastText</i>	IV-14
Gambar 4. 12 <i>Import Data Training</i>	IV-15
Gambar 4. 13 Proses Tokenisasi	IV-15
Gambar 4. 14 <i>Training Model</i>	IV-16
Gambar 4. 15 Proses <i>L1-norm</i>	IV-17
Gambar 4. 16 Proses <i>L2-norm</i>	IV-18
Gambar 4. 17 Alur Klasifikasi <i>Random Forest</i>	IV-22

Hak Cipta Dilindungi Undang-Undang

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

Gambar 4. 18 Salah satu <i>tree</i> yang terbentuk dari dataset pada algoritma <i>Random Forest</i>	IV-23
Gambar 5. 1 <i>Import Library</i>	V-3
Gambar 5. 2 <i>Import Data Training dan Testing</i>	V-3
Gambar 5. 3 <i>Load Model FastText</i>	V-4
Gambar 5. 6 <i>Proses Encoding</i>	V-4
Gambar 5. 7 <i>Kode Program Training HS</i>	V-5
Gambar 5. 8 <i>Kode Program Testing HS</i>	V-5
Gambar 5. 9 <i>Kode Program Training Abusive</i>	V-5
Gambar 5. 10 <i>Kode Program Testing Abusive</i>	V-6
Gambar 5. 11 <i>Kode Program Training Level</i>	V-6
Gambar 5. 12 <i>Kode Program Testing Level</i>	V-6
Gambar 5. 13 <i>Klasifikasi Pada Label HS</i>	V-7
Gambar 5. 14 <i>Klasifikasi Pada Label Abusive</i>	V-7
Gambar 5. 15 <i>Klasifikasi Pada Label Level</i>	V-7
Gambar 5. 16 <i>Rata-Rata Akurasi</i>	V-7
Gambar 5. 17 <i>Training Model HS, Abusive dan Level</i>	V-8
Gambar 5. 18 <i>Hasil Confusion Matrix Label HS</i>	V-9
Gambar 5. 19 <i>Hasil Confusion Matrix Label Abusive</i>	V-9
Gambar 5. 20 <i>Hasil Confusion Matrix Label Level</i>	V-9



DAFTAR TABEL

Tabel	Halaman
Tabel 2. 1 <i>Confusion Matrix</i>	II-17
Tabel 2. 2 Penelitian Terkait	II-18
Tabel 4. 1 Dataset Awal	IV-1
Tabel 4. 2 Dataset Label HS	IV-5
Tabel 4. 3 Dataset Label Abusive	IV-6
Tabel 4. 4 Dataset Label Level	IV-7
Tabel 4. 5 Data <i>Train</i> Label HS	IV-8
Tabel 4. 6 Data Test Label HS	IV-9
Tabel 4. 7 Data <i>Train</i> Label <i>Abusive</i>	IV-10
Tabel 4. 8 Data <i>Test</i> Label <i>Abusive</i>	IV-10
Tabel 4. 9 Data Train Label Level	IV-12
Tabel 4. 10 Data <i>Test</i> Label Level	IV-12
Tabel 4. 11 Hasil Pengujian L1- <i>norm</i>	IV-19
Tabel 4. 12 Hasil Pengujian L2- <i>norm</i>	IV-20
Tabel 4. 13 Tabel <i>Experiment</i> L1- <i>norm</i>	IV-20
Tabel 4. 14 Tabel <i>Experiment</i> L2- <i>norm</i>	IV-21
Tabel 4. 15 Tabel Dataset HS dan <i>Abusive</i>	IV-24
Tabel 4. 16 <i>Binary Split</i> Pertama HS	IV-24
Tabel 4. 17 <i>Binary Split</i> Kedua HS	IV-25
Tabel 4. 18 Tabel <i>Binary Split</i> Ketiga HS	IV-25
Tabel 4. 19 <i>Binary Split</i> Keempat HS	IV-26
Tabel 4. 20 <i>Binary Split</i> Kelima HS.....	IV-26

- 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.



Tabel 4. 21 Tabulasi Nilai <i>Gini Index</i> atribut HS	IV-27
Tabel 4. 22 <i>Binary Split</i> Pertama AB.....	IV-27
Tabel 4. 23 <i>Binary Split</i> Kedua AB	IV-28
Tabel 4. 24 <i>Binary Split</i> Ketiga AB	IV-28
Tabel 4. 25 <i>Binary Split</i> Keempat AB	IV-29
Tabel 4. 26 <i>Binary Split</i> Kelima AB	IV-29
Tabel 4. 27 Tabulasi Nilai Indeks atribut <i>Abusive</i>	IV-30
Tabel 4. 28 <i>Binary Split</i> Pertama Level	IV-30
Tabel 4. 29 <i>Binary Split</i> Kedua Level.....	IV-31
Tabel 4. 30 <i>Binary Split</i> Ketiga Level.....	IV-31
Tabel 4. 31 <i>Binary Split</i> Keempat Level.....	IV-32
Tabel 4. 32 <i>Binary Split</i> Kelima Level.....	IV-32
Tabel 4. 33 Tabulasi Nilai Indeks Atribut Level.....	IV-33
Tabel 5. 1 Pengujian <i>Confusion Matrix</i>	V-10

Hak Cipta Dilindungi Undang-Undang

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

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi merupakan bidang yang memiliki kemajuan sangat cepat. Setiap tahunnya selalu ada penemuan-penemuan baru yang dikembangkan oleh para penggiat teknologi. Penemuan ataupun inovasi tersebut tentunya bertujuan untuk dapat mempermudah pekerjaan manusia. Sehingga siapapun tak akan mau ketinggalan untuk mempelajari teknologi baru tersebut demi kemudahan-kemudahan yang sangat berguna bagi manusia. Salah satu perkembangan dari bidang teknologi adalah media sosial, yang pada dasarnya berguna sebagai sarana untuk berkomunikasi jarak jauh. (D. Haryanto & Argadila, 2019)

Media sosial merupakan tempat untuk berbagi informasi dan berkomunikasi dengan orang yang berada dekat maupun jauh dengan kita secara daring dengan bantuan internet. Teknologi selalu melakukan perubahan tiap tahunnya, hingga media sosial pun turut ambil bagian dengan mengembangkan fitur-fitur yang dimiliki, kini media sosial tak hanya sekedar tempat untuk berkomunikasi, bahkan media sosial bisa menjadi wadah bagi penggunanya untuk menghasilkan uang, seperti bisnis *online shop* dan sebagainya. (Abdul Rahin Bin Hamdan dkk, 2018)

Dari sekian banyak media sosial yang ada di dunia, terdapat beberapa *platform* media sosial yang populer di Indonesia. Facebook, Instagram, Twitter, dan Line, adalah *platform* yang cukup bersinar di Indonesia. Meskipun media sosial dibuat dengan tujuan untuk mempermudah manusia dan hal-hal positif lainnya, namun tak dapat dipungkiri bahwa setiap perubahan yang terjadi pada manusia tetap memiliki sisi positif dan negatif. (Katadata.co.id, 2019)

Penyebaran ujaran kebencian (*hate speech*) dan Bahasa kasar (*abusive language*) adalah bagian kecil dari sisi negatif yang dapat terjadi melalui sosial media. Meski pelanggaran dan penyalahgunaan sosial media sudah diatur didalam undang-undang, namun tidak semua manusia sadar dan patuh terhadap peraturan.

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.

Ujaran kebencian (*hate speech*) merupakan perbuatan yang dilakukan oleh individu maupun kelompok dengan tujuan ingin menjatuhkan individu atau kelompok lainnya. Provokasi, fitnah, hinaan adalah bentuk dari ujaran kebencian. Ujaran kebencian dalam ruang lingkup sosial media sering terjadi dengan konteks atau aspek ras, warna kulit, jenis kelamin, agama, dan sebagainya. (Fauzi & Yuniarti, 2018)

Bahasa kasar (*abusive language*) sering diungkapkan karena kekesalan, emosi, kecewa, atas sebuah peristiwa yang terjadi dengan individu atau kelompok tertentu. Dalam sosial media Bahasa kasar sering di implementasikan pada hal-hal yang tergolong ke dalam konteks SARA (suku, agama, ras, dan antar golongan). (Hidayatullah dkk., 2019)

Salah satu contoh kasus tentang pelanggaran dalam media sosial yaitu pada kasus istri perwira TNI yang mengunggah sebuah postingan di salah satu media sosial terkait dengan peristiwa penusukan teroris terhadap Menko Polhukam Wiranto pada November 2019 lalu. Postingan tersebut dianggap pencemaran nama baik sehingga berakhir dengan dicopotnya sang suami dari jabatannya di TNI¹. “Permasalahan yang dialami istri prajurit sebaiknya tidak berdampak pada penghukuman maupun pencopotan sang suami dari jabatannya,” ucap salah seorang pengamat militer yang merangkap Direktur Imparsial, Al-Araf.

Ada beberapa pendapat mengenai postingan istri prajurit TNI tersebut. Ada yang menganggap postingan tersebut tidak pantas dan termasuk ke dalam pencemaran nama baik karena waktunya yang bersamaan dengan peristiwa penusukan Menko Polhukam Wiranto, dan ada juga yang berpendapat bahwa postingan tersebut tidak bisa dikatakan cacian atau hinaan karena tidak langsung menjeru ke suatu objek. Untuk mencegah terjadinya kesalahpahaman antara *hate speech* dan *abusive language*, perlu dilakukan penelitian lebih lanjut mengenai *hate speech*, kategori-kategori kalimat yang termasuk ke dalam *hate speech*, dan level-level *hate speech*, agar masyarakat dapat membedakan kalimat *hate speech* dan *abusive* agar kedepannya dapat menggunakan sosial media dengan lebih bijak.

¹<https://www.kompas.com/tren/read/2019/10/13/083507365/di-balik-kasus-penusukan-wiranto-kan-penangkapan-sejumlah-terduga-teroris?page=all>

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.

Pada penelitian yang dilakukan oleh (A. T. Haryanto, 2018), terdapat informasi tentang pengguna aktif sosial media di Indonesia. Terdapat sekitar 130 juta penduduk Indonesia yang aktif dalam dunia maya atau sosial media. Twitter adalah *platform* sosial media yang cukup populer dengan fitur andalannya *tweet* dan *re-tweet*. Twitter banyak digunakan penduduk Indonesia sebagai tempat untuk berbagi cerita, informasi, dan pengalaman hidup. Penyebaran berita ataupun informasi melalui twitter dapat terbilang sangat cepat, karena fitur *re-tweet* mampu membuat pengguna lainnya ikut serta dalam menyebarkan berita atau kejadian ke pengguna lainnya.

Beberapa penelitian sebelumnya yang dijadikan sebagai referensi pada penelitian ini sebagai berikut. Penelitian yang dilakukan oleh (Putra dkk., 2018) yang menggunakan metode *Ensemble Feature* dan *Support vector machine* (SVM). (Putra dkk., 2018) mengidentifikasi ujaran kebencian pada *Facebook*, karena belum ada sebuah sistem otomatis yang dapat mendeteksi ujaran kebencian. Penelitian tersebut menggunakan metode *Ensemble feature* untuk menggabungkan beberapa fitur dari tiap teks yang ada untuk mempermudah proses identifikasi. Sedangkan metode SVM berperan dalam proses identifikasi terhadap sebuah teks berdasarkan data-data fitur yang diperoleh dari metode *Ensemble feature*. Berdasarkan hasil pengujian, didapatkan akurasi dari sistem sebesar 70%, oleh karena itu dapat ditarik sebuah kesimpulan bahwa metode *Ensemble feature* dan *Support Vector Machine* dapat digunakan dengan baik dalam mengidentifikasi ujaran kebencian pada media sosial *Facebook*.

Pada penelitian (Sudiantoro dkk., 2018) dengan judul analisis sentimen *Twitter* menggunakan *text mining* dan algoritma *Naïve Bayes Classifier*. Tujuan dari penelitian ini untuk melakukan klasifikasi terhadap data *tweet* menjadi dua bagian yaitu sentimen positif dan negatif. (Sudiantoro dkk., 2018) menggunakan data *Twitter* Bahasa Indonesia terkait pendapat masyarakat terhadap pelaksanaan pilkada Jawa barat.

Data yang akan diproses sebanyak 300 data *tweet*. Kemudian dibagi untuk data latih dan data uji, data latih 200 data, dan data uji 100 data. Dari 100 data uji

yang diklasifikasi, diperoleh 32 data yang bersentimen positif dan 68 data bersentimen negatif. Maka dapat ditarik sebuah kesimpulan bahwa berdasarkan 100 data uji yang telah dilakukan proses klasifikasi dengan hasil 68 data termasuk sentimen negatif, dan akurasi dari *naïve bayes classifier* sebesar 84%.

Berkaitan dengan judul penelitian yang akan diteliti mengenai algoritma *Random Forest*, terdapat beberapa referensi terkait diantaranya, pada penelitian (Wan & Gao, 2016) tentang analisis sentimen pelanggan maskapai penerbangan. Dengan menerapkan beberapa metode *classifier* yaitu *Naïve Bayes Classifier (NBC)*, *Support vector machine (SVM)*, Jaringan *Bayesian*, Algoritma *C4.5*, *Decision Tree* dan *Random Forest*. Berdasarkan percobaan dengan enam pendekatan klasifikasi individu dan pendekatan *ensemble* yang dilatih dan diuji menggunakan *dataset* yang sama dari 12864 *tweet*, dimana evaluasi 10 kali lipat digunakan untuk memvalidasi klasifikasi. Dalam percobaan dengan tiga kelas, *ensemble classifier* mendapatkan akurasi tertinggi dalam hal *Precision*, *Recall*, dan *F-measure*, yaitu 84,2%. Tingkat kesalahan untuk *ensemble classifier* adalah yang terendah dibandingkan dengan enam metode klasifikasi lainnya, yaitu 15,8%. Hasilnya menunjukkan bahwa usulan pendekatan *ensemble* lebih unggul untuk kasus klasifikasi dalam *dataset* twitter layanan maskapai ini.

Berdasarkan masalah yang terdapat pada latar belakang, maka penulis akan melakukan analisa terhadap metode *Random Forest* dalam mengklasifikasi *multi-class* dan *multi-label hate speech* dan *abusive language* pada *twitter* Bahasa Indonesia.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas dalam penelitian ini yaitu bagaimana cara mengklasifikasikan *multi-class* dan *multi-label hate speech* dan *abusive language* pada *twitter* Bahasa Indonesia dengan menggunakan *Random forest* sebagai *Classifier*, serta melihat performa *random forest* dalam mengklasifikasi *multi-class* dan *multi-label hate speech* dan *abusive language* pada *twitter* Bahasa Indonesia.

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.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini, diantaranya :

1. *Dataset* yang diproses sebanyak 13127 *tweet* (Ibrohim & Budi, 2019)
2. Label yang diproses yaitu *hate speech*, *abusive*, dan level
3. Level *hate speech* yang dideteksi yaitu level kuat, menengah, dan lemah.
4. Metode klasifikasi yang digunakan adalah *Random Forest*
5. Metode *embedding* kata yang digunakan adalah *FastText*
6. *Output* yang dihasilkan dari penelitian ini adalah klasifikasi *multi-class* dan *multi-label hate speech* dan *abusive language* pada Twitter Bahasa Indonesia

1.4 Tujuan Penelitian

Beberapa tujuan yang ingin dicapai dalam penelitian ini, diantaranya:

1. Implementasi metode *Random Forest* untuk mengklasifikasi *multi-class* dan *multi-label hate speech* dan *abusive language* pada Twitter Bahasa Indonesia
2. Menghitung akurasi dari algoritma *Random Forest* dalam mengklasifikasi *multi-class* dan *multi-label hate speech* dan *abusive language* pada twitter bahasa Indonesia
3. Melakukan proses *feature selection* untuk menemukan *set feature* terbaik agar mendapatkan model yang mampu memprediksi dengan lebih akurat.

1.5 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini terdiri dari beberapa bagian diantaranya:

Bab I

Pendahuluan

Pada bab ini membahas tentang latar belakang yang mendasari penelitian, rumusan masalah yang akan diteliti, tujuan penelitian, Batasan-batasan masalah dalam penelitian, dan sistematika penulisan tugas akhir.

Bab II

Landasan Teori

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.

Pada bab ini akan menjelaskan mengenai studi pustaka terhadap teori-teori yang berkaitan dengan penelitian ini, analisa serta penulisan tugas akhir.

Metodologi Penelitian

Pada bab ini akan dijelaskan bagaimana cara membuat kerangka penelitian serta uraian dari metodologi penelitian yang akan digunakan pada tugas akhir ini.

Analisa Dan Perancangan

Pada bab ini akan menjelaskan bagaimana proses analisa dan perancangan dari penelitian ini.

Implementasi Dan Pengujian

Pada bab ini akan dilakukan implementasi pada model yang telah dibangun serta dilakukan pengujian apakah model yang dibuat berjalan dengan baik atau tidak.

