

PERBANDINGAN ALGORITMA NAÏVE BAYES DENGAN REGRESI LOGISTIK UNTUK KLASIFIKASI HATE SPEECH

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika Fakultas Ilmu Komputer
Universitas Sriwijaya



Oleh :

Deri Yuliansyah
NIM : 09021181320017

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020

LEMBAR PENGESAHAN USULAN TUGAS AKHIR

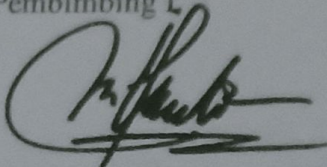
PERBANDINGAN ALGORITMA NAÏVE BAYES DENGAN
REGRESI LOGISTIK UNTUK KLASIFIKASI HATE SPEECH

Oleh :

Deri Yuliansyah
NIM : 09021181320017

Palembang, Agustus 2020

Pembimbing I,



Ir. M. Ihsan Jambak, M.Sc.
NIP. 196804052013081202

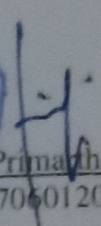
Pembimbing II,



Yunita, M.Cs.
NIP. 198306062015042002

Mengetahui,
Ketua Jurusan Teknik Informatika




Rifkie Primayha, MT
NIP. 197706012009121004

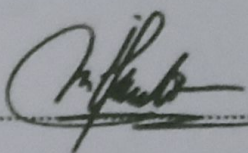
TANDA LULUS SIDANG TUGAS AKHIR

Pada hari Rabu, 29 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Deri Yuliansyah
NIM : 09021181320017
Judul : Perbandingan Algoritma Naïve Bayes dengan Regresi Logistik untuk Klasifikasi Hate Speech

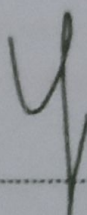
1. Pembimbing I

Ir. M. Ihsan Jambak, M.Sc.
NIP. 196804052013081202



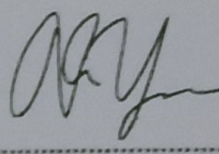
2. Pembimbing II

Yunita, M.Cs.
NIP. 198306062015042002



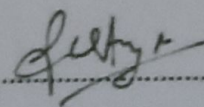
3. Penguji I

Novi Yusliani, M.T.
NIP. 198211082012122001



4. Penguji II

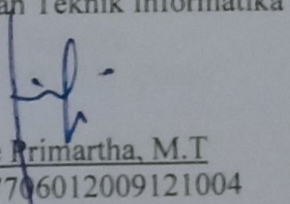
Desty Rodiah, M.T.
NIK 1671016112890005



Mengetahui,
Ketua Jurusan Teknik Informatika



Riflye Primartha, M.T.
NIP. 197706012009121004



HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Deri Yuliansyah
NIM : 09021181320017
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Algoritma Naïve Bayes dengan Regresi
Logistik Untuk Klasifikasi Hate Speech
Hasil Pengecekan Software *iThenticate/Turnitin* : 16 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Agustus 2020



Deri Yuliansyah
NIM. 09021181320017

Motto:

“Don’t compare your life to others. There’s no comparison between the sun and the moon. They shine when it’s their time”

Kupersembahkan tulisan ini kepada

:

- Allah SWT
- Kedua orang tuaku tercinta, dan nenek tercinta
- Ketiga saudaraku
- Keluargaku
- Teman-teman seperjuangan
- Almamater kebanggaanku

**PERBANDINGAN ALGORITMA *NAÏVE BAYES* DENGAN REGRESI
LOGISTIK UNTUK KLASIFIKASI *HATE SPEECH***

BY :

DERI YULIANSYAH

09021181320017

ABSTRACT

Text classification or document classification is one of the main tasks in text data mi

ning and information retrieval. Classification is used to predict group membership against an instance of data. Hate Speech is an act of communication that vilifies individual targets or specific group characteristics such as race, ethnicity, gender, religion, and others. The Naïve Bayes Algorithm is an algorithm based on the concept of Bayes probability, while logistic regression is a model used to find the relationship between 2 variables, namely two-category response variables or dichotomous variables or variables with two or more categories which are polychotomous with one or more predictors on a category or continuous scale. In this study, a comparison between the Naïve Bayes algorithm and Logistic Regression was carried out to classify hate speech. This study aims to compare the accuracy and computation time of the algorithms used.

