

Penerapan Algoritme *Naïve Bayes Classifiers* pada Sistem Pendeteksi Berita *Hoax* Berbahasa Indonesia

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Victor Ronaldo Pasaribu

NIM : 09021281722044

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN SKRIPSI

PENERAPAN ALGORITME NAÏVE BAYES CLASSIFIERS PADA SISTEM PENDETEKSI BERITA HOAX BERBAHASA INDONESIA

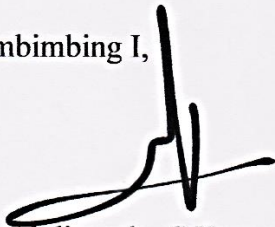
Oleh:

Victor Ronaldo Pasaribu

NIM : 09021281722044

Indralaya, 05 Juli 2021

Pembimbing I,



Dr. Abdiansah., S.Kom., M.Cs
NIP. 198410012009121005

Pembimbing II,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006043003

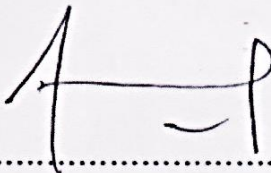
TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Rabu tanggal 23 Juni 2021 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Victor Ronaldo Pasaribu
NIM : 09021281722044
Judul : Penerapan Algoritme *Naïve Bayes Classifiers* pada Sistem Pendeteksi Berita *Hoax* Berbahasa Indonesia

1. Ketua Penguji

Muhammad Fachrurrozi, S.Si., M.T.
NIP. 198005222008121002



2. Pembimbing I

Dr. Abdiansah., S.Kom., M.Cs
NIP. 198410012009121005



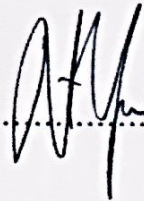
3. Pembimbing II

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



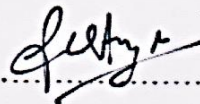
4. Penguji I

Novi Yusliani, M.T.
NIP. 198211082012122001



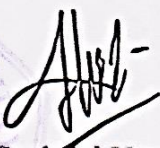
5. Penguji II

Desty Rodiah, M.T.
NIP. 198912212020122011



Mengetahui,
Ketua Jurusan Teknik Informatika




Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan dibawah ini:

Nama : Victor Ronaldo Pasaribu
NIM : 09021281722044
Program Studi : Teknik Informatika Reguler
Judul Skripsi : Penerapan Algoritme *Naïve Bayes Classifiers* pada Sistem Pendeteksi Berita *Hoax* Berbahasa Indonesia

Hasil Pengecekan Software *iThenticate/Turnitin* : 17%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek 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.



Indralaya, 05 Juli 2021



Victor Ronaldo Pasaribu
NIM. 09021281722044

MOTTO DAN PERSEMBAHAN

MOTTO:

Bersukacitalah dalam pengharapan, sabarlah dalam kesesakan, dan bertekunlah

dalam doa!

(Roma 12:12)

Kupersembahkan Karya Tulis ini kepada:

- Orang Tuaku
- Kakak dan Adikku
- Dosen Pembimbing
- Sahabat-sahabat
- Teman Seperjuanganku IF REG B 2017
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

APPLICATION OF NAÏVE BAYES CLASSIFIERS ALGORITHM ON NEWS DETECTION SYSTEM IN INDONESIAN

By :

Victor Ronaldo Pasaribu
09021281722044

ABSTRACT

Hoax is a recent issues that many people use as political weapons, irrelevant truth, or distributed purposely false news. Along with the growth of technology which is too fast, at this time fake news are very easily spread over through the internet, such as social media. Many cases of deception caused by the news that someone receives are fake news. Researches in this field has so much popped out of with variety of algorithms, however the application of algorithms in a hoax news detection system that can be used by everyone is still rare to find. Therefore, to make it easier for someone to know the category of news received, this research builds a hoax news detection system that applies Naive Bayes algorithm. The Naive Bayes Classifiers algorithm is a classification method that using probability and statistics to predict the odds based on existing data. In the application of this algorithm, been done by the test to determine the effect of the algorithm that used in this study. The test has was carried out 9 times on 1000 news texts as knowledge base. From the tested, was determined by the percentage of recall value is 82%, precision is 88%, accuracy is 78% and f-measure is 85%. With the results of that performance measurement, hopefully this hoax news detection system can be used as a reference in knowing the type of news text that will approve by someone.

