

**SISTEM KLASIFIKASI JUDUL BERITA *CLICKBAIT*
BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA
NAÏVE BAYES CLASSIFIER DAN SELEKSI FITUR CHI-SQUARE**

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



Oleh :

Nanda Tirana Mahesa
NIM: 09021381722099

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2022

LEMBAR PENGESAHAN TUGAS AKHIR

SISTEM KLASIFIKASI JUDUL BERITA *CLICKBAIT* BERBAHASA
INDONESIA MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER*
DAN SELEKSI FITUR *CHI SQUARE*

Oleh :

Nanda Tirana Mahesa
NIM : 09021381722099

Palembang, 21 Januari 2022

Pembimbing I

Novi Yusliani, M.T.
NIP. 198211082012122001

Pembimbing II

Danny Matthew Saputra, M.Cs.
NIP. 198505702015041002

Mengetahui,

Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

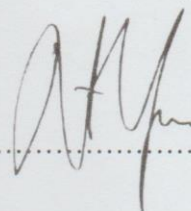
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa tanggal 4 Januari 2021 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Nanda Tirana Mahesa
NIM : 09021381722099
Judul : Sistem Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan Algoritma *Naïve Bayes Classifier* dan Seleksi Fitur *Chi Square*

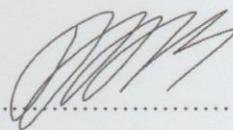
1. Pembimbing I

Novi Yusliani, M.T.
NIP. 198211082012122001



2. Pembimbing II

Danny Matthew Saputra, M.Cs.
NIP. 199110102018032001



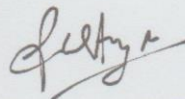
3. Penguji I

Alvi Syahrini Utami, M.Kom.
NIP. 198306062015042002

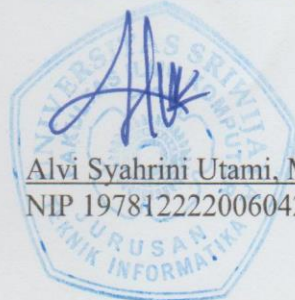


4. Penguji II

Desty Rodiah, M.T.
NIP. 198912212020122011



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Nanda Tirana Mahesa
NIM : 09021381722099
Program Studi : Teknik Informatika
Judul Skripsi : Sistem Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan Algoritma *Naïve Bayes Classifier* dan Seleksi Fitur *Chi Square*

Hasil Pengecekan *Software Ithenticate/Turnitin* : 11%

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

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



Palembang, Januari 2022



Nanda Tirana Mahesa
NIM. 09021381722099

MOTTO DAN PERSEMBAHAN

Motto :

“Sesungguhnya beserta kesulitan itu ada kemudahan.” (Q.S: Al-Insyirah: 6)

“Anythings’ possible if you’ve got enough nerve”

- J.K. Rowling

“#JalaninAja”

-Panitia Acara PBGF 2018

Kupersembahkan Karya Tulis ini kepada:

- *Allah SWT*
- *Kedua Orang Tuaku*
- *Kakakku dan kedua adikku*
- *Keluarga Besar Tercinta*
- *Sahabat dan kawan seperjuangan*
- *Universitas Sriwijaya*

**CLICKBAIT NEWS HEADLINE CLASSIFICATION SYSTEM USING
NAÏVE BAYES CLASSIFIER ALGORITHM AND CHI SQUARE
FEATURE SELECTION**

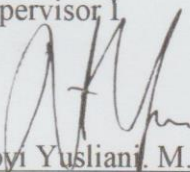
By:
Nanda Tirana Mahesa
09021381722099

ABSTRACT

Clickbait refers to a type of web content ad designed to entice readers to click on a link. Not a few news that misuse clickbait by providing news content that does not match the title. Such news has the potential to harm readers so that clickbait news title classification system is needed to make it easier for readers to filter news. In this study, clickbait headlines were classified using the Naïve Bayes Classifier algorithm and chi square feature selection. Feature selection itself is done based on the division of feature ratios as much as 25%, 50%, 75%, and 100% of all features. There are two processes: using stopwords removal in the pre-process and no stopwords removal. Systems without stopwords removal result in average accuracy values of 80.73%, 80.80%, 79.15%, and 78.33% respectively with best accuracy achieved at a feature ratio of 50%. While with stopwords removal resulted in an average accuracy of 66.29%, 67.25%, 65.73%, and 67.71% with the best accuracy at a feature ratio of 100%. The test results showed that the selection of chi square features of each ratio affected the amount of data that could be classified and the computing time and stopwords removal had an effect on system performance results.