Keywords: Hate speech classification, Naïve Bayes Algorithm, Regression Algorithm Logistics, Accuracy, Computing Time

PERBANDINGAN ALGORITMA *NAÏVE BAYES* DENGAN REGRESI LOGISTIK UNTUK KLASIFIKASI *HATE SPEECH*

OLEH :

DERI YULIANSYAH

09021181320017

ABSTRAK

Klasifikasi teks atau klasifikasi dokumen merupakan salah satu tugas utama dalam *text data mining* dan *information retrieval*. Klasifikasi digunakan untuk memprediksi keanggotaan kelompok terhadap suatu *data instances*. *Hate Speech* merupakan tindakan komunikasi yang menjelekkan dengan target individu atau karakteristik kelompok tertentu seperti ras, etnis, gender, agama, dan lain-lain. Algoritma Naïve Bayes adalah sebuah algoritma yang berdasarkan pada konsep dari probabilitas *Bayes* sedangkan Regresi logistik merupakan suatu model yang digunakan untuk mencari hubungan antara 2 variabel yakni variabel respon dua kategori atau variabel yang bersifat *dichotomous* maupun variabel dengan dua kategori atau lebih yang bersifat *polychotomous* dengan satu atau lebih prediktor berskala kategori ataupun kontinu. Dalam penelitian ini dilakukan perbandingan antara algoritma Naïve Bayes dengan Regresi Logistik untuk mengklasifikasi *hate speech*. Penelitian ini bertujuan untuk mengetahui perbandingan akurasi dan waktu komputasi dari algoritma yang digunakan.

Kata Kunci : Klasifikasi hate speech, Algoritma Naïve Bayes, Algoritma Regresi Logistik, Akurasi, Waktu Komputasi

KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkah dan rahmat-Nya penulis dapat menyelesaikan tugas akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi selama proses penelitian ini dilaksanakan. Secara khusus penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tua saya, LD. Muso dan Isnaini serta seluruh keluarga besar saya yang selalu mendokan serta memberikan dukungan baik moril maupun materiil. Terima kasih.
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika dan Dosen Pembimbing Akademik.
4. Bapak Ir. M Ihsan Jambak, M.Sc. dan Ibu Yunita M.Cs. selaku dosen pembimbing I dan II yang telah memberikan arahan, bimbingan, bantuan dan mau mendengarkan lalu mengoreksi segala pendapat penulis selama proses pengerjaan tugas akhir ini.
5. Ibu Novi Yusliani, M.T. dan Ibu Desty Rodiah, M.T. selaku dosen penguji I dan II yang telah memberikan masukan, saran dan motivasi yang tinggi agar penulis dapat memahami dan menyelesaikan tugas akhir ini dengan baik.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmunya kepada penulis selama proses perkuliahan.

7. Seluruh staf administrasi dan pegawai yang telah membantu dan mendukung Penulis dalam hal administrasi perkuliahan.
8. Terima kasih kepada kakak saya Deni Meisya Putra, S.SI. serta kedua adik saya Demi Alfarizhi dan Debi Ismu Sofuan yang selama ini telah memberikan dukungan selama proses tugas akhir ini.
9. Seluruh teman-teman mahasiswa pejuang skripsi “Teknik Informatika Reguler” dan “Teknik Informatika Bilingual” yang telah bersama-sama menjalani proses tugas akhir, berbagi nasihat, motivasi dan hiburan pada proses pengerjaan tugas akhir ini.
10. Seluruh teman-teman Setarbak Ilkom dan Masyarakat Layo yang telah memberi motivasi berbagi nasihat dan hiburan pada proses pengerjaan tugas akhir ini.
11. Kepada Pratiwi Rahmayanti, S.TP. yang setia memperhatikan saya, menemani dan memberikan dukungan, mendengarkan keluh kesah, dan kekhawatiran serta banyak membantu sepanjang mengerjakan tugas akhir ini.
12. Faris Harun Ahmad, Singgih Fajar Pratomo, Moh Azmi, Kurnia Sandi P, Robihin Permata, Yudha Sahreza dan semua teman yang selalu saya repotkan.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Agustus 2020