Key Word: Hoax Detection System, Hoax, Naïve Bayes Classifiers.

PENERAPAN ALGORITME NAÏVE BAYES CLASSIFIERS PADA SISTEM PENDETEKSI BERITA HOAX BERBAHASA INDONESIA

Oleh :

Victor Ronaldo Pasaribu
09021281722044

ABSTRAK

Hoax merupakan isu-isu terkini yang digunakan banyak orang sebagai senjata politik, kebenaran yang tidak relevan atau kabar bohong yang disebarkan secara sengaja. Seiring dengan perkembangan teknologi yang begitu pesat, pada saat ini kabar bohong sangat mudah tersebar melalui internet, seperti halnya media sosial. Banyak terjadi kasus penipuan akibat kabar yang diterima seseorang adalah kabar bohong. Penelitian dalam bidang ini banyak bermunculan dengan berbagai macam algoritme, namun penerapan algoritme pada sebuah sistem pendeteksi berita *hoax* yang dapat digunakan oleh semua orang masih jarang ditemukan. Oleh karena itu, untuk memudahkan seseorang mengetahui kategori berita yang diterima, penelitian ini membangun sistem pendeteksi berita *hoax* yang menerapkan algoritme *Naive Bayes Classifiers*. Algoritme *Naive Bayes Classifiers* sendiri merupakan sebuah metode klasifikasi yang menggunakan probabilitas dan statistik untuk memprediksi peluang berdasarkan data yang ada. Dalam penerapan algoritme ini, dilakukan pengujian untuk mengetahui pengaruh dari algoritme yang digunakan pada penelitian ini. Pengujian dilakukan sebanyak 9 kali pada 1000 teks berita sebagai *knowledge base*. Dari pengujian yang dilakukan, didapatkan persentase nilai *recall* sebesar 82%, *precision* sebesar 88%, *accuracy* sebesar 78% dan *f-measure* sebesar 85%. Dengan hasil pengukuran kinerja tersebut diharapkan sistem pendeteksi berita *hoax* ini dapat dijadikan sebagai acuan dalam mengetahui jenis teks berita yang diterima seseorang.

Kata Kunci: Sistem Deteksi *Hoax*, *Hoax*, *Naive Bayes Classifiers*.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika pada Fakultas Ilmu Komputer, Universitas Sriwijaya.

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan bantuan dan dukungan baik dalam materil maupun moril selama proses pembuatan skripsi ini. Adapun penulis merujuk secara khusus pihak yang telah membantu sebagai berikut:

1. Orang tuaku Sarip Pasaribu dan Mimin Aminah serta saudara-saudaraku yang telah membesarkan, dan selalu mendoakan serta memberikan dukungan.
2. Bapak Jaidan Jauhari, MT selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya beserta jajarannya.
3. Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
4. Bapak Dr. Abdiansyah, S.Kom., M.Cs selaku dosen pembimbing I dan Ibu Alvi Syahrini Utami, M.Kom selaku pembimbing II yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan skripsi.
5. Bapak Osvari Arsalan, S.Kom, M.T selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan skripsi.

6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Kak Ricy dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Sisil, yang telah banyak membantu, memberikan dukungan selama proses pengerjaan skripsi hingga penulis dapat menyelesaikan skripsi dengan baik.
9. Fais, Abi, Ahmad, Haris, Rosyid serta teman-teman jurusan Teknik Informatika yang tidak dapat disebutkan satu persatu yang telah berbagi keluh kesah, motivasi, semangat, canda tawa selama masa perkuliahan.