Penutup

Pada bagian ini akan menjabarkan tentang hasil akhir atau kesimpulan dari penelitian, serta saran-saran yang dapat digunakan untuk mengembangkan penelitian ini.

Hak Cipta Dilindungi Undang-Undang

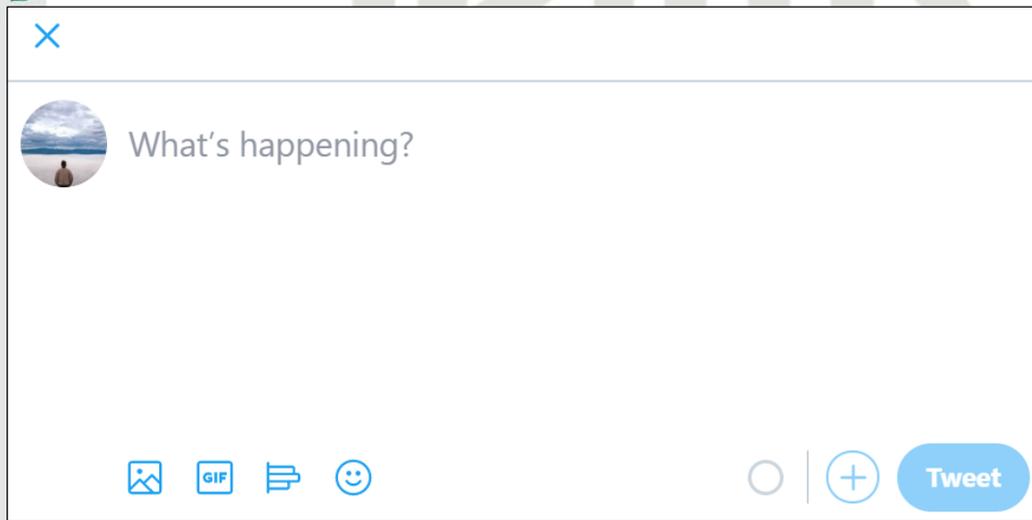
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

BAB II

LANDASAN TEORI

2.1 Twitter

Twitter adalah salah satu dari sekian banyak media sosial yang ada di Indonesia. Twitter merupakan sosial media yang cukup populer di Indonesia, dengan fitur *tweet* dan *re-tweet* yang dimilikinya, twitter dapat bersaing dengan situs besar seperti Facebook. Twitter dimulai dengan satu pertanyaan : “*What’s happening?*” dan dijawab hanya dalam 140 karakter saja. Tampilan dari halaman *user interface* twitter dapat dilihat pada gambar 2.1 dibawah ini.



Gambar 2. 1 Twitter User Interface

Tweet entity atau disebut juga sebagai konten *tweet* adalah konten/isi dari *tweet* itu sendiri. Bagian dari *tweet entity* yaitu : teks dari *tweet* itu sendiri, *hashtag* (#), *mention* (@), *retweet* (RT), url, dan *emoticons*.

2.2 Hate speech (Ujaran Kebencian)

Studi tentang ujaran kebencian dalam beberapa tahun terakhir telah menarik perhatian sejumlah besar sarjana dari berbagai bidang pengetahuan (ahli bahasa, sosiolog, filsuf, sejarawan, psikolog, antropolog, pengacara dan ilmuwan politik, dan lain-lain (Neshkovska & Trajkova, 2018). Hal yang menjadikan para pendekiawan ingin menangani masalah ini dengan pertimbangan kemungkinan

bahwa di masa-masa sulit yang kita hadapi ini, masyarakat di seluruh dunia sedang terpolarisasi secara mendalam dengan begitu banyak perbedaan seperti agama, politik, etnis, dan lain-lain, dan berada dalam keadaan yang terus berubah. Pada akhirnya membuat mereka sangat rentan terhadap kebencian yang dampak buruknya kadang-kadang bisa diluar nalar manusia.

Pertanyaan inti yang muncul dalam membahas ujaran kebencian adalah apa itu ujaran kebencian. Tinjauan literatur yang relevan mengungkapkan bahwa ujaran kebencian adalah fenomena yang kompleks dan sangat diperdebatkan, dan belum ada satu pun definisi yang dapat diterima secara bulat tentang apa sebenarnya ujaran kebencian itu.

Definisi paling sederhana bahwa ucapan kebencian adalah ucapan apa pun yang menyebabkan beberapa orang tersinggung (Lewis, 2013).

Definisi lain yang sedikit lebih luas tentang ujaran kebencian (*hate speech*) adalah tindakan komunikasi yang dilakukan oleh suatu individu atau kelompok dalam bentuk provokasi, hasutan, ataupun hinaan kepada individu atau kelompok yang lain dalam berbagai aspek seperti ras, warna kulit, jenis kelamin, agama, orientasi seksual, dan lain-lain. (Fauzi & Yuniarti, 2018)

Salah satu definisi yang paling sering dikutip tentang pidato kebencian adalah yang diusulkan oleh Dewan Eropa. Menurut Dewan Eropa “semua bentuk ekspresi yang menyebar, menghasut, mempromosikan atau membenarkan kebencian rasial, xenophobia, anti-Semitisme atau bentuk-bentuk kebencian lain berdasarkan intoleransi, termasuk intoleransi yang diungkapkan oleh nasionalisme dan etnosentrisme yang agresif, diskriminasi dan permusuhan terhadap kaum minoritas, migran dan orang-orang yang berasal dari imigran” berada di bawah payung istilah *hate speech* (Gagliardone et al., 2014).

(Gagliardone et al., 2014) lebih lanjut menyatakan bahwa ujaran kebencian mengungkapkan “sikap diskriminatif, mengintimidasi, tidak setuju, antagonis, dan atau berprasangka terhadap karakteristik tersebut, yang meliputi gender, ras, agama, etnis, warna kulit, asal kebangsaan, cacat atau orientasi seksual” dan bahwa pidato kebencian adalah dimaksudkan "untuk melukai, tidak memanusiakan, melecehkan,

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.

mengintimidasi, merendahkan, merendahkan dan menjadi korban kelompok sasaran, dan untuk menimbulkan ketidakpekaan dan kebrutalan terhadap mereka".

Ujaran kebencian (*hate speech*) dapat dilakukan melalui berbagai cara seperti orasi kegiatan kampanye, spanduk atau *banner*, media sosial, penyampaian pendapat dimuka umum (demonstrasi), ceramah keagamaan, media masa cetak maupun elektronik, dan sebagainya.

Pasal 28 ayat (2) UU No.11 Tahun 2008 tentang Informasi dan Transaksi Elektronik yang menyebutkan:

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

Istilah tindak pidana penghinaan tercantum di dalam Pasal 310 KUHP ayat (1) berbunyi:

“Barang siapa sengaja merusak kehormatan atau nama baik seseorang dengan jalan menuduh dia melakukan sesuatu perbuatan dengan maksud yang nyata akan tersiarnya tuduhan itu, dihukum karena menista, dengan hukuman penjara selamalamanya sembilan bulan atau denda sebanyak-banyaknya Rp.4500,-.”

Masalah pelanggaran atau kejahatan mencemarkan nama baik orang lain, memfitnah, menista dan perbuatan tidak menyenangkan merupakan suatu perbuatan yang melanggar hukum karena meresahkan dan melanggar hak asasi orang lain. Perbuatan tersebut tidak hanya dapat dilakukan secara langsung dengan kata-kata di muka umum tetapi juga akhir-akhir ini sering dilakukan di dunia maya atau media sosial, karena di dunia maya masyarakat merasakan kebebasan dalam hal berpendapat maupun mengkritik seseorang yang dianggap tidak akan melanggar hukum dan aman karena tidak berkontak fisik langsung dengan orang lain.

2.3 *Abusive language* (Bahasa Kasar)

Indonesia termasuk ke dalam negara dengan pengguna sosial media terbanyak. Orang Indonesia sering menggunakan sosial media untuk berbagai keperluan, seperti mencari dan berbagi informasi, menjalin komunikasi, media hiburan atau sekedar melepaskan perasaan hati (curhat). Sejumlah besar pengguna

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.

media sosial di Indonesia sering mengarah pada komunikasi yang tidak terkendali dan banyak *netizen* yang berkomunikasi dengan menggunakan Bahasa kasar.

Bahasa kasar adalah ekspresi yang berisi kata atau frasa kasar / kotor, baik lisan maupun tulisan (Tuarob & Mitranont, 2017). Menurut (Tuarob & Mitranont, 2017), penyebab digunakannya kata-kata kasar yang tidak terkontrol di media sosial adalah karena tidak adanya alat yang efektif untuk menyaring bahasa kasar di media sosial, kurangnya empati di antara warga negara, dan kurangnya bimbingan orang tua. Bahasa yang kasar di media sosial perlu disaring sehingga tidak ada anak-anak dan remaja yang belajar bahasa kasar dari media sosial yang mereka gunakan (Chen et al., 2012).

Mendeteksi Bahasa kasar di media sosial adalah masalah yang sulit untuk dipecahkan. (Nobata et al., 2016) mengatakan bahwa mendeteksi suatu Bahasa kasar di media sosial tidak bisa hanya menggunakan pencocokan kata. Karena banyak dari netizen biasanya menggunakan ejaan dan tata bahasa kasar yang sangat informal. Terutama dalam teks pendek, mengklasifikasikan teks pendek untuk mendeteksi Bahasa kasar lebih sulit untuk diselesaikan.

Misalnya dalam data twitter, ada banyak netizen yang memposting *tweet* menggunakan singkatan, karena terbatasnya jumlah kata yang diizinkan oleh twitter dalam sebuah postingan. (Hanafiah et al., 2017) mengatakan bahwa beberapa kata non-formal yang sering digunakan oleh orang indonesia adalah : kata-kata yang menunjukkan perasaan, pengulangan karakter untuk menekankan makna, menggunakan Bahasa gaul, dan mengubah huruf vokal menjadi angka.

2.4 Python

Python adalah salah satu Bahasa pemrograman tingkat tinggi yang juga banyak digunakan dalam mengembangkan sebuah aplikasi. Dikembangkan pada tahun 1990 oleh seorang ilmuwan bernama Guido Van Rossum. Meski cukup baik dalam mengembangkan aplikasi, namun python lebih banyak digunakan para penggiat *Data science* atau *Data engineer* dalam mengembangkan algoritma *machine learning*.

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.

Python dirancang untuk memberikan kemudahan yang sangat luar biasa kepada programmer baik dari segi efisiensi waktu, maupun kemudahan dalam pengembangan program dan dalam hal kompatibilitas dengan sistem (Qutsiah, Sophan and Hendrawan, 2016). Beberapa *feature* yang terdapat dalam Bahasa pemrograman Python antara lain.

1. Memiliki library yang luas, dalam distribusi Python telah disediakan modul-modul siap pakai untuk berbagai keperluan.
2. Memiliki tata bahasa yang jernih dan mudah dipelajari.
3. Memiliki aturan layout source code layout source code yang memudahkan pengecekan, pembacaan kembali, dan penulisan ulang source code tersebut.
4. Memiliki modular, mudah dikembangkan dengan menciptakan modul-modul baru, dimana modul-modul tersebut dapat dibangun dengan bahasa Python maupun C/C++

Implementasi python pada penelitian dilakukan pada proses pengolahan data. Seperti proses *training* dan *testing* data, proses *feature selection*, *klasifikasi Random Forest*, dan *Confusion Matrix*.

2.5 Natural Language Processing (NLP)

Natural Language Processing adalah suatu cabang ilmu komputer yang berkembang dari studi Bahasa dan komputasi *linguistic* dalam cabang kecerdasan buatan (*Artificial Intelligence*). *NLP* bertujuan untuk mengembangkan dan merancang sebuah aplikasi yang mampu menjadi penengah antara interaksi manusia dan mesin dengan perangkat lain melalui penggunaan Bahasa alami (Gustejevsky & Stubbs, 2012).

Pada *NLP* informasi yang akan digunakan berisi data-data yang tidak terstruktur. Sehingga diperlukan sebuah proses perubahan bentuk menjadi data yang terstruktur untuk kebutuhan penelitian (*sentiment analysis*, *topic modelling*, dan lain-lain).

Merujuk pada penelitian yang dilakukan oleh (Mujilahwati, 2016), maka pada penelitian ini akan menggunakan beberapa teknik *preprocessing* 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.

2.5.1 Tokenizing

Pemotongan string input berdasarkan tiap kata penyusunnya. Tokenisasi dilakukan untuk memisahkan setiap kata yang ada pada setiap kata yang menyusun sebuah dokumen. Contoh : “Saya sedang belajar data *science*”, setelah dilakukan tokenisasi maka kalimat tersebut akan menjadi “Saya”, “sedang”, “belajar”, “data”, “*science*”.

2.5.2 Case folding

Merupakan salah satu bentuk *text preprocessing* yang paling sederhana dan efektif. *Case folding* berperan dalam mengkonversi huruf-huruf yang berada di dalam 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*.

2.5.3 Cleaning

Tahapan untuk membersihkan dataset dari kata-kata yang tidak diperlukan untuk mengurangi *noise* yang ada pada tahap klasifikasi. Kata atau karakter yang akan dihilangkan pada proses *cleaning* seperti : simbol, angka, *link url*, *hashtag* (#), serta *mention* (@username).

2.5.4 Filtering (Stopword Removal)

Proses memilah data dengan cara mengambil kata-kata penting dari proses *preprocessing* sebelumnya dengan menggunakan algoritma *stoplist* (membuang kata penting) atau *wordlist* (menyimpan kata penting).

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. *Stopword* menghapus kata-kata yang memiliki informasi rendah dari sebuah teks agar kita dapat fokus pada kata-kata penting lainnya.

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.

2.6 Machine Learning

Machine learning adalah salah satu cabang ilmu dari kecerdasan buatan (Artificial Intelligence). *Machine learning* diperkenalkan oleh ilmuwan Matematika seperti Thomas Bayes, Adrien-Marie Legendre, dan Andrey Markov pada tahun 1920-an.

Machine learning umumnya mampu mempelajari pola atau tren tertentu pada suatu data dan dapat memberi pengetahuan yang tidak eksplisit pada pengguna mengenai data. Contoh penerapan *machine learning* seperti deteksi penipuan pada transaksi kartu kredit dan sistem yang memfilter informasi mengenai preferensi baca pengguna (Mitchell, 1997). Salah satu tugas utama dalam *machine learning* adalah menentukan algoritma mana yang tepat digunakan untuk menyelesaikan tugas komputasi tertentu.

Dalam *machine learning*, ada dua tugas umum yang dilakukan, yaitu :

1. *Supervised learning*

Pada *Supervised learning*, *dataset* yang digunakan sudah memiliki label. Label adalah *tag* atau pengenalan dari sebuah data. Misalnya terdapat sepotong buah apel yang memiliki atribut berwarna hijau, berat lebih dari 200 gram, kulitnya keras, berduri, bau yang menyengat, dan isi buahnya manis. Buah yang memiliki karakteristik seperti yang ini dikenali sebagai durian. Maka label dari atribut tersebut adalah durian.

2. *Unsupervised learning*

Pada *unsupervised learning*, *dataset* yang digunakan tidak memiliki label. Model *unsupervised* melakukan belajar sendiri untuk melabeli atau mengelompokkan data. Contoh kasus untuk *unsupervised* adalah dari data 100 pengunjung sebuah *website*, model akan belajar sendiri untuk mengelompokkan pengunjung. Misalnya dikelompokkan berdasarkan waktu kunjungan, lama kunjungan, jumlah klik, dan sebagainya.

3. *Semi-Supervised Learning*

Merupakan gabungan dari *supervised learning* dan *unsupervised learning*. Pada model *semi supervised*, *dataset* untuk pelatihan sebagian memiliki label dan sebagiannya lagi tidak memiliki label. Implementasi dari model

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.

semi supervised yang sering kita gunakan misalnya pada *Google Photos*. Pada *Google Photos*, kita bisa memberi *tag* untuk setiap orang yang terdapat pada sebuah foto. Ketika kita mengunggah foto baru dengan wajah orang yang sudah diberi *tag* sebelumnya, *Google Photos* secara otomatis akan mengenali orang tersebut.

4. *Reinforcement Learning*

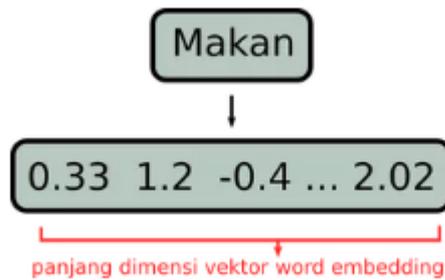
Reinforcement Learning adalah model yang belajar menggunakan sistem *reward* dan *penalties*. Alpha Go adalah contoh terkenal dari *reinforcement learning*. Sebuah program yang dikembangkan *Google Deepmind* untuk memainkan permainan Go, sebuah permainan papan yang berasal dari Cina. Alpha Go mempelajari setiap langkah dalam jutaan permainan Go untuk terus mendapatkan *reward* yaitu memenangkan sebuah permainan.