Deri Yuliansyah

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
TANDA LULUS SIDANG TUGAS AKHIR.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xvi
DAFTAR GAMBAR	xix
 BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Perumusan Masalah.....	I-3
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6

BAB II TINJAUAN PUSTAKA

2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Data Mining	II-1
2.2.2 Text Data Mining	II-1
2.2.3 Klasifikasi	II-2
2.2.4 Naïve Bayes	II-4
2.2.5 Regresi Logistik	II-6
2.2.6 Cross Validation	II-7
2.2.7 Confusion Matrix	II-8
2.2.8 Preprocessing	II-10
2.2.8.1 Case Folding	II-10
2.2.8.2 Tokenizing	II-11
2.2.8.3 Stopwords Removal	II-11
2.2.8.4 Stemming	II-11
2.2.9 Term Frequency – Inverse Document Frequency	II-12
2.2.9.1 Term Frequency	II-12
2.2.9.2 Inverse Document Frequency	II-13
2.2.10 Hate Speech	II-13
2.2.11 Metode Pengembangan Perangkat Lunak	II-14
2.2.12 Fase RUP	II-16
2.3 Penelitian yang Relevan	II-17

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan	III-1
3.2 Unit Penelitian	III-1
3.3 Metode Pengumpulan Data	III-1
3.3.1 Menentukan Ruang Lingkup dan Unit Penelitian	III-2
3.3.2 Menetapkan Kriteria Pengujian	III-2
3.3.3 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-3
3.3.4 Melaksanakan Pengujian Penelitian	III-3
3.3.5 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-8
3.4 Metode Pengembangan Perangkat Lunak	III-8
3.4.1 Fase Insepsi.....	III-8
3.4.2 Fase Elaborasi.....	III-9
3.4.3 Fase Konstruksi	III-9
3.4.4 Fase Transisi	III-10
3.5 Manajemen Proyek Penelitian.....	III-10
3.6 Kesimpulan.....	III-20

BAB IV Pengembangan Perangkat Lunak

4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis	IV-2
4.2.2 Kebutuhan Sistem.....	IV-3

4.2.3 Analisis dan Desain	IV-14
4.2.3.1 Analisis Kebutuhan Perangkat Lunak	IV-14
4.2.3.1.1 Analisis <i>Preprocessing</i>	IV-14
4.2.3.1.2 Analisis Pembobotan Kata.....	IV-17
4.2.3.1.3 Pengujian Algoritma Naïve Bayes.....	IV-18
4.2.3.1.2 Pengujian Algoritma Regresi Logistik	IV-20
4.2.3.2 Desain Kebutuhan Perangkat Lunak	IV-24
4.3 Fase Elaborasi.....	IV-25
4.3.1 Pemodelan Bisnis	IV-25
4.3.1.1 Perancangan Data	IV-35
4.3.1.2 Perancangan Antarmuka.....	IV-36
4.4 Fase Konstruksi	IV-37
4.4.1 Pemodelan Bisnis	IV-38
4.4.2 Kebutuhan.....	IV-40
4.5 Fase Transisi.....	IV-43
4.5.1 Pemodelan Bisnis	IV-43
4.5.2 Kebutuhan.....	IV-43
4.5.3 Rencana Pengujian.....	IV-44
4.5.3.1 Rencana Pengujian <i>Use Case</i> Melakukan Input Data	IV-44
4.5.3.2 Rencana Pengujian <i>Use Case</i> Melakukan Konversi Data	IV-45
4.5.3.3 Rencana Pengujian <i>Use Case</i> Melakukan Pembagian Data	IV-45