Semoga skripsi ini dapat bermanfaat bagi penulis khususnya maupun pembaca pada umumnya. Serta dapat menjadi referensi dan rujukan bagi hal-hal yang bermanfaat. Penulis menyadari bahwa dalam proses penyelesaian skripsi ini, terdapat beberapa penjelasan yang kurang sempurna. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun agar skripsi ini menjadi karya tulis yang sempurna supaya terciptanya bekal pengetahuan yang baik bagi peneliti lain di masa depan.

Indralaya, Juli 2021

Victor Ronaldo Pasaribu

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xviii
 BAB I PENDAHULUAN	 I-1
1.1 Latar Belakang Masalah	I-1
1.2 Rumusan Masalah	I-3
1.3 Tujuan Penelitian	I-3
1.4 Manfaat Penelitian	I-3
1.5 Batasan Masalah	I-4
1.6 Sistematika Penulisan	I-4
1.7 Kesimpulan.....	I-6
 BAB II KAJIAN LITERATUR	 II-1
2.1 Pendahuluan	II-1
2.2 <i>Hoax</i>	II-1
2.3 <i>Preprocessing</i>	II-2
2.4 TF-IDF.....	II-3
2.5 <i>Naïve Bayes Classifiers</i>	II-5
2.6 Pengukuran Kinerja	II-8

2.7 Penelitian Lain yang Relevan	II-10
2.8 Kesimpulan.....	II-14
BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Pendahuluan	III-1
3.2 Unit Penelitian	III-1
3.3 Pengumpulan Data.....	III-1
3.3.1 Jenis dan Sumber Data.....	III-1
3.3.2 Metode Pengumpulan Data.....	III-3
3.4 Tahapan Penelitian	III-3
3.4.1 Menentukan Ruang Lingkup dan Unit Penelitian.....	III-3
3.4.2 Menentukan Dasar Teori yang Berkualitas dengan Penelitian	III-4
3.4.3 Menetapkan Kriteria Pengujian	III-4
3.4.4 Menentukan Alat yang Digunakan dalam Penelitian	III-6
3.4.5 Menentukan Format Data Pengujian	III-7
3.4.6 Melakukan Pengujian Penelitian	III-7
3.4.7 Menentukan Analisis Hasil Pengujian dan Membuat Kesimpulan	III-8
3.5 Metode Pengembangan Perangkat Lunak	III-9
3.5.1 Fase Insepsi.....	III-10
3.5.2 Fase Elaborasi	III-10
3.5.3 Fase Konstruksi.....	III-10
3.5.4 Fase Transisi	III-11
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis.....	IV-1
4.2.2 Kebutuhan Sistem	IV-2
4.2.3 Analisis dan Desain	IV-4
4.2.3.1 Analisis dan Kebutuhan Perangkat Lunak.....	IV-4
4.2.3.2 Analisis Data.....	IV-4
4.2.3.3 Analisis Pra-Pengolahan Data	IV-5
4.2.3.4 Analisis Pembobotan Kata TF-IDF	IV-10
4.2.3.5 Analisis Algoritme <i>Naïve Bayes Classifiers</i>	IV-13