2.6.1 *Machine Learning Workflow*

Dalam sebuah *machine learning*, ada tahapan-tahapan yang perlu dilalui sebelum *project* bisa diimplementasikan ke tahap produksi. Berikut adalah tahapan-tahapan yang dimaksud menurut buku *Hands on Machine Learning* karya Aurelien Geron :

1. *Exploratory Data Analysis*
Exploratory data analysis atau EDA bertujuan sebagai analisa awal terhadap data dan melihat bagaimana kualitas data.
2. *Data Preprocessing* dan *Data Cleaning*
Data preprocessing dan *cleaning* adalah tahap dimana data diolah lebih lanjut sehingga data siap digunakan dalam pengembangan *machine learning*.
3. *Model Selection*
Pada tahap ini kita mulai memilih model yang akan dipakai serta melakukan optimasi parameter dari model tersebut.
4. *Model Evaluation*
Lalu kita melakukan evaluasi terhadap model dengan melihat performanya terhadap data *testing*.
5. *Deployment*

Kedekatan nilai dibandingkan kata “marah” dan “bahagia”. Ilustrasi dari *word embedding* dapat dilihat pada gambar dibawah ini.



Gambar 2. 2 Ilustrasi Word Embedding

Penelitian ini menggunakan metode *word embedding* yang diberi nama *FastText* yang dibuat oleh Facebook AI Research (FAIR)². *FastText* adalah perpustakaan untuk mempelajari *embeddings* kata dan klasifikasi teks. Model ini memungkinkan untuk membuat pembelajaran tanpa pengawasan atau algoritma pembelajaran terawasi untuk mendapatkan representasi *vector* untuk kata-kata. Facebook menyediakan model pra-pelatihan untuk 294 bahasa, termasuk bahasa Indonesia³. *FastText* menggunakan jaringan saraf untuk penyematan kata.

2.8 Random Forest Classifier

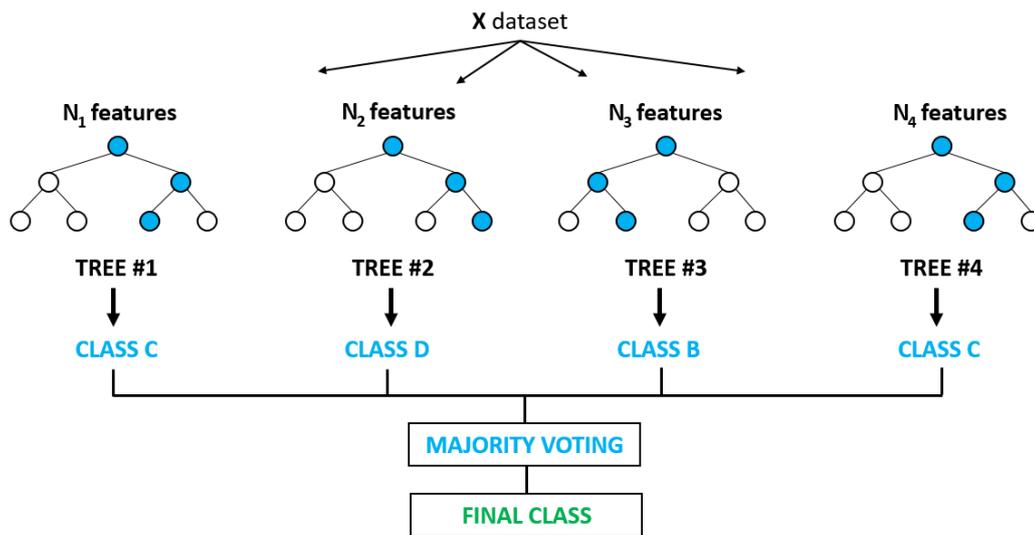
Random Forest merupakan sebuah metode *ensemble*. Metode *ensemble* merupakan cara untuk meningkatkan akurasi metode klasifikasi dengan cara mengkombinasikan metode klasifikasi (Saputra, Wahyu S. et al., 2011). *Random Forest* diawali dengan teknik dasar *data mining* yaitu *decision tree*. Pada *decision tree input* dimasukkan pada bagian atas (*root*) kemudian turun kebagian bawah (*leaf*) untuk menentukan data tersebut termasuk kedalam kelas apa. *Random Forest* adalah pengklasifikasi yang terdiri dari kumpulan pengklasifikasi pohon terstruktur dimana masing-masing pohon melemparkan unit suara untuk kelas paling populer di input x (Breiman, 2001). Dengan kata lain *Random Forest* terdiri dari sekumpulan *decision tree* (pohon keputusan), dimana kumpulan *decision tree* tersebut digunakan untuk mengklasifikasi data ke dalam suatu kelas.

² <https://en.wikipedia.org/wiki/FastText>

³ Sabirin, Dyani. “Facebook Membuat Program AI tersedia dalam 294 Bahasa”.

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. 3 Ilustrasi *Random Forest*

Random forest adalah salah satu metode yang *supervised learning*. *Supervised learning* adalah salah satu tipe pembelajaran mesin di mana kita sudah memiliki variabel masukan (*input*) dan keluaran (*output*). Tujuan dari tipe pembelajaran ini adalah untuk memperkirakan fungsi pemetaannya, sehingga disaat kita memiliki *inputan* baru, mesin dapat memprediksi *output* dari inputan tersebut.

Random Forest merupakan metode klasifikasi yang *supervised*. Sesuai dengan namanya, metode ini menciptakan sebuah hutan (*forest*) dengan sejumlah pohon (*tree*). Secara umum, semakin banyak pohon (*tree*) pada sebuah hutan (*forest*) maka semakin kuat juga hutan tersebut terlihat. Semakin banyak *tree*, maka semakin besar pula akurasi yang didapatkan (Polamuri et al., 2019). *Decision tree* menggunakan *information gain* dan *gini index* untuk perhitungan dalam menentukan *root node* dan *rule*. Sama halnya dengan *Random Forest* yang akan menggunakan *information gain* dan *gini index* untuk perhitungan dalam membangun *tree*, hanya saja *Random Forest* akan membangun lebih dari satu *tree*. Masing-masing *tree* dibangun menggunakan set data dengan atribut yang diambil secara acak dari *data training*. Dengan kata lain setiap *tree* akan bergantung pada nilai dari sampel vektor yang independen dengan distribusi yang sama pada setiap *tree*. Selama proses klasifikasi setiap *tree* akan memberikan *voting* kelas yang paling populer. (Sitepu, 2019)

Hak Cipta Dilindungi Undang-Undang

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

Cara kerja dari *Random Forest* dapat digambarkan dengan kasus nyata dalam kehidupan sehari-hari yang dianalogikan seperti kasus berikut ini (Polamuri, 2017). Seorang mahasiswa ingin berlibur ke suatu tempat tapi ia tidak tahu ingin berlibur kemana. Mahasiswa tersebut memutuskan untuk bertanya pada sahabatnya. Kemudian sahabatnya tersebut memberikan beberapa pertanyaan untuk memutuskan rekomendasi tempat yang cocok untuk mahasiswa tersebut. Sejauh ini kasus tersebut menggambarkan metode *Decision Tree*, dimana seorang sahabat dari mahasiswa tersebut menggambarkan *tree* yang dibangun untuk memutuskan rekomendasi tempat. Kemudian mahasiswa tersebut memutuskan untuk bertanya ke teman yang lain dengan jumlah yang banyak dan setiap orang mengajukan pertanyaan-pertanyaan yang acak, setiap teman tentu akan memberikan rekomendasi yang berbeda dan juga ada yang sama. Lalu mahasiswa tersebut akan memutuskan untuk pergi ke tempat yang paling banyak di rekomendasikan dari teman-temannya. Inilah yang menggambarkan proses dan cara kerja metode *Random Forest*, dimana akan membentuk *tree* yang banyak untuk memutuskan suatu keputusan, lalu keputusan akhir akan ditentukan oleh hasil keputusan terbanyak dari *tree* yang telah di bangun. Konsep *voting* ini disebut dengan *majority voting*.

2.8.1 Keuntungan dan Kerugian *Random Forest*

Pada proses pengklasifikasian dengan metode *Random Forest* memiliki keuntungan dan kerugian sebagai berikut:

1. Keuntungan dari *Random Forest* adalah:
 - a. Salah satu algoritma paling akurat yang tersedia. Pada *dataset* yang banyak, menghasilkan klasifikasi yang sangat akurat.
 - b. Pada *database* yang besar berjalan dengan efisien.
 - c. *Random Forest* dapat menangani ribuan variabel input tanpa penghapusan variabel.
 - d. *Random Forest* memberikan perkiraan tentang variabel apa yang penting dalam klasifikasi.
 - e. *Random Forest* menghasilkan perkiraan internal yang tidak biasa dari kesalahan generalisasi ketika pembangunan hutan berlangsung.

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.

- f. *Random Forest* memiliki metode yang efektif untuk memperkirakan data yang hilang dan mempertahankan akurasi ketika sebagian besar data hilang.
 - g. *Random Forest* merupakan metode yang menyeimbangkan kesalahan dalam kumpulan data populasi kelas yang tidak seimbang.
 - h. Hutan yang dihasilkan dapat disimpan untuk digunakan dimasa depan pada data lain.
 - i. Perhitungan *Prototype* yang memberikan informasi tentang hubungan antara variabel dan klasifikasi.
 - j. Perhitungan *proximities* antara pasangan kasus yang dapat digunakan dalam pengelompokkan, mencari *outlier*, atau skala memberikan pandangan yang sangat menarik dari sampel data.
2. Kerugian dari *Random Forest* adalah:
- a. *Random Forest* telah diamati secara berlebihan untuk beberapa *dataset* dengan cara pengklasifikasian yang acak.
 - b. Pada data variabel dengan jumlah tingkat yang berbeda, mendukung atribut tersebut dengan banyak level. Maka dari itu, nilai terpenting dari *variable* hutan acak tidak dapat diandalkan untuk jenis data ini.

2.8.2 Parameter *Random Forest*

Random Forest memiliki banyak parameter yang digunakan dalam membuat sebuah pohon acak. Namun, terdapat beberapa parameter yang paling penting dalam prediksi fitur. Parameter atau atribut tersebut dapat membantu dalam menangani terjadinya *overfitting*.

a. *n_estimators*

Jumlah pohon yang ada pada hutan. Semakin banyak jumlah pohon yang dibuat maka semakin tinggi pula akurasi yang dihasilkan. Nilai *n_estimators* dapat diubah dari 10 sampai 100.

b. *Max_depth*

Kedalaman pohon bervariasi tergantung pada ukuran dan sifat dari sampel *dataset*. *Parameter* ini digunakan untuk membatasi ukuran pohon. Proses pembuatan pohon tidak dilanjutkan ketika kedalaman

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.

pohon sama dengan kedalaman maksimal. Jika nilainya diatur ke ‘-1’, *parameter* kedalaman maksimal tidak menempatkan batas pada kedalaman pohon, pohon dengan kedalaman maksimum dihasilkan. Jika nilainya disetel ke ‘1’, pohon dengan satu simpul dihasilkan.

c. *Criterion*

Fungsi untuk mengukur kualitas *split*. Kriteria yang didukung adalah “*gini*” untuk ketidakmurnian Gini dan “*entropy*” untuk perolehan informasi.

d. *Minimal samples split*

Merupakan parameter yang memberi tahu pohon keputusan di hutan acak jumlah pengamatan minimum yang diperlukan di sembarang simpul yang diberikan untuk membaginya. Nilai defaultnya adalah 2. Ini berarti bahwa jika ada simpul terminal yang memiliki lebih dari dua pengamatan dan bukan simpul murni, kami dapat membaginya lebih jauh menjadi beberapa *subnode*. Dengan meningkatkan nilai *min_samples_split*, kita dapat mengurangi jumlah pemisahan yang terjadi di pohon keputusan dan membantu mencegah model mengalami *overfitting*.

e. *Max features*

Jumlah fitur yang perlu dipertimbangkan saat mencari pemisahan terbaik. Terdapat beberapa nilai dari *max_features*, yaitu “*auto*”, “*sqrt*”, dan “*log2*”. Pencarian *split* tidak berhenti sampai setidaknya satu partisi sampel *node* yang *valid* ditemukan, bahkan jika diperlukan untuk secara efektif memeriksa lebih dari fitur *max_features*.

2.8.3 Algoritma Random Forest

1. Diberikan jumlah kasus pelatihan dinotasikan dengan N. dan jumlah variable dalam classifier dinotasikan dengan M.
2. Jumlah m’ untuk variabel input yang digunakan untuk menentukan keputusan pada suatu *node* dari *tree*. m’ harus jauh lebih sedikit dibanding M.
3. Pilih satu set pelatihan untuk tree ini dengan memilih N kali dengan penggantian dari semua N kasus pelatihan yang tersedia. Gunakan sisa

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.

kasus untuk memperkirakan *estimate error* dari *tree*, dengan memprediksi kelas mereka.

4. Untuk setiap *node* dari pohon, secara acak memilih variabel m' yang menjadi dasar keputusan di *node* tersebut. Kemudian menghitung split terbaik berdasarkan variabel m' dalam training set.
5. Setiap pohon telah dibangun dan tidak dipangkas.

Random forest adalah *classifier* yang terdiri dari kumpulan pengklasifikasi pohon terstruktur $\{h(x, \theta_k), \theta_k = 1 \dots k\}$. Dimana k secara independen, identik dan distribusikan pohon acak dan setiap pohon melemparkan suara unit untuk klasifikasi akhir input x . Seperti 'CART', *Random forest* menggunakan Gini Index untuk menentukan kelas akhir di setiap pohon. Pada final class dari setiap pohon dikumpulkan dan dipilih oleh nilai-nilai weight untuk membangun classifier akhir.

Random forest menggunakan Gini Index diambil dari sistem CART untuk membangun *decision trees*. Gini Index pada *node impurity* adalah ukuran yang paling umum dipilih untuk masalah klasifikasi.

Gini Index, $Gini(T)$ didefinisikan sebagai :

$$Gini(T) = 1 - \sum_{j=1}^n (p_j)^2$$

Jika T dataset dibagi menjadi dua himpunan bagian T_1 dan T_2 dengan ukuran N_1 dan N_2 masing-masing, indeks pada split data berisi contoh-contoh dari kelas n , indeks gini (T) didefinisikan sebagai :

$$Gini_{split}(T) = \frac{N_1}{N} gini(T_1) + \frac{N_2}{N} gini(T_2)$$

2.3.4 Kerja Algoritma Random Forest

1. Sebuah *seed* acak dipilih yang diambil keluar secara acak dari kumpulan sampel dari X *training dataset* dengan tetap mempertahankan distribusi *class*.
2. Dengan *dataset* yang dipilih ini, satu set acak atribut dari set data asli yang dipilih berdasarkan pada nilai-nilai yang ditetapkan pengguna. Semua variabel input tidak dianggap karena perhitungan yang sangat besar dan kemungkinan besar terjadi 'over fitting'.

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. Dalam *dataset* dimana M adalah jumlah total masukan atribut dalam *dataset*, hanya atribut R yang dipilih secara acak untuk setiap *tree* dimana $R < M$.
4. Atribut dari set ini memungkinkan menciptakan *best split* menggunakan Gini Index untuk mengembangkan model *decision tree*. Proses pengulangan untuk setiap cabang sampai kondisi penghentian menyatakan bahwa daun dari cabang sampai kondisi penghentian menyatakan bahwa daun *node* terlalu kecil untuk di *split*.

2.9 Feature Selection

Feature Selection atau seleksi fitur merupakan sebuah proses dalam *machine learning* yang menggunakan sekumpulan fitur yang dimiliki oleh data untuk digunakan dalam proses algoritma. (Rokach dkk., 2006) mengatakan bahwa *feature selection* telah menjadi bidang penelitian aktif yang digunakan dalam pengenalan pola, statistik, dan data *mining*.

Feature selection merupakan salah satu faktor penting yang dapat mempengaruhi tingkat akurasi dari klasifikasi. Karena 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 (Rokach dkk., 2006) :

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

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.

atau atribut yang mengganggu dan dapat mengurangi akurasi. *Noisy feature* atau fitur yang tidak dipakai tersebut harus dihapus agar dapat meningkatkan akurasi yang diperoleh. Selain itu, fitur atau atribut yang terlalu banyak dapat memperlambat proses komputasi.

Pada penelitian ini, menggunakan cara manual untuk melakukan tahapan seleksi fitur. Yaitu dengan melakukan beberapa percobaan dengan mengkombinasikan dataset, *word embedding*, *text preprocessing*, dan model *random forest*. Hasil dari percobaan tersebut akan menentukan *feature* mana yang memiliki akurasi terbaik.