4.5.3.4 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma <i>Naïve Bayes</i>	IV-46
4.5.3.5 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma Regresi Logistik	IV-47
4.5.4 Implementasi.....	IV-48
4.5.4.1 Pengujian <i>Use Case</i> Melakukan Input Data	IV-49
4.5.4.2 Pengujian <i>Use Case</i> Melakukan Konversi Data.....	IV-50
4.5.4.3 Pengujian <i>Use Case</i> Melakukan Pembagian Data.....	IV-51
4.5.4.4 Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma <i>Naïve Bayes</i>	IV-53
4.5.4.5 Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma Regresi Logistik	IV-55

BAB V Hasil dan Analisis Penelitian

5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan	V-2
5.2.1 Konfigurasi Percobaan.....	V-1
5.2.2 Data Hasil Konfigurasi Percobaan I	V-3
5.2.3 Data Hasil Konfigurasi Percobaan II.....	V-4
5.2.4 Data Hasil Konfigurasi Percobaan III.....	V-5
5.3 Analisis Hasil Penelitian	V-6
5.4 Waktu Komputasi	V-7

5.5 Kesimpulan	V-9
BAB VI Kesimpulan dan Saran	
6.1 Pendahuluan	VI-1
6.2 Kesimpulan.....	VI-2
6.3 Saran	VI-2
DAFTAR PUSTAKA	xx
LAMPIRAN.....	L-1

DAFTAR TABEL

	Halaman
III-1 Rancangan Tabel Hasil Pengujian Klasifikasi Hatespeech	III-5
III-3 Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS)	III-16
IV-1 Kebutuhan Fungsional.....	IV-4
IV-2 Kebutuhan Non-Fungsional	IV-4
IV-3 Definisi Aktor.....	IV-6
IV-4 Definisi <i>Use Case</i>	IV-6
IV-5 Skenario <i>Use Case</i> Melakukan Input Data.....	IV-8
IV-6 Skenario <i>Use Case</i> Melakukan Konversi Data	IV-9
IV-7 Skenario <i>Use Case</i> Melakukan Pembagian Data	IV-11
IV-8 Skenario <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma <i>Naïve</i> <i>Bayes</i>	IV-12
IV-9 Skenario <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma Regresi Logistik	IV-13
IV-10 <i>Tokenizing Tweet</i>	IV-15
IV-11 <i>Stopword Removal Tweet</i>	IV-16
IV-12 <i>Stemming Tweet</i>	IV-16
IV-13 Pembobotan Kata <i>Tweet</i>	IV-17
IV-14 Data Latih untuk <i>Naïve Bayes</i>	IV-19
IV-15 Data Latih untuk Regresi Logistik	IV-22
IV-16 Model Klasifikasi Regresi Logistik.....	IV-23

IV-17 Implementasi Kelas	IV-40
IV-18 Rencana Pengujian <i>Use Case</i> Melakukan Input Data	IV-44
IV-19 Rencana Pengujian <i>Use Case</i> Melakukan Konversi Data.....	IV-45
IV-20 Rencana Pengujian <i>Use Case</i> Melakukan Pembagian Data.....	IV-45
IV-21 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma <i>Naïve Bayes</i>	IV-46
IV-22 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma Regresi Logistik	IV-47
IV-23 Pengujian <i>Use Case</i> Melakukan Input Data.....	IV-49
IV-24 Pengujian <i>Use Case</i> Melakukan Konversi Data.....	IV-50
IV-25 Pengujian <i>Use Case</i> Melakukan Pembagian Data	IV-52
IV-26 Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma <i>Naïve Bayes</i>	IV-54
IV-27 Pengujian <i>Use Case</i> Melakukan Klasifikasi Menggunakan Algoritma Regresi Logistik	IV-56
V-1 Konfigurasi Percobaan I	V-3
V-2 Konfigurasi Percobaan II.....	V-3
V-3 Konfigurasi Percobaan III.....	V-3
V-4 Hasil Akurasi Percobaan Konfigurasi I	IV-4
V-5 Hasil Akurasi Percobaan Konfigurasi II.....	IV-5
V-6 Hasil Akurasi Percobaan Konfigurasi III.....	IV-6