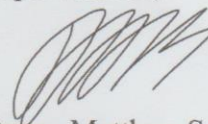
Key Word : Classification, Clickbait, Naïve Bayes Classifier, Feature Selection, Chi Square

Supervisor I



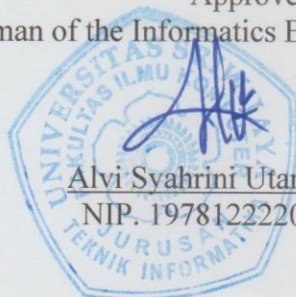
Novi Yusliani, M.T.
NIP. 198211082012122001

Palembang, Januari 2022
Supervisor II,



Danny Matthew Saputra, M.Cs.
NIP. 198505702015041002

Approved,
Chairman of the Informatics Engineering Department



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

**SISTEM KLASIFIKASI JUDUL BERITA MENGGUNAKAN
ALGORITMA NAÏVE BAYES CLASSIFIER DAN SELEKSI FITUR CHI
SQUARE**

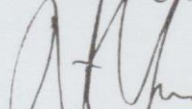
**Oleh:
Nanda Tirana Mahesa
09021381722099**

ABSTRAK

Clickbait mengacu pada jenis iklan konten web yang dirancang untuk memikat pembaca agar mengklik tautan. Tidak sedikit berita yang menyalahgunakan *clickbait* dengan memberikan konten berita yang tidak sesuai dengan judulnya. Berita yang seperti itu berpotensi merugikan pembaca sehingga dibutuhkan sistem klasifikasi judul berita *clickbait* agar memudahkan pembaca dalam memfilter berita. Pada penelitian ini, judul berita *clickbait* diklasifikasikan menggunakan metode *Naïve Bayes Classifier* dan seleksi fitur *Chi Square*. Seleksi fitur sendiri dilakukan berdasarkan pembagian rasio fitur sebanyak 25%, 50%, 75%, dan 100% dari seluruh fitur. Dilakukan dua proses yaitu terdapat penghapusan *stopwords* dalam pra-prosesnya dan tanpa penghapusan *stopwords*. Sistem tanpa penghapusan *stopwords* menghasilkan nilai rata-rata *accuracy* masing-masing dari rasio fitur sebesar 80,73%, 80,80%, 79,15%, dan 78,33% dengan *accuracy* terbaik dicapai pada rasio 50%. Sedangkan dengan penghapusan *stopwords* menghasilkan rata-rata *accuracy* sebesar 66,29%, 67,25%, 65,73%, dan 67,71% dengan *accuracy* terbaik pada rasio 100%. Hasil pengujian menunjukkan seleksi fitur *chi square* setiap rasio mempengaruhi jumlah data yang bisa terklasifikasi dan waktu komputasi serta *stopwords removal* memberikan pengaruh pada hasil kinerja sistem.

Kata Kunci : Klasifikasi, *Clickbait*, *Naïve Bayes Classifier*, Seleksi Fitur, *Chi Square*

Pembimbing I,

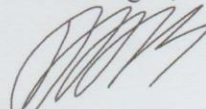


Novi Yulistiani, M.T.

NIP. 198211082012122001

Palembang, Januari 2022

Pembimbing II,



Danny Matthew Saputra, M.Cs.

NIP. 198505702015041002

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003

KATA PENGANTAR

Sujud syukur kusembahkan kepada Allah ta'ala, Tuhan Yang Maha Agung dan Maha Tinggi. Atas izin-Nya penulis bisa menyelesaikan Tugas Akhir yang berjudul “**Sistem Klasifikasi Judul Berita *Clickbait* Berbahasa Indonesia Menggunakan Algoritma *Naïve Bayes Classifier* dan Seleksi Fitur *Chi Square***” ini. Tugas Akhir ini disusun untuk memenuhi syarat kelulusan demi menyanggah gelar Sarjana Komputer program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini, banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih kepada:

1. Kedua orang tua tercinta, Ayah Abdul Fatah dan Ibu Noerfima, yang memberikan dukungan tiada henti, doa yang tak ada habisnya, menjadi rumah untuk pulang, dan kasih sayang sehingga membuat penulis tetap kuat hingga saat ini.
2. Kakak dan adikku tersayang, Ayuk Icha, Rafly, dan Zanza, yang selalu menyalurkan energi positif dengan memberikan guyonan dan lawakan sehingga penulis selalu tertawa dan selalu bersyukur memiliki kalian.
3. Pak Jaidan Jauhari, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, pak Samsuryadi, M.Kom., Ph.D., pak Julian Supardi, M.T., pak

Mgs. Afriyan Firdaus. S.Si., MIT, serta pak Fathoni. ST., MMSI selaku Wakil Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.