2.10 Confusion Matrix

Pada tahap klasifikasi tentunya diharapkan agar semua *dataset* yang ada dapat melakukan proses klasifikasi dengan tepat. Oleh karena itu diperlukan sebuah cara untuk mengukur hasil kerja klasifikasi, hal tersebut dapat dilakukan menggunakan *Confusion Matrix*.

Berikut contoh perhitungan menggunakan *confusion matrix* :

Tabel 2. 1 Confusion Matrix

Predicted Values	1 (Positive)	Actual Values	
		1 (Positive)	0 (Negative)
0 (Negative)	TP (True Positive)	FP (False Positive)	
	FN (False Negative)	TN (True Negative)	

Rumus untuk menghitung akurasi *Confusion Matrix* sebagai berikut:

$$\begin{aligned}
 accuracy &= \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi keseluruhan}} \\
 &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\%
 \end{aligned}$$

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.

2.11 Penelitian Terkait

Tabel 2.2 berikut ini merupakan pembahasan yang berisi penelitian terdahulu dari jurnal-jurnal yang berkaitan dengan penelitian tugas akhir ini.

Tabel 2. 2 Penelitian Terkait

No	Author	Tahun	Judul Penelitian	Hasil
1	Muftia Chalida	2019	Analisis Sentimen Ujaran Kebencian Pemilihan Presiden 2019 Menggunakan Algoritma <i>Naïve Bayes Classifier</i>	Hasil klasifikasi pada <i>sentiment yang irrelevant</i> terdapat 11,3% dengan 573 data, <i>sentiment negatif</i> 35,4% dengan 1786 data, <i>sentiment netral</i> 26,7% dengan 1343 data, kecenderungan pada <i>sentiment negatif</i> dengan nilai terbesar yaitu 35,4%.
2	Yusra, dkk,	2016	Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode <i>Naïve Bayes Classifier</i> dan <i>K-Nearest Neighbor</i>	Akurasi yang didapatkan dengan menggunakan metode KNN sebesar 85.00% dan 87.00% pada metode <i>Naïve Bayes Classifier</i> .
3	Nurjanah, dkk,	2017	Analisis Sentimen Terhadap Tayang Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter	Akurasi yang didapatkan yaitu <i>precision</i> 72,28%, <i>recall</i> 100%, dan <i>f-measure</i> 83,91%.
4	Suprpto	2017	Sistem Klasifikasi Opini Pengguna Maskapai Penerbangan di Indonesia Pada Jejaing Sosial Twitter menggunakan Metode <i>K- Nearest Neighbor</i>	Hasil yang diperoleh yaitu <i>sentiment</i> positif sebesar 86,674% dan 93,345% untuk <i>sentiment</i> negatif.
5	Buntoro	2017	Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter	Akurasi yang diperoleh adalah sebesar 95% untuk <i>Naïve Bayes Classifier</i> dan SVM sebesar 90%.
6	Imelda A.Muis,	2015	Penerapan Metode <i>Support vector</i>	Hasil akurasi sebesar 97,54% sebelum dilakukan

Hak Cipta Dilindungi Undang-Undang

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

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

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

	Muhammad Affandes, M.T		<i>machine</i> (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi <i>Tweet</i>	pemilihan <i>feature</i> , 99.12% setelah pemilihan <i>feature</i> .
7	Umi Rofiqoh, Rizal Setya Perdana, M. Ali Fauzi	2017	Analisa Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode <i>Support vector machine</i> dan <i>Lexicon Based Features</i>	Akurasi dengan menggunakan SVM dan <i>Lexicon Based Features</i> adalah 79.00%, dan untuk akurasi dengan menggunakan SVM saja sebesar 84%.
8	Ghulam Asrofi Buntoro	2016	Analisa Sentimen <i>Hatespeech</i> pada Twitter dengan Metode <i>Naive Bayes Classifier</i> dan <i>Support vector machine</i>	Nilai rata-rata akurasi tertinggi dicapai 66,6% menggunakan metode SVM.
9	Aji Primajaya dan Betha Nurina Sari	2018	<i>Random Forest for Prediction of Precipitation</i>	Dengan menggunakan Teknik 10-fold cross validation diperoleh akurasi 71,09%, dan dengan menggunakan seluruh data sebesar 99,45%.
10	Yusuf Sulistyono Nugroho	2017	Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode <i>Random Forest</i>	Pada penelitian ini, sistem pengklasifikasi menggunakan metode <i>Random Forest</i> telah diterapkan pada data tingkat penerimaan konsumen terhadap mobil dan dapat menghasilkan sebuah pohon keputusan. Pohon Keputusan menunjukkan bahwa atribut Estimated Safety menempati sebagai root node sehingga dapat diketahui bahwa atribut ini merupakan atribut paling signifikan bagi konsumen dalam menentukan jenis mobil yang akan dipilih kemudian diikuti oleh atribut-atribut lainnya yang menempati sebagai internal

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.

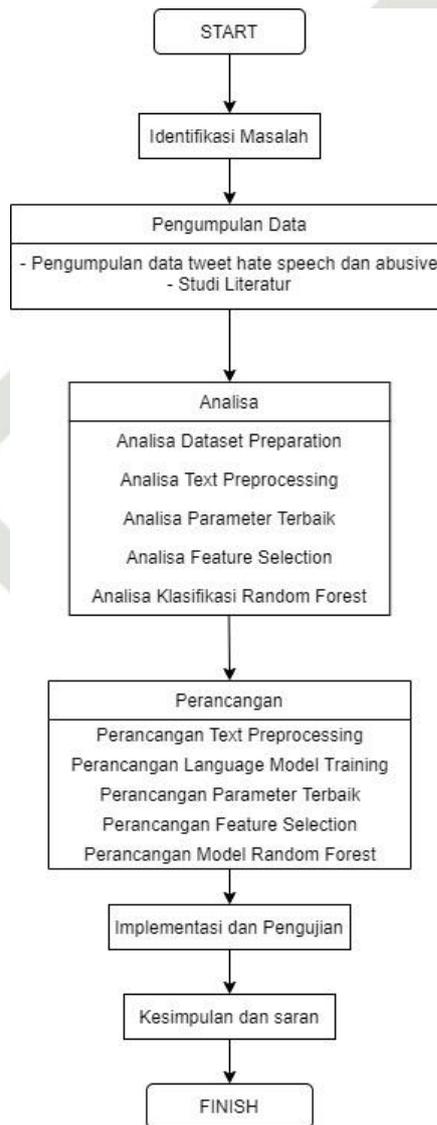
				node. Atribut yang menjadi root node dapat dijadikan sebagai pertimbangan produsen mobil dalam proses produksi mobil di masa mendatang.
11	Nariswari Karina Dewi, dkk.	2018	PENERAPAN METODE <i>RANDOM FOREST</i> DALAM <i>DRIVER ANALYSIS</i>	<i>Random Forest</i> dengan rukunan contoh peubah penjelas sebesar 4 dan ukuran <i>Random Forest</i> lebih dari 500 diperoleh hasil prediksi yang tinggi dan stabil, yaitu dengan tingkat kesalahan klasifikasi antara 33% dan 35.5% dengan nilai rata-rata 34.5%.
12	Meri Febriyani	2018	ANALISIS FAKTOR PENYEBAB PELAKU MELAKUKAN UJARAN KEBENCIAN (<i>HATE SPEECH</i>) DALAM MEDIA SOSIAL	Terdapat beberapa faktor penyebab pelaku melakukan ujaran kebencian, yaitu faktor dari dalam diri individu, keadaan psikologis dan kejiwaan, faktor lingkungan individu, faktor kurangnya kontrol sosial, kepentingan masyarakat, ketidaktahuan masyarakat, serta faktor fasilitas dan kemajuan teknologi.
13	Adhi Viky Sudiantoro dan Eri Zuliarso	2018	ANALISIS SENTIMEN TWITTER MENGGUNAKAN <i>TEXT MINING</i> DENGAN ALGORITMA <i>NAÏVE BAYES CLASSIFIER</i>	Hasil akurasi yang diperoleh sebesar 84.00%, dan <i>Naïve Bayes Classifier</i> sangat efektif dalam melakukan klasifikasi <i>tweet sentiment</i> ini.

Hak Cipta Dilindungi Undang-Undang

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

BAB III METODOLOGI PENELITIAN

Metodologi penelitian merupakan pedoman atau tahapan dalam melakukan penelitian. Metodologi penelitian ini dilakukan bertujuan untuk mendapatkan hasil yang sesuai dengan yang diinginkan. Berikut tahapan-tahapan dalam penelitian :



Gambar 3. 1 Tahapan Metodologi Penelitian

Hak Cipta Dilindungi Undang-Undang

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

3.1 Identifikasi Masalah

Tahap ini merupakan tahapan awal dari metodologi penelitian untuk menentukan latar belakang dan rumusan masalah serta tujuan dari penelitian ini. Selain itu pada tahap ini juga menentukan batasan-batasan pada penelitian yang dilakukan. Adapun hal-hal yang dilakukan pada tahap identifikasi masalah yaitu sebagai berikut:

1. Latar Belakang

Latar belakang adalah dasar yang membuat penelitian ini menjadi perlu dilakukan, dalam penelitian ini fenomena *hate speech* pada media sosial dan kecenderungan pengguna sosial dalam menyampaikan komentar berupa ujaran kebencian, hinaan, cacian, dan lain-lain.

2. Tujuan Penelitian

Tujuan penelitian meliputi hal-hal yang ingin dicapai dalam penelitian. Dalam penelitian ini tujuan utamanya yaitu untuk mengklasifikasikan *tweet* apakah tergolong ke dalam *hate speech* atau *abusive* dan level *hate speech* dari *tweet* tersebut apakah kuat, sedang, atau lemah.

3. Rumusan Masalah

Rumusan masalah adalah masalah-masalah yang harus diselesaikan dalam penelitian, contohnya metode apa yang digunakan, dan bagaimana memperoleh data.

4. Batasan Masalah

Batasan masalah menjadi salah satu yang harus dilakukan pada tahap identifikasi agar penelitian menjadi lebih fokus sehingga tujuan penelitian dapat tercapai.

3.2 Pengumpulan Data

Pada tahap ini akan menjelaskan tentang bagaimana proses pengumpulan data dilakukan dan referensi apa saja yang digunakan dalam pengumpulan data. Penjelasan lengkapnya sebagai berikut :

3.2.1 Pengumpulan Data *tweet Hate Speech* dan *Abusive*

Data yang diperlukan pada penelitian ini adalah data twitter berupa *tweet* yang memiliki kemungkinan termasuk ke dalam *tweet hate speech* dan *abusive*.

Proses pengumpulan data tidak dilakukan dengan cara memilih dan memberi label pada data *tweet* secara manual, karena pada penelitian sebelumnya yang dilakukan oleh (Ibrohim & Budi, 2019) tentang *hate speech* dan *abusive language* telah melakukan *crawling* data *tweet hate speech* dan *abusive*.

(Ibrohim & Budi, 2019) mempersilakan siapa saja yang ingin melakukan penelitian dengan topik yang berhubungan dan memerlukan dataset tentang *hate speech* dan *abusive* dapat menggunakan dataset yang sama secara gratis (*Free*). Oleh karena itu peneliti meminta izin kepada (Ibrohim & Budi, 2019) untuk menggunakan dataset beliau dengan cara menghubungi melalui email, dan beliau mengizinkan serta memberikan link untuk mengunduh dataset tersebut melalui *email*.

Dalam penelitian ini, menggunakan *dataset* kebencian dan penyalahgunaan bahasa Twitter dari beberapa penelitian sebelumnya yang terdiri dari (Alfina dkk., 2017, 2018), (Putri, 2018), dan (Ibrohim & Budi, 2019). Selain menggunakan *dataset* Twitter dari penelitian sebelumnya, (Ibrohim & Budi, 2019) juga merangkak *tweet* untuk memperkaya *dataset* sehingga dapat mencakup jenis penulisan pidato kebencian dan bahasa kasar yang mungkin belum ada dalam data dari penelitian sebelumnya.

Untuk proses anotasi, pada penelitian sebelumnya membangun sistem anotasi berbasis web untuk memudahkan dalam membuat anotasi data sehingga dapat mempercepat proses anotasi dan meminimalkan kesalahan anotasi. Kami juga melakukan anotasi standar emas untuk menguji apakah bahasa sudah memahami tugas atau tidak. Dalam penelitian ini, kami melakukan diskusi dan konsultasi dengan ahli Bahasa (Nurasijah, 2018) untuk mendapatkan pedoman anotasi yang valid dan anotasi standar emas. Data *Twitter* yang digunakan untuk standar emas berasal dari penelitian sebelumnya (Alfina dkk., 2017; Ibrohim & Budi, 2018) dan buku pedoman ucapan kebencian (Komnas HAM, 2015)

Pada tahap anotasi pertama, berhasil mengumpulkan 16.500 *tweet* dari proses perayapan dan penelitian sebelumnya (Alfina dkk., 2017, 2018; Putri, 2018). Dari fase ini, kita mendapatkan 11.292 (68,44% total *tweet* yang dijelaskan dalam fase pertama) yang terdiri dari 6.187 *tweet* ucapan tidak kebencian dan 5.105 *tweet*

Hak Cipta Dilindungi Undang-Undang

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

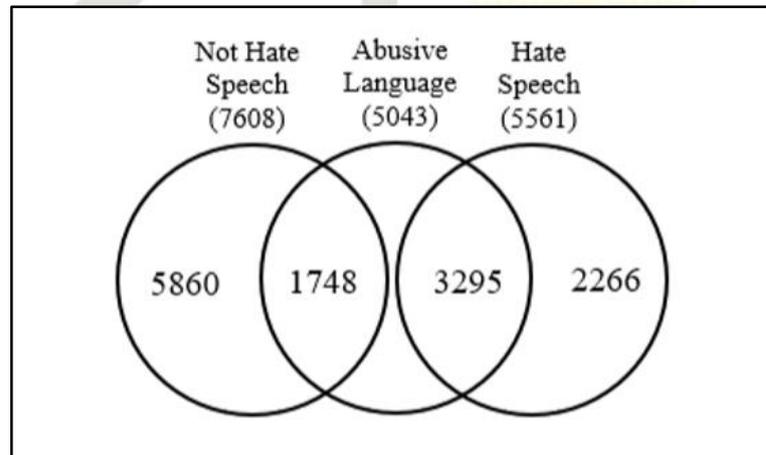
Hak Cipta Dilindungi Undang-Undang

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.

ucapan benci yang memiliki perjanjian 100% (*dataset* dapat diandalkan). Menurut (McHugh, 2012), jumlah persentase ini dari *dataset* yang dapat diandalkan (data dapat digunakan untuk eksperimen penelitian) menunjukkan bahwa penilaian tersebut menetapkan tingkat persetujuan yang meningkat.

Selanjutnya, dalam fase anotasi kedua, tercatat 5.700 *tweet* ucapan kebencian (5.105 *tweet* dari anotasi fase pertama dan 595 *tweet* dari (Ibrohim & Budi, 2019)).

Dari proses anotasi dua fase ini, berhasil mendapatkan 13.169 *tweet* yang telah digunakan untuk eksperimen penelitian yang terdiri dari 7.608 *tweets* ucapan tidak membenci (6.187 *tweets* dari anotasi tahap pertama dan 1.421 *tweets* dari (Ibrohim & Budi, 2018)) dan 5.561 *tweet* ucapan kebencian . Distribusi Bahasa kasar ke *tweets* ucapan tidak membenci pidato dan *tweets* ucapan benci dari *tweet* yang dikumpulkan dapat dilihat pada Gambar 3.2. Dari Gambar 3.2, kita dapat melihat bahwa tidak semua pidato kebencian adalah Bahasa yang kasar. Sebaliknya, bahasa yang kasar juga tidak harus berupa pidato kebencian.



Gambar 3. 2 Distribusi *tweet* Abusive ke *tweet* HS dan Not HS

3.2.2 Studi Literatur

Merupakan tahap untuk memperoleh semua informasi seperti mencari referensi dari jurnal, buku, *paper* internasional, *youtube* dan referensi lainnya yang berhubungan dengan penelitian ini. Referensi yang digunakan adalah referensi yang

berhubungan dengan penelitian, seperti teori-teori seputar penelitian yang pernah dilakukan sebelumnya.

3.3 Analisa

Setelah melakukan identifikasi masalah dan pengumpulan data, proses selanjutnya yaitu analisa. Analisa merupakan tahapan untuk mempelajari dan melakukan evaluasi terhadap suatu permasalahan serta bertujuan untuk mengetahui gambaran jelas mengenai penelitian yang dilakukan. Tahapan analisa yang dilakukan sebagai berikut :

3.3.1 Analisa Dataset Preparation

Awalnya *dataset* yang diberikan oleh (Ibrohim & Budi, 2019) terdiri dari 4 label, dimana ada label untuk *hate speech*, label target *hate speech* yang akan dicapai, label level dari *hate speech*, dan label *abusive*. Namun pada penelitian ini label-label tersebut dibagi menjadi tiga label yaitu label *hate speech*, *abusive*, dan *leve hate speech*. Untuk kelas target *hate speech* yang akan dicapai tidak digunakan dalam penelitian ini.