4.2.3.6 Analisis <i>Confusion Matrix</i>	IV-14
4.2.3.7 Desain Perangkat Lunak	IV-14
4.3 Fase Elaborasi.....	IV-23
4.3.1 Pemodelan Bisnis.....	IV-23
4.3.2 Perancangan Data	IV-23
4.3.3 Kebutuhan Sistem	IV-23
4.3.4 Diagram	IV-24
4.3.3.1 Diagram Aktivitas.....	IV-24
4.3.3.2 <i>Sequence Diagram</i>	IV-27
4.3.3.3 Perancangan Antarmuka	IV-31
4.4 Fase Konstruksi	IV-31
4.4.1 Kebutuhan Sistem	IV-31
4.4.2 Diagram Kelas	IV-32
4.4.3 Implementasi.....	IV-32
4.5.3.1 Implementasi Kelas	IV-32
4.5.3.2 Implementasi Antarmuka.....	IV-33
4.5 Fase Transisi	IV-34
4.5.1 Pemodelan Bisnis.....	IV-34
4.5.2 Kebutuhan Sistem	IV-34
4.5.3 Rencana Pengujian.....	IV-35
4.5.3.1 Rencana Pengujian <i>Use Case</i> Melakukan <i>Input Data</i>	IV-35
4.5.3.2 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi dengan <i>Naïve Bayes Classifiers</i>	IV-35
4.5.3.3 Rencana Pengujian <i>Use Case</i> Melakukan <i>Clear Data</i>	IV-36
4.5.3.4 Rencana Pengujian <i>Use Case</i> Melakukan <i>Input Fold</i>	IV-36
4.5.3.5 Rencana Pengujian <i>Use Case</i> Melakukan Perhitungan Akurasi dengan <i>Confusion Matrix</i>	IV-37
4.5.3.6 Rencana Pengujian <i>Use Case</i> Melakukan <i>Close Sistem</i>	IV-37
4.5.4 Implementasi.....	IV-37
 BAB V HASIL DAN ANALISIS PENELITIAN.....	 V-1
5.1 Pendahuluan	V-1
5.2 Hasil Program	V-1
5.3 Hasil Pengujian.....	V-2
5.4 Analisis Hasil Pengujian.....	V-4

5.5 Kesimpulan.....	V-7
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran	VI-2
DAFTAR PUSTAKA	xix
LAMPIRAN	L-1

DAFTAR TABEL

	Halaman
Tabel II-1. <i>Confusion Matrix</i>	II-9
Tabel II-2. Data Penelitian Sebelumnya	II-12
Tabel III-1. Format Data Pengujian <i>Confusion Matrix</i> Hasil Klasifikasi	III-7
Tabel III-2. Hasil Pengujian Klasifikasi.....	III-8
Tabel IV-1. Kebutuhan Fungsional.....	IV-3
Tabel IV-2. Kebutuhan Non-Fungsional.....	IV-3
Tabel IV-3. Contoh Data Teks Berita	IV-5
Tabel IV-4. Hasil <i>Case Folding</i> dari Contoh Teks Berita	IV-6
Tabel IV-5. Hasil <i>Tokenizing</i> dari Contoh Teks Berita.....	IV-6
Tabel IV-6. Hasil <i>Filtering</i> dari Contoh Teks Berita.....	IV-8
Tabel IV-7. Hasil <i>Stemming</i> dari Contoh Teks Berita	IV-9
Tabel IV-8. Hasil TF-IDF dari Contoh Teks Berita.....	IV-10
Tabel IV-9. Hasil Pembobotan Kata dari Contoh Teks Berita.....	IV-12
Tabel IV-10. Contoh <i>Confusion Matrix</i>	IV-14
Tabel IV-11. Definisi Aktor	IV-16
Tabel IV-12. Definisi <i>Use Case</i>	IV-16
Tabel IV-13. Skenario <i>Use Case</i> Melakukan <i>Input Data</i>	IV-17
Tabel IV-14. Skenario <i>Use Case</i> Melakukan Klasifikasi dengan <i>Naïve Bayes</i> <i>Classifiers</i>	IV-18
Tabel IV-15. Skenario <i>Use Case</i> Melakukan <i>Clear Data</i>	IV-20
Tabel IV-16. Skenario <i>Use Case</i> Melakukan <i>Input Fold</i>	IV-20
Tabel IV-17. Skenario <i>Use Case</i> Melakukan Perhitungan Akurasi dengan <i>Confusion Matrix</i>	IV-21
Tabel IV-18. Skenario <i>Use Case</i> Melakukan <i>Close Sistem</i>	IV-22
Tabel IV-19. Implementasi Kelas	IV-33
Tabel IV-20. Rencana Pengujian <i>Use Case</i> Melakukan <i>Input Data</i>	IV-35