DAFTAR GAMBAR

	Halaman
II-1 Arsitektur RUP.....	II-15
III-2 Tahapan Pengujian Penelitian.....	III-7
III-3 Penjadwalan untuk Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....	III-16
III-4 Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian dan Menentukan Kriteria Pengujian	III-17
III-5 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Insepsi	III-17
III-6 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Elaborasi	III-18
III-7 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Konstruksi.....	III-18
III-8 Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase Transisi.....	III-19
III-2 Penjadwalan untuk Tahap Melakukan Pengujian Penelitian, Analisa Hasil Pengujian Penelitian dan Membuat Kesimpulan	III-19
IV-1 Diagram <i>Use Case</i> Perangkat Lunak	IV-5
IV-2 Diagram Aktivitas Melakukan Input Data	IV-26
IV-3 Diagram Aktivitas Melakukan Konversi Data	IV-27
IV-4 Diagram Aktivitas Melakukan Pembagian Data	IV-28

IV-5 Diagram Aktivitas Melakukan Klasifikasi Menggunakan Algoritma Naïve	
Bayes	IV-29
IV-6 Diagram Aktivitas Melakukan Klasifikasi Menggunakan Algoritma	
Regresi Logistik	IV-30
IV-7 <i>Sequence</i> Diagram Melakukan Input Data	IV-31
IV-8 <i>Sequence</i> Diagram Melakukan Konversi Data	IV-32
IV-9 <i>Sequence</i> Diagram Melakukan Pembagian Data	IV-33
IV-10 <i>Sequence</i> Diagram Melakukan Klasifikasi Menggunakan Algoritma Naïve	
Bayes	IV-34
IV-11 <i>Sequence</i> Diagram Melakukan Klasifikasi Menggunakan Algoritma	
Regresi Logistik	IV-35
IV-12 Rancangan Antarmuka Menu Utama	IV-36
IV-13 Rancangan Antarmuka Menu <i>Training - Testing</i>	IV-37
IV-14 Class Diagram	IV-29
V-1 Perbandingan Akurasi <i>Naïve Bayes</i> dan Regresi Logistik dengan Konfigurasi	
1, 2, dan 3.....	V-7
V-2 Perbandingan Waktu Komputasi <i>Naïve Bayes</i> dan Regresi Logistik dengan	
Konfigurasi 1, 2, dan 3.....	V-8

BAB 1

PENDAHULUAN

1.1. Pendahuluan

Bab 1 akan menjelaskan gambaran secara umum penelitian. Penjelasan dimulai dari latar belakang masalah, rumusan masalah, manfaat penelitian, Batasan penelitian hingga sistematika penulisan.

1.2. Latar Belakang Masalah

Klasifikasi teks atau klasifikasi dokumen merupakan salah satu tugas utama dalam *text data mining* dan *information retrieval* (Jain dan Jyoti, 2016). Klasifikasi digunakan untuk memprediksi keanggotaan kelompok terhadap suatu *data instances*.

Naive Bayes merupakan salah satu algoritma pembelajaran induktif yang paling efektif dan efisien dalam mengolah data (Zhang, 2004). *Naive Bayes* sering digunakan sebagai dasar dalam pengklasifikasian teks karena cepat dan mudah diimplementasikan (Rennie *et al.*, 2003). *Naive Bayes* adalah pengklasifikasi probabilistik sederhana berdasarkan teorema Bayes. Dalam prakteknya, kebebasan asumsi sering dilanggar, tetapi pada metode *Naive Bayes* masih cenderung dapat bekerja dengan sangat baik di bawah asumsi yang tidak realistis dan sangat populer hingga saat ini (Jain dan Jyoti, 2016). Untuk menangani tugas klasifikasi teks, dokumen akan ditandai dengan kata-kata yang muncul didalamnya (Zhang *et al.*, 2016).