Pada tahap ini peneliti akan melakukan proses persiapan terhadap *dataset* yang sudah ada, yaitu dengan membagi *dataset* menjadi dua bagian. Data untuk *training* dan *testing*. Agar pada saat proses produksi *dataset* sudah dalam keadaan tetap dan tidak berubah-ubah.

3.3.2 Analisa Text Preprocessing

Kemudian tahap *text preprocessing*. Dimana data dan model yang sudah dibuat pada tahap sebelumnya akan digunakan pada tahap ini. Dengan menggunakan data dan model yang telah dibuat, *dataset* akan di proses lagi untuk mendapatkan *dataset* yang bersih untuk tahap produksi.

Tahapan *text preprocessing* yang dilakukan yaitu tokenisasi, *filtering stopword*, *case folding*, dan *cleaning*. Tokenisasi merupakan tahap awal dari *text preprocessing*, dimana setiap *dataset* yang terdiri dari sebuah *sentence* atau kalimat, agar dapat diproses dalam model *machine learning* harus ditokenisasi terlebih dahulu. Yaitu dengan memisahkan setiap kata yang ada dalam suatu kalimat.

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.

Filtering stopword, mengambil kata-kata yang penting dan membuang kata-kata yang tidak penting untuk penelitian. Kemudian dilakukan *case folding* yaitu dengan mengubah semua huruf yang ada dalam dataset menjadi huruf kecil, hal ini memiliki pengaruh besar untuk tahap produksi. Dan yang terakhir yaitu *cleaning*. Untuk penelitian yang menggunakan kalimat sentimen, *cleaning* sangat berpengaruh dalam proses produksi. Karena sebuah kalimat sentimen umumnya mengandung karakter-karakter khusus yang dapat memperkuat kalimat tersebut termasuk ke dalam sentimen positif atau negatif.

3.3.3 Analisa Language Model Training

Merupakan proses untuk membuat model yang akan digunakan dalam proses produksi. Model dibuat berdasarkan *dataset* yang telah dibagi sebelumnya. Untuk pembuatan *language model training* menggunakan library dari Facebook yaitu *FastText*. *FastText* adalah pengembangan dari library *Word2Vec* yang lebih dulu dikenal sebagai library untuk *word embedding*.

⁴*FastText* memiliki keunggulan dibandingkan *Word2Vec*. Salah satunya adalah kemampuan *FastText* dalam menangani kata yang tidak pernah ditemui sebelumnya (*Out Of Vocabulary Word* atau dikenal dengan *OOV*). Misalnya kata-kata yang tidak baku seperti “Pengoptimalisasian” tetap akan diperoleh *vectornya*.

3.3.4 Analisa Parameter Terbaik

Dalam menemukan parameter terbaik yang nantinya akan digunakan untuk proses *feature selection*, peneliti melakukan beberapa kombinasi parameter *Random Forest* yang akan digunakan. Setelah mendapatkan parameter dengan nilai terbaik, maka nilai dari parameter tersebut yang akan digunakan dalam proses *feature selection*.

Parameter yang digunakan berdasarkan parameter yang dimiliki oleh *Random Forest*, yaitu *n_estimators* (banyaknya pohon), *max_depth* (kedalaman pohon), *criterion* (fungsi untuk mengukur kualitas split), *minimal_samples_split*

<https://structilmy.com/2019/04/word-embedding-bahasa-indonesia-menggunakan-FastText-part->

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.

(jumlah sampel minimum untuk membelah simpul), dan `max_features` (jumlah fitur yang dipertimbangkan saat mencari pemisahan terbaik).

Pada saat proses *training*, masing-masing parameter dilakukan beberapa percobaan untuk mendapatkan nilai terbaik pada setiap parameter. Setelah mendapatkan parameter dengan nilai terbaik, maka nilai parameter tersebut akan digunakan untuk proses selanjutnya.

3.3.5 Analisa Feature Selection

Pada tahap ini akan dilakukan proses seleksi fitur terhadap data yang akan digunakan nantinya. Seleksi fitur berfungsi untuk menemukan fitur-fitur apa saja yang penting dan kurang penting di dalam data. Sehingga setelah dilakukannya proses seleksi fitur diharapkan hasil dari seleksi fitur adalah fitur-fitur terbaik yang dapat menghasilkan performa terbaik untuk proses klasifikasi.

Proses *feature selection* dimulai dari dataset awal yang masih berupa sebuah kalimat. Untuk dapat diproses oleh model *word embedding*, dataset ditokenisasi terlebih dahulu ke dalam bentuk *word* (kata). Setelah ditokenisasi, data akan diubah menjadi sebuah vektor, dengan panjang dimensi maksimal 128. Sekarang setiap data telah menjadi vektor dengan maksimum dimensi 128.

Setelah vektor data didapatkan, maka data akan melalui tahap *encoding*. Sebuah tahap pemrosesan untuk setiap kalimat yang akan diubah menjadi sebuah kata dan kemudian di *convert* menjadi vektor. Hasil dari proses *encoding* yaitu hasil dari rata-rata nilai untuk setiap vektor dalam setiap kata dan kalimat. Nilai inilah yang akan digunakan untuk proses training.

Proses *training* untuk menemukan fitur terbaik yaitu dengan menggabungkan model *word embedding*, dengan teknik *encoding* (L1, dan L2 *norm*) dengan beberapa teknik *text preprocessing* seperti *case folding*, *stopwords*, dan *filtering (removal punctuation)*.

3.3.6 Analisa Klasifikasi Random Forest

Setelah semua proses dilakukan, tahap akhir yaitu melakukan klasifikasi menggunakan metode *Random Forest*. Proses klasifikasi dilakukan dengan

menggunakan parameter terbaik dan *feature selection* terbaik yang telah diperoleh sebelumnya untuk mendapatkan hasil akurasi Random Forest yang optimal.

3.4 Perancangan

Pada tahap ini aplikasi dirancang berdasarkan hasil analisa yang telah dilakukan pada proses sebelumnya dengan tujuan untuk memudahkan dalam pembuatan model yang terstruktur dengan baik.

Dalam melakukan perancangan, peneliti menggunakan Jupyter notebook sebagai tempat untuk mengimplementasikan kode python. Dalam membuat notebook untuk proses perancangan model, tahapan-tahapan dalam perancangan dibuat ke dalam lima notebook. Dilakukannya pemisahan terhadap setiap proses bertujuan agar tahap demi tahap dalam pemodelan dapat terstruktur dan juga dapat mempermudah perubahan terhadap salah satu notebook apabila terjadi kesalahan dalam pengkodean.

Tahapan dalam merancang model pada penelitian ini terdiri dari Lima tahap yaitu:

1. Perancangan Text Preprocessing
Setelah dilakukan analisa terhadap dataset preparation, maka selanjutnya akan dilakukan proses text preprocessig terhadap data agar data bersih dari *noise* dan siap untuk digunakan pada proses selanjutnya.
2. Perancangan *Language Model Training*
Agar data dapat digunakan dalam proses klasifikasi, maka dilakukan training terhadap data. Karena pada proses klasifikasi data yang diproses adalah data dalam bentuk numerik, oleh karena itu dilakukan *language model training*.
3. Perancangan Parameter Terbaik
Dalam menemukan *feature selection* terbaik, diperlukan parameter-parameter terbaik yang memiliki pengaruh besar pada tahap klasifikasi.
4. Perancangan Feature Selection
Agar proses klasifikasi dapat menghasilkan akurasi yang baik, maka dilakukan proses *feature selection*. Agar fitur-fitur yang tidak

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.

berpengaruh dengan data dapat diseleksi, sehingga tersisa fitur-fitur terbaik yang memiliki peran penting pada data.

5. Perancangan Model Klasifikasi menggunakan Random Forest

Klasifikasi dilakukan untuk mengetahui performa dari random forest terhadap dataset yang ada, serta untuk mengetahui akurasi yang didapatkan yang nantinya akan diuji kembali menggunakan *confusion matrix* untuk memastikan apakah hasil akurasi yang diperoleh sama dengan hasil perhitungan dari *confusion matrix*.

3.5 Implementasi

Implementasi adalah proses penerapan dari hasil perancangan ke dalam sebuah sistem. Proses ini membutuhkan perangkat pendukung berupa perangkat keras dan perangkat lunak.

1. Perangkat keras

Perangkat keras yang akan digunakan dalam penelitian ini yaitu Laptop Lenovo Ideapad 100 dengan *Processor Intel Core i3*, *RAM 6 GB* dengan sistem operasi windows.

2. Perangkat lunak

Perangkat lunak yang akan digunakan dalam penelitian ini yaitu menggunakan Python. Python merupakan Bahasa pemrograman *open source* dan telah menduduki posisi 4 yang paling sering digunakan di seluruh dunia (Triasanti, 2000) dan memiliki library yang luas, dalam distribusi Python telah menyediakan modul-modul siap pakai untuk berbagai keperluan.

3.6 Pengujian

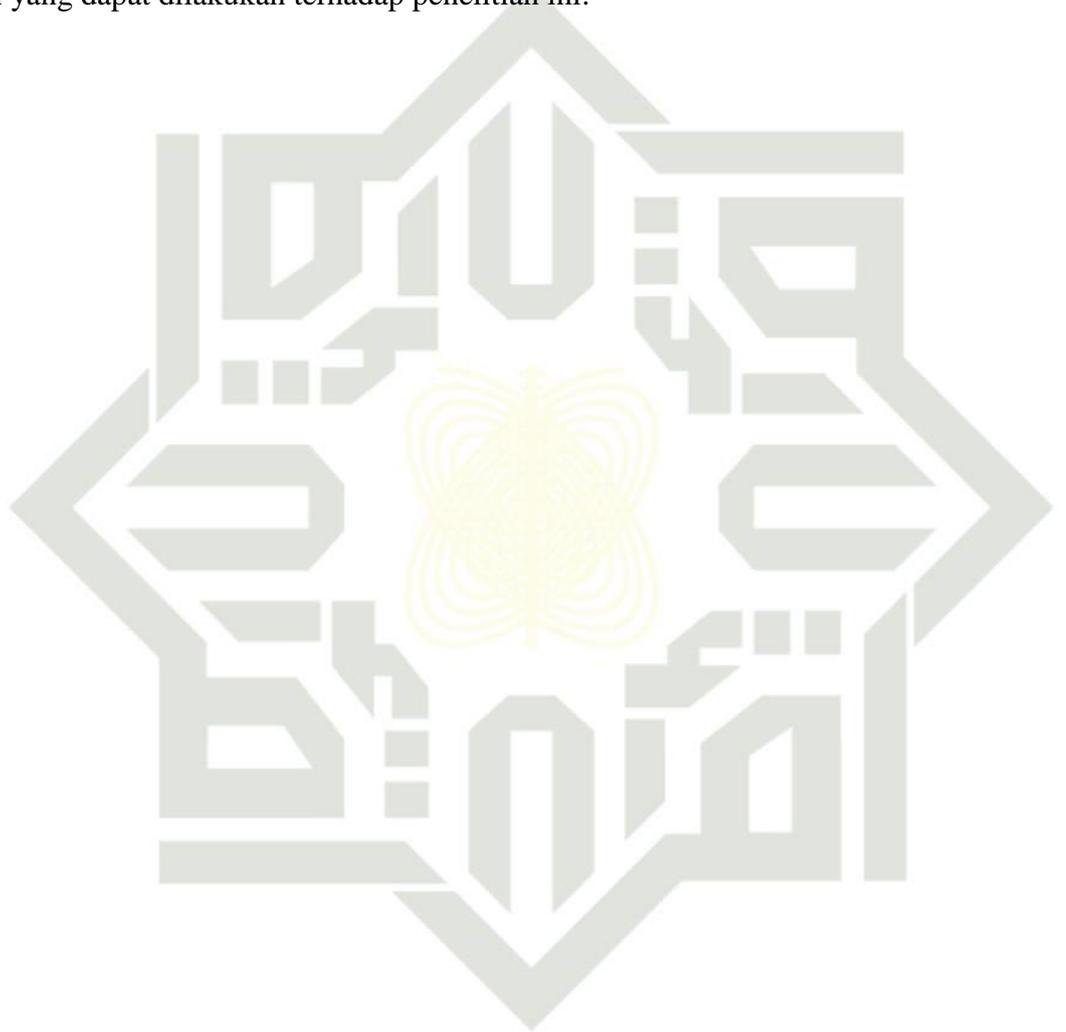
Pengujian adalah tahapan untuk menguji kinerja dari suatu mesin yang telah dibangun. Untuk menghitung akurasi dari model *machine learning* yang telah dibuat dan mengidentifikasi klasifikasi *Random Forest* menggunakan *Confusion matrix*.

Hak Cipta Dilindungi Undang-Undang

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

3.7 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 *Random Forest* berhasil dan mengetahui tingkat akurasi. Pada bagian saran berisi kemungkinan pengembangan yang dapat dilakukan terhadap penelitian ini.



UIN SUSKA RIAU

BAB VI PENUTUP

6.1 Kesimpulan

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan, maka dapat ditarik beberapa kesimpulan sebagai berikut :

1. Metode *Random Forest* dapat diterapkan untuk klasifikasi *multi-class* dan *multi-label hatespeech* dan *abusive* pada twitter Bahasa Indonesia.
2. Kombinasi fitur terbaik yang diperoleh adalah *L1-norm + Stopwords + Case folding* dengan akurasi sebesar **76,07%**.
3. Kombinasi parameter terbaik adalah pada *feature selection* CO-L1 dengan nilai $n_estimators=100$, $max_depth=10$, $criterion="entropy"$, $min_samples_split=10$, dan $max_features="auto"$. Dengan hasil akurasi sebesar **76,20%**.
4. Hasil akurasi model *Random Forest* dengan menggunakan fitur terbaik dan parameter terbaik sebesar **75,96%**.

6.2 Saran

Saran yang dapat diberikan untuk penelitian lanjutan terkait dengan penelitian ini adalah :

1. Klasifikasi dapat dilakukan dengan menggunakan teknik word embedding lainnya seperti Word2Vec dan GloVe, untuk menemukan teknik *word embedding* yang terbaik.
2. Pada penelitian selanjutnya dapat menggunakan metode klasifikasi lainnya untuk dapat melihat hasil perbandingan akurasi seperti *Naïve Bayes Classifier*, *Support Vector Machine*, atau bahkan metode *deep learning* seperti *Convolutional Neural Network*.

Hak Cipta Dilindungi Undang-Undang

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

DAFTAR PUSTAKA

- Abdul Rahin Bin Hamdan, Ahmad Johari Bin Hj Sihes, Jamaluddin Bin Ramli, Rosliza Binti Hamzah, Bernama, Ezaleila Mustafa, S., Hamid, S. A., Ibrahim, J., Nasrullah, R., Amir, O. A., Saidi, F. S., Adrianto, m, W., Mohd Hilmi Mohd Hashim, Muhd Zulkifli Ismail, Nik Farhan Mustafa, Mohamad, H., Paruzzaman, N. A., Does, H. O. W., Apply, I. T., ... Sonenthal, A. (2018). Media Sosial. *Www.Moe.Gov.My*.
- Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2017). Hate speech detection in the indonesian language: A dataset and preliminary study. *2017 International Conference on Advanced Computer Science and Information Systems (ICACISIS)*, 233–238.
- Alfina, I., Pratiwi, S. H., Budi, I., Mulia, R., & Ekanata, Y. (2018). Detecting hate speech against religion in the indonesian language. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*.
- Breiman, L. (2001). Documentation for R package randomForest. *Machine Learning*. <https://doi.org/10.1023/A:1010933404324>
- Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012). Detecting offensive language in social media to protect adolescent online safety. *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, 71–80.
- Huuzi, M. A., & Yuniarti, A. (2018). Ensemble method for indonesian twitter hate speech detection. *Indonesian Journal of Electrical Engineering and Computer Science*, 11(1), 294–299. <https://doi.org/10.11591/ijeecs.v11.i1.pp294-299>
- Gagliardone, I., Patel, A., & Pohjonen, M. (2014). *Mapping and analysing hate speech online: Opportunities and challenges for Ethiopia*.
- Hanafiah, N., Kevin, A., Sutanto, C., Arifin, Y., Hartanto, J., & others. (2017). Text normalization algorithm on twitter in complaint category. *Procedia Computer Science*, 116, 20–26.
- Haryanto, A. T. (2018). 130 Juta Orang Indonesia Tercatat Aktif di Medsos. 12

Hak Cipta Dilindungi Undang-Undang

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

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.

Maret.