Tabel IV-21. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi dengan <i>Naïve Bayes Classifiers</i>	IV-35
Tabel IV-22. Rencana Pengujian <i>Use Case</i> Melakukan <i>Clear Data</i>	IV-36
Tabel IV-23. Rencana Pengujian <i>Use Case</i> Melakukan <i>Input Fold</i>	IV-36
Tabel IV-24. Rencana Pengujian <i>Use Case</i> Melakukan Perhitungan Akurasi dengan <i>Confusion Matrix</i>	IV-37
Tabel IV-25. Rencana Pengujian <i>Use Case</i> Melakukan <i>Close Sistem</i>	IV-37
Tabel IV-26. Rencana Pengujian <i>Use Case</i> Melakukan <i>Input Data</i>	IV-38
Tabel IV-27. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi dengan <i>Naïve Bayes Classifiers</i>	IV-38
Tabel IV-28. Rencana Pengujian <i>Use Case</i> Melakukan <i>Clear Data</i>	IV-39
Tabel IV-29. Rencana Pengujian <i>Use Case</i> Melakukan <i>Input Fold</i>	IV-40
Tabel IV-30. Rencana Pengujian <i>Use Case</i> Melakukan Perhitungan Akurasi dengan <i>Confusion Matrix</i>	IV-41
Tabel IV-31. Rencana Pengujian <i>Use Case</i> Melakukan <i>Close Sistem</i>	IV-41
Tabel V-1. <i>Confusion Matrix</i> Hasil Pengujian Klasifikasi	V-3
Tabel V-2. Hasil Klasifikasi Pengujian	V-4
Tabel V-3. Perbandingan <i>Naïve Bayes Classifiers</i> dan <i>Support Vector Machine</i>	V-6

DAFTAR GAMBAR

	Halaman
Gambar III-1 Tampilan awal turnbackhoax.id.....	III-2
Gambar III-2 Tampilan awal cekfakta.com	III-2
Gambar III-3 Diagram Alur Sistem Pendeteksi Berita <i>Hoax</i>	III-5
Gambar III-5 Arsitektur <i>Rational Unified Process</i> (RUP)	III-9
Gambar IV-1. Diagram <i>Use Case</i>	IV-15
Gambar IV-2. Diagram Aktivitas <i>Input Data</i>	IV-24
Gambar IV-3. Diagram Aktivitas Klasifikasi dengan <i>Naïve Bayes</i> <i>Classifiers</i>	IV-25
Gambar IV-4. Diagram Aktivitas <i>Clear Data</i>	IV-25
Gambar IV-5. Diagram Aktivitas <i>Input Fold</i>	IV-26
Gambar IV-6. Diagram Aktivitas Perhitungan Akurasi dengan <i>Confusion</i> <i>Matrix</i>	IV-26
Gambar IV-7. Diagram Aktivitas <i>Close Sistem</i>	IV-26
Gambar IV-8. Diagram Alur <i>Input Data</i>	IV-27
Gambar IV-9. Diagram Alur Klasifikasi dengan <i>Naïve Bayes Classifiers</i>	IV-28
Gambar IV-10. Diagram Alur <i>Clear Data</i>	IV-29
Gambar IV-11. Diagram Alur <i>Input Fold</i>	IV-29
Gambar IV-12. Diagram Alur Perhitungan Akurasi dengan <i>Confusion Matrix</i>	IV-30
Gambar IV-13. Diagram Alur <i>Close Sistem</i>	IV-30
Gambar IV-14. Perancangan Antarmuka	IV-31
Gambar IV-15. Diagram Kelas Perangkat Lunak	IV-32
Gambar IV-16. Implementasi Antarmuka.....	IV-34
Gambar V-1. Hasil Pengujian Sistem Cek Akurasi	V-2
Gambar V-2. Hasil Pengujian Sistem Cek Berita	V-2
Gambar V-3. Diagram Analisis Hasil Pengujian	V-5

Gambar V-4. Diagram Perbandingan <i>Naïve Bayes Classifiers</i> dan <i>Support</i>	
<i>Vector Machine</i>	V-7