Regresi logistik merupakan suatu model analisis statistik yang merupakan perkembangan lebih lanjut dari metode probabilitas linier. Regresi logistik digunakan untuk dapat menentukan estimasi probabilitas suatu fenomena dengan cara mereduksi kelemahan-kelemahan pada model probabilitas linier, regresi logistik digunakan untuk menguji probabilitas terjadinya variabel terikat dapat diprediksi dengan variabel bebasnya (Maski, 2010).

Pada era sekarang internet atau media sosial cukup banyak memberikan dampak terhadap masyarakat. Dampak yang ditimbulkan dari perkembangan teknologi bukan hanya berupa dampak positif tetapi juga dapat berdampak negatif terhadap masyarakat. Perkembangan kejahatan semakin meluas mulai dari internet *abuse, hacking, cracking*, hingga *hate speech* atau berupa bentuk ujaran kebencian. Salah satu jenis kejahatan yang sering terjadi di media sosial adalah *hate speech* yang sangat umum terjadi.

Hate speech adalah tindakan komunikasi yang menjelekkan dengan target individu atau karakteristik kelompok tertentu seperti ras, etnis, gender, agama, dan lain-lain. (Djuric *et al.*, 2015). Untuk itu perlu pengklasifikasian apakah suatu pernyataan atau ucapan termasuk dalam kelompok *hate speech* atau bukan. Maka perlu diuji algoritma mana yang lebih baik dalam pengklasifikasian *hate speech*, antara *Naïve Bayes* dengan Regresi Logistik.

Klasifikasi berdasarkan topik semantik, merupakan masalah komptasi yang sering dijumpai pada *information retrieval, machine learning*, serta pemrosesan bahasa alami. Klasifikasi *hate speech* didalam praktiknya diharapkan dapat digunakan untuk membantu sosial media dalam melacak ujaran kebencian, bahkan

diperlukan upaya preventif terhadap *hate speech* tersebut agar dapat menghindari *hate speech* tersebut dapat tampil pada sosial media. Sehingga analisa terhadap data *hate speech* dapat dilakukan secara ter program agar dapat digunakan untuk membuat model pembelajaran mesin yang dapat mengidentifikasi dengan akurat dalam mendeteksi *hate speech* untuk dapat dimanfaatkan secara luas (Kevin *et al.*, 2019). Penelitian yang dilakukan akan mengukur kinerja dari algoritma *Naïve Bayes* dan Regresi Logistik yang akan diketahui dari nilai akurasi dan waktu komputasi. Dataset berupa *tweet* dari *twitter* yang diambil dari penelitian terdahulu.

1.3. Perumusan Masalah

Rumusan masalah pada penelitian yang dilakukan berupa bagaimana perbandingan antara algoritma *Naïve Bayes* dengan algoritma regresi logistik untuk klasifikasi *hate speech*. Perbandingan dilakukan pada tingkat akurasi dan waktu komputasi yang paling efektif digunakan untuk mengidentifikasi terjadinya *hate speech*.

Untuk melakukan penelitian tersebut, maka rumusan masalah diatas diuraikan menjadi beberapa pertanyaan penelitian (*research question*) yaitu sebagai berikut :

1. Bagaimana mekanisme algoritma *Naïve Bayes* dan Regresi Logistik untuk mengidentifikasi *hate speech* ?
2. Apa persamaan dan perbedaan antara *Naïve Bayes* dengan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*?

3. Bagaimana cara mengukur kinerja algoritma *Naïve Bayes* dan algoritma Regresi Logistik dalam mengidentifikasi *hate speech* ?