- Haryanto, D., & Argadila, D. (2019). Jurnal Teknik Informatika Juni 2019. *Jurnal Teknik Informatika*.
- Hidayatullah, A. F., Fadila, A. A., Juwairi, K. P., & Nayoan, R. A. (2019). Identifikasi Konten Kasar Pada Tweet Bahasa Indonesia. *Jurnal Linguistik Komputasional (JLK)*, 2(1), 1. <https://doi.org/10.26418/jlk.v2i1.15>
- Ibrohim, M. O., & Budi, I. (2018). A dataset and preliminaries study for abusive language detection in indonesian social media. *Procedia Computer Science*, 135, 222–229.
- Ibrohim, M. O., & Budi, I. (2019). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. *ALW3: 3rd Workshop on Abusive Language Online*.
- Katadata.co.id. (2019). Berapa Pengguna Media Sosial Indonesia? *Katadata.Co.Id*.
- Komnas HAM. (2015). Buku Saku Penanganan Ujaran Kebencian (Hate Speech). *Jakarta: KOMNAS HAM*, 10.
- Lewis, M. (2013). A Cognitive Linguistics Overview of Offense and Hate Speech. *SSRN Electronic Journal*, 1–25. <https://doi.org/10.2139/ssrn.2205178>
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica: Biochemia Medica*, 22(3), 276–282.
- Mitchell. (1997). *Cyberporn: bisnis atau kriminal*. Elex Media Komputindo.
- Mujilawati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*, 18–19.
- Neshkovska, S., & Trajkova, Z. (2018). *The Essential of Hate Speech*.
- Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive language detection in online user content. *Proceedings of the 25th International Conference on World Wide Web*, 145–153.
- Nurasijah, M. (2018). *Personal Communication*.
- Okky Ibrohim, M., Sazany, E., & Budi, I. (2019). Identify abusive and offensive

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.

language in Indonesian twitter using deep learning approach. *Journal of Physics: Conference Series*, 1196(1), 12041. <https://doi.org/10.1088/1742-6596/1196/1/012041>

Polamuri, S. R., Srinivas, K., & Krishna Mohan, A. (2019). Stock market prices prediction using random forest and extra tree regression. *International Journal of Recent Technology and Engineering*. <https://doi.org/10.35940/ijrte.C4314.098319>

Ristejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. “O’Reilly Media, Inc.”

Putra, A. K. B. A., Fauzi, M. A., Setiawan, B. D., & Setiawati, E. (2018). Identifikasi Ujaran Kebencian Pada Facebook Dengan Metode Ensemble Feature Dan Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(12).

Putri, T. T. A. (2018). Analisis dan deteksi hate speech pada sosial twitter berbahasa Indonesia. *Master’s Thesis, Faculty of Computer Science, Universitas Indonesia*.

Rokach, L., Chizi, B., & Maimon, O. (2006). Feature selection by combining multiple methods. In *Advances in Web Intelligence and Data Mining* (pp. 295–304). Springer.

Saputra, Wahyu S., J., Sujatmika, A. R., & Arifin, A. Z. (2011). Seleksi Fitur Menggunakan Random Forest Dan Neural Network. *The 13 Th Industrial Electronics Seminar 2011 (IES 2011)*,.

Stepu, B. (2019). *Analisis Algoritma Decision Tree dengan Algoritma Random Forest pada Discretize By Frequency*.

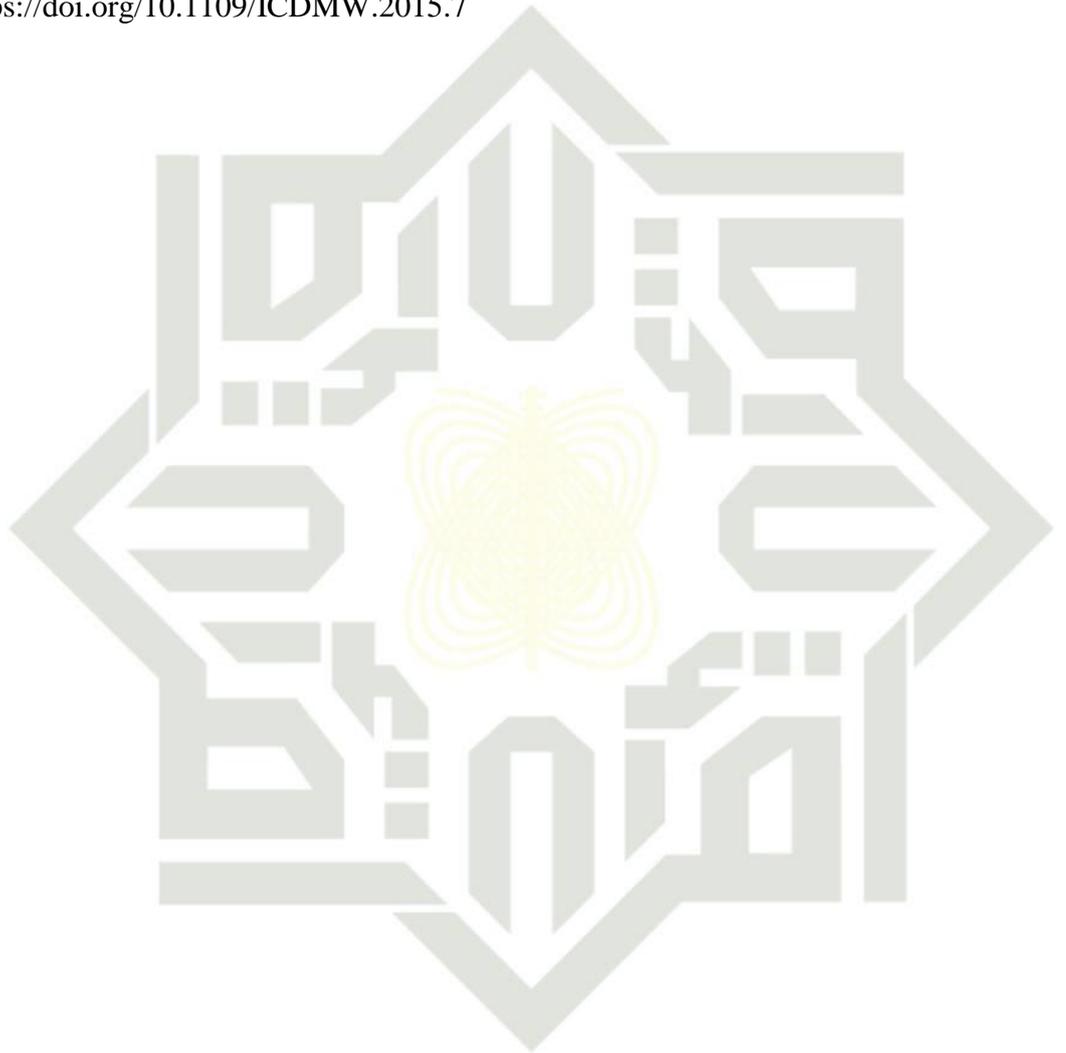
Sudiantoro, A. V., Zuliarso, E., Studi, P., Informatika, T., Informasi, F. T., Stikubank, U., & Mining, T. (2018). *Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma NAÏVE BAYES CLASSIFIER*. 398–401.

Tiasanti, D. (2000). Konsep dasar python. *Python*, 1–6. <https://doi.org/aa>

Tuarob, S., & Mitranont, J. L. (2017). Automatic discovery of abusive Thai

language usages in social networks. *International Conference on Asian Digital Libraries*, 267–278.

Zhan, Y., & Gao, Q. (2016). An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis. *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015*, 1318–1325. <https://doi.org/10.1109/ICDMW.2015.7>



UIN SUSKA RIAU

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

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

LAMPIRAN A

DATASET AWAL DENGAN LABEL

Berikut ini merupakan dataset awal beserta label dan kelas yang digunakan pada penelitian ini :

No	Tweet	HS	Ab usi ve	HS _W eak	HS_ Mod erate	HS_ Stron g	L ev el
1	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
2	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	0
3	41. Kadang aku berfikir, kenapa aku tetap percaya pada Tuhan padahal aku selalu jatuh berkali-kali. Kadang aku merasa Tuhan itu ninggalkan aku sendirian. Ketika orangtuaku berencana berpisah, ketika kakakku lebih memilih jadi Kristen. Ketika aku anak ter	0	0	0	0	0	0
4	USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIT TAPI DILIAT DARI MANA ITU AKU'	0	0	0	0	0	0
5	USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah'	1	1	0	1	0	2
6	USER Ya bani taplak dkk \xf0\x9f\x98\x84\xef\x9f\x98\x84\xef\x9f\x98\x84'	1	1	0	1	0	2
7	deklarasi pilkada 2018 aman dan anti hoax warga dukuh sari jabon	0	0	0	0	0	0
8	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	0
9	Nah admin belanja satu lagi port terbaik nak makan Ais Kepal Milo, Ais Kepal Horlicks atau Cendol Topping kaw kaw. ð??; ; Docket mano tu ? Gerai Rojak Mertuaku - Taipan 2 (depan TWINS BABY & ROMANTIKA / Bank Islam Senawang) ð???	0	0	0	0	0	0
10	USER Enak lg klo smbil ngewe'	0	1	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.

1	Setidaknya gw punya jari tengah buat lu, sebelum gw ukur nyali sama bacot lu \xf0\x9f\x98\x8f'	1	1	1	0	0	1
2	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	1
3	Kalo belajar ekonomi mestinya jago memprivatisasi hati orang. Duh.. ironi USER	0	0	0	0	0	0
4	Aktor huruhara 98 Prabowo S ingin lengserkan pemerintahan Jokowi.... Nyata	1	0	1	0	0	1
5	USER Bu guru enakan jadi jablay atau guru esde sih.\nKayaknya menikmati jadi pecun ini guru.'	1	1	1	0	0	1
6	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 sah. Kafir kyk loe pasti menolak, makanya loe ahlunnar.'	1	1	1	0	0	1
17	Belakangan ini kok fikiran ampas banget ya'	0	0	0	0	0	0
18	Ari sarua beki mah reph monyet\xfb\x9f\x98\x86\xfb\x9f\x98\x86'	0	1	0	0	0	0
19	Jadi cowo itu harus Gantle kalo ga Gantle itu namanya BANCI !'	1	1	0	1	0	2
20	USER Slga mnr bom \xf0\x9f\x98\x82'	0	0	0	0	0	0
21	Asw ya tapi gua jarang ngambek, tacut wkkwkwkwk gua kan bucin.'	0	1	0	0	0	0
22	USER kalo kamu noob pasti peluang disakitin nya lebih gede si'	0	0	0	0	0	0
23	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	0
24	PELAJAR SMA KEC BILAH HILIR DEKLARASI ANTI HOAX & PILKADA 2018 DAMAI	0	0	0	0	0	0
25	Bandara Udara Internasional Kertajati dibangun oleh Gubernur Ahmad Heryawan tanpa gunakan APBN pemerintah pusat. Insy Allah siap beroperasi bulan depan;	0	0	0	0	0	0
26	Siapa pun gubernur dan presidennya, rakyatnya, ya kitaÂ² juga...	0	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 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.

27	Ini si USER kerjaannya delay mulu! Setan!	1	1	0	1	0	2
28	menurutku pintu sorga ada yaitu pintu sorga yang asli dan pintu hatimu modusbanget	0	0	0	0	0	0
29	RT USER USER USER PKI hanya muncul jika jelang pemilihan kepala daerah dan pemilihan presiden krn digoreng sampai gosong.	1	0	0	1	0	2
30	USER USER Itu mah sdh nenek-nenek sy heran sama cebong biasay bohong terus'	1	1	1	0	0	1
31	USER USER USER USER USER USER USER USER USER Islam Nusantara produk JIL, dipasarkan dg gencar oleh antek\xc2\xb2 JIL.\n\nSementara JIL anti Arab, tapi pro Aseng!\n\nKata (alm) KH. Hasyim Muzadi,	1	0	0	1	0	2
32	Dari habis sahur sampe jam 10. Sibayik udah nete 4x. Skg rasanya lemas tak terkira \xf0\x9f\xa4\xa2'	0	0	0	0	0	0
33	USER Gak kak emak mah gak demen ama yang sipit wkwk'	0	0	0	0	0	0
34	RT USER: Kelakuan homok jaman now, ngentot aja sambil live di Blued #gayvid #gvid #lokalhangat URL	0	1	0	0	0	0
35	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	0
36	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	2
37	\xe2\x80\x9dPara Bajingan yang telah melukai Lady harus menerima ganjaran seratus kali lipat!!\xe2\x80\x9d(sanji) ##HBD17thPelautJKT48'	0	1	0	0	0	0
38	ajakan menolak berita hoax dan sukseskan pilkada di wilayah kota keidri:	0	0	0	0	0	0
39	USER USER Presiden RI Jokowi luarBiasa kerja nya tulus IHLAs untuk #RakyatB3rsamaJKW	0	0	0	0	0	0
40	Edan emang goollnya fabiano!!'	0	1	0	0	0	0
41	GAPERNAH MENDALAMI AL-QURAN YA BANG??? PANTESAN MULUTNYA KAYA ORANG KAFIR BEJAT HINA farhatabbaslaw'	1	1	1	0	0	1
42	Sangat terlihat beda sekali Kewibawaannya ; yg sangat berwibawa	0	0	0	0	0	0

	cuma gubernur Indonesia USER ; Massya Allah...						
43	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	2
44	#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	0
45	USER USER Apa iya Islam itu ga boleh dilokalisasi? Lha Kristen aja mas ada Nusantarnya 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	0
46	Ciptakan Situasi Kamtibmas Menjelang Pilkada 2018, Kapolsek Konang Silaturahmi ke Tokoh Masyarakat	0	0	0	0	0	0
47	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	0
48	USER USER USER USER Kata Sembahyang mengadopsi tradisi lokal pada masa kerjaan Hindu.\n\nIstilah\xc2\xa0sembahyang\xc2\xa0pun lahir dari tradisi	0	0	0	0	0	0
49	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	1
50	USER USER BTP the best ,,,'abaikan kaum kampret otak micin ,,,'	1	1	0	1	0	2
51	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	1
52	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!". \xf0\x9f\x98\x8b	1	1	1	0	0	1

Hak Cipta Dilindungi Undang-Undang

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

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.

53	pinokio umur 3 taonan tuh. #2019GantiPresiden	1	0	1	0	0	1
54	USER Cerita silat lae \xf0\x9f\x98\x80 semacam Kho Ping Hoo yang dari Indonesia, tapi Liang Ie Shen penulis dari China'	0	0	0	0	0	0
55	RT USER benci sekali dgn Umat Islam	1	0	0	1	0	2
56	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	0	0	1
57	USER dari pada gua nete'	0	1	0	0	0	0
58	USER Lol bani lowo'	1	1	0	1	0	2
59	Takabur bisa lahir dari siapa saja. Penguasa, rakyat, ulama maupun cendikia. Bahkan iblis mahluq yg alim & telah beribadah beratus2 th dilaknat krn hasud takaburnya. Apalagi hanya seorang ahli ilmu filsafat yg tdk failosof; ; Mari kita demo diri sendi	0	0	0	0	0	0
60	A orang yang suka ke konser-konser, sedikit boros.	0	0	0	0	0	0
61	[SAMPAI KAPAN MAKAN RACUN ?]; _; Banyak ulama yang mengatakan hukum rokok mubah dan ada jugaâ?'	0	0	0	0	0	0
62	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
63	USER udah biasa kali. Gue kpoper tapi kalo ngomonv sama temen2 gue, idol gua banci homo'	1	1	1	0	0	1
64	Saya suka video USER URL Awal Kisah Perseteruan Haji Bolot & Malih Tongtong'	0	0	0	0	0	0
65	Budaya kafir bersungguh, klau hal2 islam tidak diendahkan...	0	0	0	0	0	0
66	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\xab\xe2\x99\xaa\xe2\x99\xaa\xe2\x99\xaa'	0	0	0	0	0	0
67	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

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.

68	USER USER USER USER USER USER USER USER Klw ente atheis..ngapain sibuk dgn agama orang...	0	0	0	0	0	0
69	Lengserkan Jokowi Bangsat!!	1	1	0	0	1	3
70	USER Selamat hari kartini ; ; #2019GANTIPRESIDEN; #2019PRESIDENBARU	0	0	0	0	0	0
71	RT USER: Bajingan Homo!!!\nHati2 terhadap anak2 kita terutama di kamar mandi laki2, jgn biarkan sendiri! URL	1	1	1	0	0	1
72	USER Iya cocot'	1	1	1	0	0	1
73	USER karena budaya indonesia.	0	0	0	0	0	0
74	USER Yang ngajarin solat'a ajah itu orang aseng dia hanya ingin memecah belah islam"	0	0	0	0	0	0
75	USER x Lo ngomong aja kaga bener bangke'	1	1	1	0	0	1
76	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	0
77	tapi masih menikmati kebijakan yang dulu diusulkan PKI.. lha yo opo...	1	0	0	1	0	2
78	Kalo seumpama orang cina di usir paksa dari indonesia gimana	1	0	0	1	0	2
79	USER bunda ngentot yuk'	0	1	0	0	0	0
80	RT USER: china babi... china maling...\n#UsirChina https://t.co/9y4HIWHJp2	1	1	0	0	1	3
81	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	0
82	PROYEK KORUPSI REZIM SBY URL	1	0	1	0	0	1
83	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	0
84	AB: pribadi yg senang berkompetisi dalam hal yang baik dan positif	0	0	0	0	0	0
85	USER USER Terus menurut lo kitab suci agAma apa yg di bilang om USER ; ; #Mendadakkalap	1	0	1	0	0	1
86	USER Suka nutup mata. Kaga ditutup pun udah terlihat tertutup gara2 sipit /?'	0	0	0	0	0	0
87	USER Sy sbg muslim tdk merasa pak USER menistakan kitab suci Alqur'an,	0	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.