DAFTAR LAMPIRAN

	Halaman
Tabel L-1. Data Teks Berita <i>Hoax</i> dan <i>Non Hoax</i>	L-1
Tabel L-2. 5 Contoh Data <i>True Positive</i> (TP).....	L-53
Tabel L-3. 5 Contoh Data <i>False Positive</i> (FP).....	L-54
Tabel L-4. 5 Contoh Data <i>True Negative</i> (TN).....	L-56
Tabel L-5. 5 Contoh Data <i>False Negative</i> (FN).....	L-58
Tabel L-6. Kode Program	L-62

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Hoax merupakan isu-isu terkini yang digunakan sebagai senjata politik, kebenaran yang tidak relevan, atau kabar bohong yang disebarkan secara sengaja (Berghel, 2017). Kabar-kabar bohong tersebut apabila tersebar secara luas akan mempengaruhi banyak orang dan dapat menodai suatu citra dan kredibilitas (Chen, Yong, dan Ishak, 2014). *Hoax* sendiri dapat bertujuan untuk mempengaruhi pembaca untuk mengambil tindakan sesuai dengan informasi palsu yang didapat. Informasi palsu dan menyesatkan ini berbahaya karena dapat menyesatkan persepsi manusia dengan menyampaikan informasi yang tidak benar sebagai sebuah kebenaran sehingga sangat memungkinkan untuk membawa dampak pada pemikiran manusia (Rasywir dan Purwarianti, 2016).

Berita *hoax* lebih populer dan tersebar luas melalui media sosial (Balmas, 2014). Jumlah penyebaran informasi yang disebarkan menggunakan media sosial berjumlah ratusan bahkan ribuan tiap harinya (Juliaswara, 2017). *Hoax* dapat mudah disebarkan karena bergantung pada pembaca yang secara sengaja mengirimkan berita tersebut pada korban yang berpotensi melakukan hal serupa (Hintzbergen et al., 2010). Beberapa tahun belakangan ini, untuk membantu pengguna online mengidentifikasi informasi yang memiliki nilai dan manfaat, telah ada penelitian yang luas tentang pembentukan sebuah *framework* yang otomatis dan efektif dalam deteksi berita *hoax* online (Shu et al., 2017).

Algoritme klasifikasi yang populer seperti *Naïve Bayes Classifiers* (NBC), dan *Support Vector Machine* (SVM) banyak digunakan oleh peneliti untuk deteksi berita *hoax*. *Support Vector Machine* (SVM) memiliki kelebihan seperti keakuratan dan kinerja yang baik pada kumpulan data yang kecil, fleksibel karena dapat untuk mengklasifikasikan data yang berdimensi tinggi dan cenderung efisien dalam hal memori (Ray et al., 2017). Akan tetapi algoritme ini memiliki kesulitan dengan *dataset* yang besar karena waktu *training* yang dibutuhkan bisa saja tinggi sehingga itu kurang efektif pada *dataset* yang *noise* (Brambrick, 2018).

Naïve Bayes Classifiers (NBC) dapat mengatasi masalah pada SVM yang kesulitan dengan *dataset* yang besar karena memang algoritme NBC membutuhkan data learning yang lebih banyak agar mendapatkan hasil yang baik dan algoritme ini juga dapat menggunakan pengukuran efektifitas *Confusion Matrix* seperti algoritme klasifikasi yang lain untuk mengetahui tingkat akurasi dari NBC dalam mengklasifikasikan teks berita. Namun kekurangan algoritme NBC ini masih butuh penyempurnaan algoritme dalam tokenisasi kalimat untuk menghasilkan klasifikasi dengan akurasi yang lebih tinggi (Informatikalogi, 2017).