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengetahui mekanisme algoritma *Naïve Bayes* dan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*.
2. Mengetahui persamaan dan perbedaan antara algoritma *Naïve Bayes* dengan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*.
3. Untuk mengetahui tolak ukur kinerja algoritma *Naïve Bayes* dan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*

1.5. Manfaat Penelitian

Manfaat dari penelitian adalah:

1. Memahami algoritma *Naive Bayes* dan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*.
2. Memahami perbandingan dan persamaan antara algoritma *Naive Bayes* dengan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*.
3. Memahami tolak ukur kinerja dari algoritma *Naive Bayes* dan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*.

1.6. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan menggunakan Bahasa Indonesia.
2. Data yang digunakan berupa teks.
3. Jumlah *tweet* yang digunakan untuk dataset adalah 700 *tweet*.
4. *Tweet* akan di klasifikasikan kedalam *hate speech* dan non *hate speech*.

1.7. Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I. PENDAHULUAN

Dalam bab ini menjelaskan mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan,

BAB II. KAJIAN LITERATUR

Pada bab ini menjelaskan mengenai beberapa algoritma yang digunakan yaitu algoritma *Naive Bayes* dan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*.

BAB III. METODELOGI PENELITIAN

Pada bab ini akan membahas mengenai tahapan yang akan dilakukan dalam penelitian. Masing-masing rencana tahapan penelitian akan dideskripsikan secara rinci dengan mengacu pada suatu kerangka kerja, selanjutnya pada akhir bab berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan membahas mengenai perancangan dan lingkungan implementasi algoritma *Naive Bayes* dan algoritma Regresi Logistik dalam mengidentifikasi *hate speech*, hasil *training* dan hasil *testing*.

BAB V. HASIL DAN PEMBAHASAN

Pada bab ini berisi kesimpulan dari berbagai uraian yang terdapat pada bab sebelumnya dan juga berisi saran-saran yang diharapkan dapat berguna dalam identifikasi *hate speech* menggunakan algoritma *Naive Bayes* dan algoritma Regresi Logistik.

1.8. Kesimpulan

Penelitian ini akan berfokus untuk melihat perbandingan hasil klasifikasi antara algoritma *Naive Bayes* dengan algoritma Regresi Logistik dalam mengidentifikasi *hate speech* yang didapat melalui hasil *testing*.

DAFTAR PUSTAKA

- Beel, J., Langer, S., & Gipp, B. (2017). TF-IDuF : A Novel Term-Weighting Scheme for User Modeling based on Users' Personal Document Collections. Proceedings of the tConference, 1-7.
- Darujati, C., & Gumelar, A.B. (2012). Pemanfaatan Teknik Supervised untuk Klasifikasi Teks Bahasa Indonesia. Jurnal Link, 5 (1)- 5 (8).
- Efendi, R., Malik, R. F., Milasari, J. (2012). Klasifikasi Dokumen Berbahasa Indonesia Menggunakan Naïve Bayes Classifier. Journal of Research in Computer Science and Applications, 7-13.
- Han, J., Kambor, M., & Pei, J. (2006). Data Mining : Concepts and Techniques. Morgan Kaufmann.
- Hosmer, D. W & Lemeshow, S (2000). Applied Logistic Regression. New York John Wiley & Sons, Inc.
- Nazief, B.A., & Adriani, M. (1996). Confix Stripping. Approach to Stemming Algorithm for Bahasa Indonesia. International Publication Faculty of Computer and Science, University of Indonesia, Depok, Jakarta. 41
- Patil, T. R., & Sherekar, R. R. (2013). Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification. International Journal of Computer Science And Applications, 256-261.

Pressman, R.S. (2005). *Software engineering: a practitioner's approach*. Palgrave Macmillan

Raharjo, S., & Winarko, E. (2014). Klasterisasi, Klasifikasi, dan Peringkasan Teks Berbahasa Indonesia. *Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2014)* (pp. 391-401). Depok: Universitas Gunadarma.

Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1725-1732.

Srividhya, V., & Anitha, R. (2010). Evaluating Preprocessing Techniques in Text Categorization. *International Journal of Computer Science and Application Issue*, 49- 51.

Tan, Ah-Hwen. (1999). Text Mining: The state of the art and the challenges. *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*