	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						
88	USER eh kasar kau kampang'	1	1	1	0	0	1
89	BENTAR LAGI GUA RILIS ALBUM THE ONTA'S WKWKWKWKWK"	0	0	0	0	0	0
90	USER Ada cebong bang dia merasa paling benar'	1	1	1	0	0	1
91	Najis banget gw ngetwit apa barusan wey'	0	1	0	0	0	0
92	Kamu transgender atau gmn anjing :(URL	1	1	1	0	0	1
93	USER awas bintitan lu ngintipin titit kucing'	0	1	0	0	0	0
94	USER USER Bawel, baper, bolot, baik, bodor , gaje wkwk apa ya udah itu aja deh'	0	1	0	0	0	0
95	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 bla... pasti hatimu puas deh ..'	1	1	1	0	0	1
96	Jgn salah target bkn ahok... agenda utama lengserkan Jokowi ..	1	0	0	0	1	3
97	USER USER masa baru kena hestek aja udah kepanasan ??? Kepo banget ; ; #2019GantiPresiden	1	0	1	0	0	1
98	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	0
99	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	0
100	Ha .. ha .. Banyak baca biar pintar ... JEPANG, AMERIKA, CHINA... Utang sama siapa ... Anda tau enggak ..?	0	0	0	0	0	0

131	USER Mana situ ngasih("): itu cuma foto ya kutil onta'	1	1	1	0	0	1

LAMPIRAN B

DATASET SETELAH PENGHAPUSAN LABEL

Berikut ini merupakan hasil penghapusan label dimana label yang tersisa adalah label HS :

No	Tweet	HS
1	disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego !!!'	1
2	RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga'	0
3	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
4	USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIT TAPI DILIAT DARI MANA ITU AKU'	0
5	USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah'	1
6	USER Ya bani taplak dkk \xf0\x9f\x98\x84\xf0\x9f\x98\x84\xf0\x9f\x98\x84'	1
7	deklarasi pilkada 2018 aman dan anti hoax warga dukuh sari jabon	0
8	Gue baru aja kelar re-watch Aldnoah Zero!!! paling kampret emang endingnya! 2 karakter utama cowonya kena friendzone bray! XD URL	0
9	Nah admin belanja satu lagi port terbaik nak makan Ais Kepal Milo, Ais Kepal Horlicks atau Cendol Topping kaw kaw. Å??Å-; ; Docket mano tu ? Gerai Rojak Mertuaku - Taipan 2 (depan TWINS BABY & ROMANTIKA / Bank Islam Senawang) Å???	0
10	USER Enak lg klo smbil ngewe'	0
11	Setidaknya gw punya jari tengah buat lu, sebelum gw ukur nyali sama bacot lu \xf0\x9f\x98\x8f'	1
12	USER USER USER USER BANCI KALENG MALU GA BISA JAWAB PERTANYAAN KAMI DARI 2 HARI LALU.... NYUNGSEP KOE USER URL	1
13	Kalo belajar ekonomi mestinya jago memprivatisasi hati orang. Duh.. ironi USER	0
14	Aktor huruhara 98 Prabowo S ingin lengserkan pemerintahan Jokowi.... Nyata	1
15	USER Bu guru enakan jadi jablay atau guru esde sih.\nKayaknya menikmati jadi pecun ini guru.'	1
16	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 sah. Kafir kyk loe pasti menolak, makanya loe ahlunnar.'	1
17	Belakangan ini kok fikiran ampas banget ya'	0
18	Ari sarua beki mah reph monyet\xf0\x9f\x98\x86\xf0\x9f\x98\x86'	0
19	Jadi cowo itu harus Gantle kalo ga Gantle itu namanya BANCI !'	1
20	USER Slga mnr bom \xf0\x9f\x98\x82'	0
21	Asw ya tapi gua jarang ngambek, tacut wkkwkwkwkw gua kan bucin.'	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 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.

22	USER kalo kamu noob pasti peluang disakitin nya lebih gede si'	0
23	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 g	0
24	PELAJAR SMA KEC BILAH HILIR DEKLARASI ANTI HOAX &	0
25	Bandara Udara Internasional Kertajati dibangun oleh Gubernur Ahmad Heryawan tanpa gunakan APBN pemerintah pusat. Insya Allah siap beroperasi bulan depan	0
26	Siapapun gubernur dan presidennya, rakyatnya, ya kita, juga...	0
27	Ini si USER kerjanya delay mulu! Setan!	1
28	menurutku pintu sorga ada yaitu pintu sorga yang asli dan pintu hatimu modusbanget	0
29	RT USER USER USER PKI hanya muncul jika jelang pemilihan kepala daerah dan pemilihan presiden krn digoreng sampai gosong.	1
30	USER USER Itu mah sdh nenek-nenek sy heran sama cebong biasay bohong terus'	1
31	USER USER USER USER USER USER USER USER Islam Nusantara produk JIL, dipasarkan dg gencar oleh antek\xc2\xb2 JIL.\n\nSementara JIL anti Arab, tapi pro Aseng!\n\nKata (alm) KH. Hasyim Muzadi,	1
32	Dari habis sahur sampe jam 10. Sibayik udah nete 4x. Skg rasanya lemas tak terkira \xf0\x9f\xa4\xa2'	0
33	USER Gak kak emak mah gak demen ama yang sipit wkwk'	0
34	RT USER: Kelakuan homok jaman now, ngentot aja sambil live di Blued #gayvid #gvid #lokalhangat URL	0
35	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
36	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
37	\xe2\x80\x9dPara Bajingan yang telah melukai Lady harus menerima ganjaran seratus kali lipat!!\xe2\x80\x9d(sanji) ##HBD17thPelautJKT48'	0
38	ajakan menolak berita hoax dan sukseskan pilkada di wilayah kota keidri:	0
39	USER USER Presiden RI Jokowi luarBiasa kerja nya tulus IHLAs untuk #RakyatB3rsamaJKW	0
40	Edan emang goollnya fabiano!!'	0
41	GAPERNAH MENDALAMI AL-QURAN YA BANG??? PANTESAN MULUTNYA KAYA ORANG KAFIR BEJAT HINA farhatabbaslaw'	1
42	Sangat terlihat beda sekali Kewibawaannya ; yg sangat berwibawa cuma gubernur Indonesia USER ; Massya Allah...	0
43	Anjir baru kali ini lihat orang bego mau debat pilkada. Kerjanya nyerang karakter. Shame on you mpok silvy and agus. #DebatFinalPilkadaJKT	1
44	#GubernurZamanNow #GusIpulPuti2 #GanjarYasin1 #DjarotSihar2 #HasanAnton2 #KosterAce1 #NurdinSudirman3 #KarolinGidot2 RT USER: "Keanekaragaman budaya, suku, agama, itulah sesungguhnya kekayaan kita,Äç?? tutur Djarot di hadapan peserta RakercabsuÄç?Ä'	0
45	USER USER Apa iya Islam itu ga boleh dilokalisasi? Lha Kristen aja mas ada Nusantarnya 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

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.

68	USER USER USER USER USER USER USER Klw ente atheis..ngapain sibuk dgn agama orang...	0
69	Lengserkan Jokowi Bangsat!!	1
70	USER Selamat hari kartini #2019GANTIPRESIDEN #2019PRESIDENBARU	0
71	RT USER: Bajingan Homo!!!\nHati2 terhadap anak2 kita terutama di kamar mandi laki2, jgn biarkan sendiri! URL	1
72	USER Iya cocot'	1
73	USER karena budaya indonesia.	0
74	USER Yang ngajarin solat'a ajah itu orang asing dia hanya ingin memecah belah islam"	0
75	USER x Lo ngomong aja kaga bener bangke',1	
76	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
77	tapi masih menikmati kebijakan yang dulu diusulkan PKI.. lha yo opo...	1
78	Kalo seumpama orang cina di usir paksa dari indonesia gimana	1
79	USER bunda ngentot yuk'	0
80	RT USER: china babi... china maling...\n#UsirChina https://t.co/v9y4HIWHJp2	1
81	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
82	PROYEK KORUPSI REZIM SBY URL	1
83	USER Wajar saja, lihat saja tingkah bbrp petinggi negeri ini yg suka mencampur bhs indonesia dgn bhs asing setiap kali bicara'	0
84	AB: pribadi yg senang berkompetisi dalam hal yang baik dan positif	0
85	USER USER Terus menurut lo kitab suci agAma apa yg di bilang om USER #Mendadakkalap	1
86	USER Suka nutup mata. Kaga ditutup pun udah terlihat tertutup gara2 sipit /?'	0
87	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
88	USER eh kasar kau kampang'	1
89	BENTAR LAGI GUA RILIS ALBUM THE ONTA'S WKWKWKWKWK"	0
90	USER Ada cebong bang dia merasa paling benar'	1
91	Najis banget gw ngetwit apa barusan wey'	0
92	Kamu transgender atau gmn anjing :(URL	1
93	USER awas bintitan lu ngintipin titit kucing'	0
94	USER USER Bawel, baper, bolot, baik, bodor , gaje wkwk apa ya udah itu aja deh'	0
95	USER USER USER USER Gue saranin gk perlu bnyk bacot maling ayat, langsung to the point aja lu jadi orang... Yg tdk sepaham lu teriakn aja kafir sesat, liberal, pemuja dukun, bla blaaa.. pasti hatimu puas deh ..'	1
96	Jgn salah target bkn ahok... agenda utama lengserkan Jokowi ..	1
97	USER USER masa baru kena hestek aja udah kepanasan ??? Kepo banget	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.

98	Maju lu sini duit tapi kau goyang kaki tak buat pape , mana la duit nak maju kat kau bodo URL	
99	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
100	Ha .. ha .. Banyak baca biar pintar ... JEPANG, AMERIKA, CHINA... Utang sama siapa ... Anda tau enggak ..?	0
...
3127	USER Mana situ ngasih(": itu cuma foto ya kutil onta'	1

Berikut ini merupakan hasil penghapusan label dimana label yang tersisa adalah label Abusive :

No	Tweet	Abusiv e
1	- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego !!!'	1
2	RT USER: USER siapa yang telat ngasih tau elu?edan sarap gue bergaul dengan cigax jifla calis sama siapa noh licew juga'	0
3	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
4	USER USER AKU ITU AKU\n\nKU TAU MATAMU SIPIT TAPI DILIAT DARI MANA ITU AKU'	0
5	USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah'	1
6	USER Ya bani taplak dkk \xf0\x9f\x98\x84\xf0\x9f\x98\x84\xf0\x9f\x98\x84'	1
7	deklarasi pilkada 2018 aman dan anti hoax warga dukuh sari jabon	0
8	Gue baru aja kelar re-watch Aldnoah Zero!!! paling kampret emang endingnya! 2 karakter utama cowonya kena friendzone bray! XD URL	0
9	Nah admin belanja satu lagi port terbaik nak makan Ais Kepal Milo, Ais Kepal Horlicks atau Cendol Topping kaw kaw. ã°??Â-; ; Doket mano tu ? Gerai Rojak Mertuaku - Taipan 2 (depan TWINS BABY & ROMANTIKA / Bank Islam Senawang) ã°???	0
10	USER Enak lg klo smbil ngewe'	0
11	Setidaknya gw punya jari tengah buat lu, sebelum gw ukur nyali sama bacot lu \xf0\x9f\x98\x8f'	1
12	USER USER USER USER BANCI KALENG MALU GA BISA JAWAB PERTANYAAN KAMI DARI 2 HARI LALU.... NYUNGSEP KOE USER URL	1
13	Kalo belajar ekonomi mestinya jago memprivatisasi hati orang. Duh.. ironi USER	0
14	Aktor huruhara 98 Prabowo S ingin lengserkan pemerintahan Jokowi.... Nyata	0
15	USER Bu guru enakan jadi jablay atau guru esde sih.\nKayaknya menikmati jadi pecun ini guru.'	1

Hak Cipta Dilindungi Undang-Undang

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

16	USER USER USER USER USER USER Lawan bicara gw gak intelektual kyk loe, yg otak gak punya. Ttg kencing onta, gw mengakui hadis nabi dan itu sah. Kafir kyk loe pasti menolak, makanya loe ahlunnar.'	1
17	Belakangan ini kok fikiran ampas banget ya'	0
18	Ari sarua beki mah refeh monyet\xf0\x9f\x98\x86\xf0\x9f\x98\x86'	1
19	Jadi cowo itu harus Gantle kalo ga Gantle itu namanya BANCI !'	1
20	USER Slga mnr bom \xf0\x9f\x98\x82'	0
21	Asw ya tapi gua jarang ngambek, tacut wkkwkwkwkw gua kan bucin.'	1
22	USER kalo kamu noob pasti peluang disakitin nya lebih gede si'	0
23	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 g	0
24	PELAJAR SMA KEC BILAH HILIR DEKLARASI ANTI HOAX & PILKADA 2018 DAMAI	0
25	Bandara Udara Internasional Kertajati dibangun oleh Gubernur Ahmad Heryawan tanpa gunakan APBN pemerintah pusat. Insyallah siap beroperasi bulan depan	0
26	Siapapun gubernur dan presidennya, rakyatnya, ya kita, juga...	0
27	Ini si USER kerjanya delay mulu! Setan!'	1
28	menurutku pintu sorga ada yaitu pintu sorga yang asli dan pintu hatimu modusbanget	0
29	RT USER USER USER PKI hanya muncul jika jelang pemilihan kepala daerah dan pemilihan presiden krn digoreng sampai gosong.	0
30	USER USER Itu mah sdh nenek-nenek sy heran sama cebong biasay bohong terus'	1
31	USER USER USER USER USER USER USER USER Islam Nusantara produk JIL, dipasarkan dg gencar oleh antek\xc2\xb2 JIL.\n\nSementara JIL anti Arab, tapi pro Aseng!\n\nKata (alm) KH. Hasyim Muzadi,	0
32	Dari habis sahur sampe jam 10. Sibayik udah nete 4x. Skg rasanya lemas tak terkira \xf0\x9f\xa4\xa2'	0
33	USER Gak kak emak mah gak demen ama yang sipit wkwk'	0
34	RT USER: Kelakuan homok jaman now, ngentot aja sambil live di Blued #gayvid #gvid #lokalhangat URL	1
35	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
36	Pegawai Guardian training lagi kek Mentang2 gue kek gembel cuman kaos sama Flat shoes nanya dijutekin bangsada emang males belinya juga kabur gua ke matahari'	1
37	\xe2\x80\x9dPara Bajingan yang telah melukai Lady harus menerima ganjaran seratus kali lipat!!\xe2\x80\x9d(sanji) ##HBD17thPelautJKT48'	1
38	ajakan menolak berita hoax dan sukseskan pilkada di wilayah kota keidri:	0
39	USER USER Presiden RI Jokowi luarBiasa kerja nya tulus IHLAs untuk #RakyatB3rsamaJKW	0
40	Edan emang goollnya fabiano!!'	1

41	GAPERNAH MENDALAMI AL-QURAN YA BANG??? PANTESAN MULUTNYA KAYA ORANG KAFIR BEJAT HINA farhatabbaslaw'	1
42	Sangat terlihat beda sekali Kewibawaannya ;, yg sangat berwibawa cuma gubernur Indonesia USER ; Massya Allah...	0
43	Anjir baru kali ini lihat orang bego mau debat pilkada. Kerjanya nyerang karakter. Shame on you mpok silvy and agus. #DebatFinalPilkadaJKT	1
44	#GubernurZamanNow #GusIpulPuti2 #GanjarYasin1 #DjarotSihar2 #HasanAnton2 #KosterAce1 #NurdinSudirman3 #KarolinGidot2 RT USER: "Keanekaragaman budaya, suku, agama, itulah sesungguhnya kekayaan kita,Ã¿?? tutur Djarot di hadapan peserta RakercabsuÃ¿?Ã¿!	0
45	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
46	Ciptakan Situasi Kamtibmas Menjelang Pilkada 2018, Kapolsek Konang Silaturahmi ke Tokoh Masyarakat	0
47	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\nNurl	0
48	USER USER USER USER Kata Sembahyang mengadopsi tradisi lokal pada masa kerjaan Hindu.\n\nIstilah\xc2\xa0sembahyang\xc2\xa0pun lahir dari tradisi ritus\xc2\xa0menyembahHyang (Yang Tunggal), sama dengan\xc2\xa0shalat menyembah All	0
49	KAU PENDIKAN TINGGI TAPI BODOH...GK ADA LARANG BAWA AGAMA..BIAR ORANG ISLAM YG DI BAYAR AHOK SADAR GK JADI MILIH AHOK GLA	1
50	USER USER BTP the best ,,,'abaikan kaum kampret otak micin ,,,'	1
51	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
52	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.. Png Sambil ky nhn ngeden.."dungu!". \xf0\x9f\x98\x8b	1
53	pinokio umur 3 taonan tuh. #2019GantiPresiden	0
54	USER Cerita silat lae \xf0\x9f\x98\x80 semacam Kho Ping Hoo yang dari Indonesia, tapi Liang Ie Shen penulis dari China'	0
55	RT USER benci sekali dgn Umat Islam	0
56	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
57	USER dari pada gua nete'	1
58	USER Lol bani lowo'	1
59	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 hanya seorang ahli ilmu filsafat yg tdk failosof; ; Mari kita demo diri sendi	0
60	A orang yang suka ke konser-konser, sedikit boros.	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 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.