Berita *hoax* dapat dianalisis dari kata-kata yang ada di dalam berita dengan cara membuat aturan untuk mengetahui berita tersebut termasuk kategori *hoax* atau *non hoax*. Akan tetapi, membuat aturan dengan cara manual dari sebuah berita tidak mudah karena banyak mengandung kata-kata sehingga membutuhkan waktu yang lama. Oleh karena itu pada penelitian ini akan menerapkan algoritme *Naïve Bayes Classifiers* (NBC) untuk melakukan klasifikasi berita kemudian hasil dari klasifikasi tersebut akan dilakukan pengukuran efektifitas dengan menggunakan

confusion matrix untuk mengetahui tingkat akurasi dari algoritme NBC dalam mengklasifikasikan teks berita.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah dijelaskan pada latar belakang maka rumusan masalah dari penelitian ini adalah:

1. Bagaimana penerapan algoritme *Naïve Bayes Classifiers* dalam mendeteksi berita *hoax* dan *non hoax*.
2. Bagaimana cara mengukur persentase dari tingkat keakuratan sistem pendeteksi *hoax* dalam mengklasifikasikan berita *hoax* dan *non hoax*.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membuat perangkat lunak dengan menerapkan algoritme *Naïve Bayes Classifiers* untuk mendeteksi berita *hoax* dan *non hoax*.
2. Mengukur kinerja perangkat lunak dengan menggunakan pengukuran efektifitas *confusion matrix* pada hasil klasifikasi algoritme *Naïve Bayes Classifiers*.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah dapat membantu setiap orang untuk mengetahui kebenaran berita yang diterima, sehingga dapat memberikan informasi apakah berita yang diterima termasuk kategori *hoax* atau *non hoax*.

1.5 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Tidak akan menerjemahkan kata apabila ada masukan yang dianggap berbahasa asing.
2. *Tools* yang digunakan adalah bahasa pemrograman *Python* dan *database* MySQL.

1.6 Sistematika Penulisan

Penulisan skripsi ini akan disusun mengikuti sistematika penulisan skripsi Fakultas Ilmu Komputer Universitas Sriwijaya, diantaranya:

BAB I. PENDAHULUAN

Adapun pendahuluan bab ini akan menguraikan pokok-pokok pemikiran yang dilandasi rancangan penelitian, diantaranya latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Adapun kajian literatur bab ini akan menjelaskan dasar-dasar teori yang dipergunakan dalam penelitian ini, seperti definisi *hoax*, *preprocessing*, TF-IDF, algoritme *Naïve Bayes Classifiers*, *confusion matrix*, serta beberapa kajian literatur mengenai penelitian lain yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Adapun metodologi penelitian bab ini akan membahas terkait tahapan-tahapan yang kelak dilaksanakan pada penelitian ini, diantaranya pengumpulan data, analisis data, serta perancangan sistem yang akan dibangun. Masing-masing rencana tahapan penelitian ini dideskripsikan secara detail dan berfokus pada suatu kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Adapun uraian bab ini akan dibahas mengenai analisis dan perancangan perangkat lunak yang kelak digunakan sebagai alat penelitian. Diawali dengan pengumpulan dan analisa kebutuhan, rancangan dan konstruksi perangkat lunak serta pengujian untuk memastikan semua kebutuhan pengembangan perangkat lunak sesuai dengan dengan kebutuhan.

BAB V. HASIL DAN ANALISA PENELITIAN

Adapun hasil dan analisa penelitian bab ini menguraikan hasil pengujian berdasarkan rencana langkah-langkah yang telah dibuat. Penyajian tabel hasil dari pengujian serta analisisnya ditunjukkan sebagai dasar dari kesimpulan yang akan diambilkan dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Adapun uraian kesimplan dari bab ini berisikan kesimpulan dari uraian pokok-pokok penting berdasarkan bab-bab sebelumnya dan juga terdapat saran yang diberikan dari hasil penelitian yang telah dilakukan.

1.7 Kesimpulan

Dari bab ini, telah dijabarkan secara umum pokok-pokok pikiran yang melandasi penelitian ini yang meliputi latar belakang dan rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

DAFTAR PUSTAKA

- Balmas, M. 2014. *When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism*. Communication Research 41(3):430–454.
- Berghel, H. 2017. “*Alt-News and Post-Truths in the “Fake News” Era*”. IEEE Computer 50(4):110-114.
- Brambrick, Aylien, N. (n.d.). KDnuggets. Retrieved February 20, 2018, from <https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html>. Diakses pada tanggal 23 Februari 2020 pukul 20.00 WIB.
- Bustami, B. (2013). Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. TECHSI-Jurnal Teknik Informatika, 5(2).
- Chen, Y. Y., Yong, S. P., and Ishak, A. 2014. *Email Hoax Detection System Using Levenshtein Distance Method*, Journal of Computers Volume 2 Number 2, Academy Publisher.
- Hintzbergen, J., Hintzbergen, K., Smulders, A., and Baars, H. 2010. Foundations of Information Security Based on ISO27001 and ISO27002 (2nd Edition). Zaltbommel: Van Haren Publishing, Hertogenbosch, Netherlands.
- Informatikalogi, "Informatikalogi.com," 27 November 2017. Tersedia: <https://informatikalogi.com/text-preprocessing/>. Diakses 23 Februari 2020 pukul 20.35 WIB.

- Juliaswara, V. 2017. “Mengembangkan Model Literasi Media yang Berkebhinekaan dalam Menganalisis Informasi Berita Palsu (Hoax) di Media Sosial”. *Jurnal Pemikiran Sosiologi* 4(2):65-68.
- Kroll, P., and Kruchten, P. 2003. *The Rational Unified Process Made Easy: A Practitioner's Guide to the RUP: A Practitioner's Guide to the RUP*. Addison-Wesley Professional.
- Nugroho, A. S. 2020. Tips Menghadapi Hoax dan Spam. Tersedia: www.ilmukomputer.com. Diakses 15 April 2021 pukul 20.10 WIB.
- Patil, T. R., and Sherekar, S. S. 2013. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International journal of computer science and applications*, 6(2), 256-261.
- Pattekari, S. A., & Parveen, A. 2012. Prediction system for heart disease using Naïve Bayes. *International Journal of Advanced Computer and Mathematical Sciences*, 3(3), 290-294.
- Prasetyo, E. 2012. *Data Mining: Konsep Dan Aplikasi Menggunakan MATLAB*. Penerbit ANDI. Yogyakarta.
- Rasywir, E dan Purwarianti, A. 2016. “Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin”. *Jurnal Cybermatika* 3(2):1-8.
- Ray, S., Srivastava, T., Dar, P., dan Shaikh, F. 2017. *Understanding Support Vector Machine Algorithm from Examples (along with code)*. Retrieved March 2, 2018, from <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>.

- Ridwan, M., Suyono, H., dan Sarosa, M. 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *jurnal EECCIS*, 7(1), 59-64.
- Saadah, M. N., Atmagi, R. W., Rahayu, D. S., & Arifin, A. Z. (2013). Information Retrieval Of Text Document With Weighting TF-IDF And LCS. *Jurnal Ilmu Komputer dan Informasi*, 6(1), 34-37.
- Shu, K., Sliva, A., Wang, S., Tang, J., dan Liu, H. 2017. *Fake News Detection on Social Media: A Data Mining Perspective*. ACM SIGKDD Explorations Newsletter 19(1):22–36.
- Silverman, C. 2015. Lies, damn lies and viral content. Tersedia: <https://academiccommons.columbia.edu/doi/10.7916/D8Q81RHH>. Diakses 15 April 2021 pukul 20.05 WIB
- Sokolova, M., and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4), 427-437.
- Xhemali, D., Hinde, C., and Stone, R. 2009. Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages", *International Journal of Computer Science Issues*, IJCSI, 4(1), 16-23.
- Zaman, B., dan Winarko, E. 2011. Analisis Fitur Kalimat untuk Peringkat Teks otomatis pada Bahasa Indonesia. *Indonesian Journal of Computing and Cybernetics Systems* 5(2):60-68.