61	[SAMPAI KAPAN MAKAN RACUN ?]_ Banyak ulama yang mengatakan hukum rokok mubah dan ada juga?!	0
62	Susilo Bambang Yudhoyono Presiden RI ke 6 meminta Khofifah Emil untuk dekat dengan masyarakat . No 1 Bisa ! BaGusS #KhofifahEmilNomer1	0
63	USER udah biasa kali. Gue kpoper tapi kalo ngomov sama temen2 gue, idol gua banci homo'	1
64	Saya suka video USER URL Awal Kisah Perseteruan Haji Bolot & Malih Tongtong'	0
65	Budaya kafir bersungguh, klau hal2 islam tidak diendahkan...	0
66	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\xab\xe2\x99\xab\xe2\x99\xab\xe2\x99\xab'	0
67	Apabila setan membisikkan was-was kepadamu, mintalah perlindungan kepada Allah. (QS. al-A\xe2\x80\x99raf: 200 dan Fushilat: 36).'	0
68	USER USER USER USER USER USER USER Klw ente atheis..ngapain sibuk dgn agama orang...	0
69	Lengserkan Jokowi Bangsat!!	1
70	USER Selamat hari kartini #2019GANTIPRESIDEN #2019PRESIDENBARU	0
71	RT USER: Bajingan Homo!!!\nHati2 terhadap anak2 kita terutama di kamar mandi laki2, jgn biarkan sendiri! URL	1
72	USER Iya cocot'	1
73	USER karena budaya indonesia.	0
74	USER Yang ngajarin solat'a ajah itu orang aseng dia hanya ingin memecah belah islam"	0
75		
76	USER x Lo ngomong aja kaga bener bangke'	1
77	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
78	tapi masih menikmati kebijakan yang dulu diusulkan PKI.. lha yo opo...	0
79	Kalo seumpama orang cina di usir paksa dari indonesia gimana	0
80	USER bunda ngentot yuk'	1
81	RT USER: china babi... china maling...\n#UsirChina https://t.co/v9y4HIWHJp2	1
82	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
83	PROYEK KORUPSI REZIM SBY URL	0
84	USER Wajar saja, lihat saja tingkah bbrp petinggi negeri ini yg suka mencampur bhs indonesia dgn bhs asing setiap kali bicara'	0
85	AB: pribadi yg senang berkompetisi dalam hal yang baik dan positif	0
86	USER USER Terus menurut lo kitab suci agAma apa yg di bilang om USER #Mendadakkalap	0
87	USER Suka tutup mata. Kaga ditutup pun udah terlihat tertutup gara2 sipit /?'	0

88	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
89	USER eh kasar kau kampung'	0
90	BENTAR LAGI GUA RILIS ALBUM THE ONTA'S WKWKWKWKWK"	0
91	USER Ada cebong bang dia merasa paling benar'	0
92	Najis banget gw ngetwit apa barusan wey'	1
93	Kamu transgender atau gmn anjing :(URL	1
94	USER awas bintitan lu ngintipin titit kucing'	1
95	USER USER Bawel, baper, bolot, baik, bodor , gaje wkwk apa ya udah itu aja deh'	1
96	USER USER USER USER Gue saranin gk perlu bnyk bacot maling ayat, langsung to the point aja lu jadi orang... Yg tdk sepaham lu teriakn aja kafir sesat, liberal, pemuja dukun, bla blaaa.. pasti hatimu puas deh ..'	1
97	Jgn salah target bkn ahok... agenda utama lengserkan Jokowi ..	0
98	USER USER masa baru kena hestek aja udah kepanasan ??? Kepo banget #2019GantiPresiden	0
99	Maju lu sini duit tapi kau goyang kaki tak buat pape , mana la duit nak maju kat kau bodo URL	1
100	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
...
1312 7	USER Mana situ ngasih(": itu cuma foto ya kutil onta'	1

Berikut ini merupakan hasil penghapusan label dimana label yang tersisa adalah label Level :

No	Tweet	Level
1	- disaat semua cowok berusaha melacak perhatian gue. loe lantas remehkan perhatian yg gue kasih khusus ke elo. basic elo cowok bego !!!'	1
2	USER USER Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah'	2
3	USER Ya bani taplak dkk \xf0\x9f\x98\x84\xfo\x9f\x98\x84\xfo\x9f\x98\x84'	2
4	Setidaknya gw punya jari tengah buat lu, sebelum gw ukur nyali sama bacot lu \xf0\x9f\x98\x8f'	1
5	USER USER USER USER BANCI KALENG MALU GA BISA JAWAB PERTANYAAN KAMI DARI 2 HARI LALU.... NYUNGSEP KOE USER URL	1
6	Aktor huruhara 98 Prabowo S ingin lengserkan pemerintahan Jokowi.... Nyata	1
7	USER Bu guru enakan jadi jablay atau guru esde sih.\nKayaknya menikmati jadi pecun ini guru.'	1

Hak Cipta Dilindungi Undang-Undang

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

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.

8	USER USER USER USER USER USER Lawan bicara gw gak intelektual 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
9	Jadi cowo itu harus Gantle kalo ga Gantle itu namanya BANCI !'	2
10	Ini si USER kerjanya delay mulu! Setan!'	2
11	RT USER USER USER PKI hanya muncul jika jelang pemilihan kepala daerah dan pemilihan presiden krn digoreng sampai gosong.	2
12	USER USER Itu mah sdh nenek-nenek sy heran sama cebong biasay bohong terus'	1
13	USER USER USER USER USER USER USER USER Islam Nusantara produk JIL, dipasarkan dg gencar oleh antek\xc2\xb2 JIL.\n\nSementara JIL anti Arab, tapi pro Aseng!\n\nKata (alm) KH. Hasyim Muzadi,	2
14	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'	2
15	GAPERNAH MENDALAMI AL-QURAN YA BANG??? PANTESAN MULUTNYA KAYA ORANG KAFIR BEJAT HINA farhatabbaslaw'	1
16	Anjir baru kali ini lihat orang bego mau debat pilkada. Kerjanya nyerang karakter. Shame on you mpok silvy and agus. #DebatFinalPilkadaJKT	2
17	KAU PENDIKAN TINGGI TAPI BODOH...GK ADA LARANG BAWA AGAMA..BIAR ORANG ISLAM YG DI BAYAR AHOK SADAR GK JADI MILIH AHOK GLA	1
18	USER USER BTP the best ,,abaikan kaum kampret otak micin ,,'	2
19	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
20	USER Awaasss... Jgn sampe beritanya sampe ke akhi gerung! Ntar susah nyari kosakatanya. 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!". \xf0\x9f\x98\x8b	1
21	pinokio umur 3 taonan tuh. #2019GantiPresiden	1
22	RT USER benci sekali dgn Umat Islam	2
23	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
24	USER Lol bani lowo'	2
25	USER udah biasa kali. Gue kper tapi kalo ngomong sama temen2 gue, idol gua banci homo'	1
26	Lengserkan Jokowi Bangsaat!!	3
27	RT USER: Bajingan Homo!!!\nHati2 terhadap anak2 kita terutama di kamar mandi laki2, jgn biarkan sendiri! URL	1
28	USER Iya cocot'	1
29	USER x Lo ngomong aja kaga bener bangke'	1
30	tapi masih menikmati kebijakan yang dulu diusulkan PKI.. lha yo opo...	2
31	Kalo seumpama orang cina di usir paksa dari indonesia gimana	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.

32	RT USER: china babi... china maling...\n#UsirChina https://t.co/v9y4HIWHJp2	3
33	PROYEK KORUPSI REZIM SBY URL	1
34	USER USER Terus menurut lo kitab suci agAma apa yg di bilang om USER #Mendadakkalap	1
35	USER eh kasar kau kampung'	1
36	USER Ada cebong bang dia merasa paling benar'	1
37	Kamu transgender atau gmn anjing :(URL	1
38	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
39	Jgn salah target bkn ahok... agenda utama lengserkan Jokowi ..	3
40	USER USER masa baru kena hestek aja udah kepanasan ??? Kepo banget #2019GantiPresiden	1
41	Prabowo Sudah Kalah Menyebut Bantuan Jokowi Hanya Pencitraan Adalah Ratapan Pulu	1
42	USER USER Yg goblog itu adalah bani cebong tukang tipu penjilat penguasa yg ketahuan gerakin masa yg dibayar pakai nasi bungkus. Propaganda nasi bungkus memang selalu gagal.'	2
43	USER USER USER USER Wuih... cebong sewot...\n\x0f\x9f\x98\x82\x0f\x9f\x98\x82\x0f\x9f\x98\x82'	1
44	Padahal gubernur saat ini Djarot. Mayoritas partai politik di DPRD juga pendukung Ahok Djarot. Payah ya mereka USER URL	1
45	USER Anak pecun... ga jauh2 dr hobi zina.. haha'	1
46	Pendaftaran Capres untuk #Pilpres2019 saja belum dibuka sampai Agustus 2018...yang normatif begini politikus PDIP masih belum paham? #2019GantiPresiden !!	2
47	Si tolol udah tau pemerintah rezim komunis ini bobrok koruptor bisa pimpinan sidang DPR paya cebong URL	1
48	Cina perusak bangsa!! Usir !! Stuju??	3
49	sional hak setiap rakyat woiii!! . Gini nih kalo planga plongo tapi kelakuan otoriter.	1
50	USER kroasia itu kapir tauuu,ngapain blajar sm kapir...udh lupa sm kapling sorga?'	2
51	USER USER loh bukan nya mentri suharto ada yg homo?'	1
52	Kejar setorang memang cukup bangsad, mana rupiah turun \x0f\x9f\x98\xa5'	1
53	Moga ini ada tindak lanjutnya!; ; Pelaporan terhadap ?penista agama? Ade Armando akan dilakukan langsung oleh Ketua Umum FPI KH. Shobri Lubis ke Bareskrim Mabes Polri, Gambir, Jakarta Pusat pada Selasa (10/4) pukul 13.00 WIB.; #Perhaps USER;	1
54	USER USER USER USER USER USER USER USER Bangke.. hajar!!!!'	3
55	USER USER USER .\nOnta mati klo kebanyakan Minum'	2
56	URL Ahok tng mrka trus was-was hdpnya USER USER USER USER USER USER USER	1
57	Itu baru di Sul Teng.. !!! Ini eksekusi akibat tersandera hutang china !!! #2019GantiPresiden #2019PresidenBaru #2019TetapJokowiGanti	1
58	USER Tak hanya kurang bertaji nya kpk era ini...tapi era ini juga komitmen dlm pemberantasan korupsi tdk spt di era Pak sby....	2

59	RT USER: Admin tolol di partai tolol\nPartai Tolol untuk orang Tolol'	2
60	Jokowi aja yg mundur. USER tutup. URL	1
61	USER Ngomong opo itu.mulut kok seperti banci...beli kaca dulu lihat dirimu,. Broo'	1
62	USER Ada amin rasis.\nAda amat dani\nAda ratna si paaiit.\nAda si onta.\nAda si lobang anus.\nJdi jngan takut wo wo..\nKehabisan obat penenang yh.'	1
63	USER Tunggu komando IMAM BEJAT CABUL USER . TAIK BEER \xf0\x9f\x8d\xba\n\nApe2 bawa ULAMA...\nDASAR Banj SEMPAK'	1
64	Kabinet berantakan \nPemecah belah persatuan hanya ad di rezim ini'	2
65	dih! Ngaku lebih pintar dr orang politik, NGACA DULU WOYY tampang loe aja k\xc3\xa6k orang BEGO!!! USER ckakakak'	1
66	#PDIPerjuanganMenang Dengan menekan perusahaan tambang dgn target proyek yang impossible, lalu dilelang murah pada antek Cina. NTB contoh kasus. PDIP mau menang? Dasar bajingan'	2
67	USER USER ga caya sama mulut buaya mu bwang'	1
68	RT USER USER Jangan kayak BANCI KALENG kamu GATOT sering buat Gaduh gentleman lah MUNDUR siap bersaing di pemilihan presidens 2019 melawan Jokowi	1
69	USER USER umuh > persib\npersib kunyuk!!!!\n\xf0\x9f\x98\x82\xf0\x9f\x98\x82\xf0\x9f\x98\x82'	2
70	USER USER Buat film muslim pilih ahok mati nggak di sholatin..... dijamin laris	1
71	Pemberitaan saracen diframing katak seakan tim prabowo dan anies-sandi. Seberulnya itu media menumpahkan sakit hati katak dan cebong. *Ehe	2
72	USER USER USER USER Kan ane bukan cebong udeh dibilangin dari maren masih ngeyel\nSitu maksa ane jadi cebong?'	1
73	USER Kelemahan gerindra itu ada di orang ini,...Prabowo turun elektabilitasnya karena CONGOR fadlizon spt DURNO'	1
74	USER USER USER mau menasehati kali supaya ga ugal2an nih rezim.. eh tapi ini negeri sdh keburu ancor..nah skrg saja ekonomi morat marit,, kacaooooo!!'	1
75	USER tahun 2019 Rakyat makan ASPAL dan BETON; ini adalah solusi jitu mengurangi kemiskinan; ; KEAHLIAN CEBONG HANYA 1, TUKANG TIPU !!!; cebong mah gitu, goblok nggak ketulungan. Plonga plongo mengakibatkan optimisme berlebihan; ; #2019GantiPresiden #	2
76	USER Kemarin di hujat sekarang mengharap dasar para bani kamprey'	2
77	USER Penjelasan sy bukan FITNAH, silahkan DIBANTAH USER Apollinaris Darmawan Sebut Al-Quran Kitab Sampah	2
78	Monyet aja koloe dbilang saudra sma elo pasti ogah, hdupmu kok gini amat ya.. Toe Hati apa comberan? USER'	1
79	Becak Warga Binjai Hendak Disita karena Gunakan Tenda Bertuliskan #2019GantiPresiden, PKS MERADANG; ; Bukti pemerintah panik dan takut?; ;	2
80	lebih baik jokowi mundur..	1
81	Oleh rezim skrg atas kendali Negara Cina Komunis yg mjd patron tunggal rezim anti islam (Jokowi)	2
82	USER Apaansi tai baru gt doang belagunya udah selangit. Ga inget ada karma? Ntar bbrp taun lagi kalo idol lu flop, gua ketawa paling keras.'	1

Hak Cipta Dilindungi Undang-Undang

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



DAFTAR RIWAYAT HIDUP

Nama Lengkap : Aulil Amri
 Tempat/Tanggal Lahir : Sekeladi, 31 Juli 1998
 Agama : Islam
 Golongan Darah : B+
 Anak ke : 4
 Jumlah Saudara : 4

Alamat : Jl. Merpati Sakti No.97, Kelurahan Simpang Baru, Kecamatan Tampan, Pekanbaru.

Email : 11651103437@students.uin-suska.ac.id

PENDIDIKAN

2004-2010 : SD Negeri 017 Sedingin, Tanah Putih, Rokan Hilir
 2010-2013 : SMP Negeri 1 Tanah Putih, Rokan Hilir
 2013-2016 : SMA Negeri 1 Tanah Putih, Rokan Hilir
 2016-2020 : Universitas Islam Negeri Sultan Syarif Kasim Riau, Jurusan Teknik Informatika

PENGALAMAN ORGANISASI

2018-2019 : Generasi Baru Indonesia (GenBI Riau) Divisi Informasi dan Komunikasi (Infokom)
 2018-2019 : Dewan Eksekutif Mahasiswa Fakultas Sains dan Teknologi Divisi Dinas Pemuda dan Olahraga (DISPORA)

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.