4. Ibu Novi Yusliani, M.T. dan Pak Danny Matthew Saputra, M.Cs. selaku pembimbing I dan pembimbing II yang sudah memberikan banyak bantuan, bimbingan, dan arahan sehingga penulis bisa menyelesaikan tugas akhir ini dengan baik.
5. Penguji Tugas Akhir, ibu Alvi Syahrini Utami, M.Kom. dan ibu Desty Rodiah, M.T. yang sudah meluangkan waktunya dan memberikan saran serta masukan kepada penulis. Lalu tak lupa pak Dr. Abdiansyah, M.Cs. selaku ketua penguji penulis.
6. Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Ibu Nabila Rizky Oktadini, M.T. selalu dosen pembimbing akademik yang sudah dengan sabar memberikan bimbingan kepada penulis setiap kali penulis kebingungan selama perkuliahan.
8. Mba Wiwin yang selalu siap sedia menjawab pertanyaan dan telah membantu penulis dalam kelancaran proses administrasi selama masa perkuliahan.
9. Salah satu *support system*, Kak Theo Rosadi, yang sudah menemani penulis, menjadi penghibur, tempat berkeluh kesah, penyemangat, pendengar dan pemberi saran yang sangat baik, partner bertengkar, *movie date*, dan menggendut bersama.
10. Teddy Septian R, *my 911 call, human diary*. Sahabat terbaik yang masih setia dan tidak pernah pergi sejak SMP, teman bercerita semua hal selama 24/7,

pendengar dan pemberi saran yang hebat. Serta sahabatku lainnya, Ndut, Alqo, dan Faris, terima kasih sudah hadir dan ada.

11. Kak Malian Zikri, *leader* bukan bos, *brother from different mother*. Terima kasih atas semua bimbingan dan jasanya, serta kesempatan yang diberikan sehingga penulis banyak mendapatkan pengalaman untuk bekal di dunia kerja kedepannya, dan semua kebaikan-kebaikan yang penulis terima. Lalu kepada *team*, Salman, Taufik, dan Kresna, yang sudah penulis anggap seperti keluarga sendiri, siap nge-*back up*, dan selalu memberikan tawa ketika lelah.
12. Destia Asri Felliani, kakakku lainnya, *my unbiological sister*. Terima kasih sudah banyak membantu bahkan tanpa diminta, menjadi tempat bercerita, dan semua kebaikan yang penulis rasakan dan tidak bisa disebutkan semuanya.
13. Ciwi ciwi; Kak Dedes, Aisyah, Mardiana, Dinda, Haliza, dan Ecik, yang sudah mau menjadi teman baik penulis selama di perantauan, sirkel berbagi info promo, nongkrong, sahabat saling menguat. Mungkin setelah ini kita akan sibuk dengan kehidupan masing-masing, namun tetap, aku menyayangi kalian.
14. Yasmin dan Mega, sahabat berbagi keluh kesah baik selama kuliah dan penyusunan tugas akhir, selalu siap untuk direpotkan, yang tidak pernah lepas memberikan semangat dan dorongan.
15. Sahabat seperjuangan Teknik Informatika Bilingual; Hafif, Yasmin, Mega, Chia, Fitri, Hani, Muti, Yani, Suci, Putri, Fajar, Adit, Kak Iwan, Jodi, Rizky, Beben, Isan, Vandi, Dede, Maman, Septi, Sausan, Bunga, Wiwik, dan semuanya yang tidak bisa disebut satu persatu yang sudah banyak membantu selama dan di akhir perkuliahan.

16. Kakak-kakak yang penulis harap bisa akrab dari dulu, Kak Dina Mayang Sari dan Kak Bill Brie, terima kasih semua kalimat-kalimat manisnya dan membuat penulis semakin yakin bahwa masih banyak sekali orang baik di dunia ini.
17. BEM KM Fasilkom Unsri 2018 dan 2019, tempat penulis banyak bertemu orang-orang baik dan hebat di perantauan, tempat mendapatkan pengalaman dan kenangan yang sangat berharga. Senang sekali bisa bertemu kalian, Fatoni, Wiranto, Abi, kak Rabbani, kak Barlian, kak Tami, kak Farid, kak Alif, kak Ko, kak Farah, kak Herlan, dan lain-lain yang tidak bisa disebutkan semuanya. Anak-anak PPSDM-ku tersayang; Ditya, Sena, Tasya, Alen, Iqbal, dan Reyhan.
18. Kak Wira, kakak kosan yang sudah sangat baik memberikan bantuan kepada penulis berupa pinjaman fasilitas yang sangat berguna dalam proses penyelesaian tugas akhir.
19. Raihan Abdurrahim, partner berbagi tawa dan drama di dunia virtual, penghilang penat dari beratnya tekanan skripsi. Juga teman-teman Payne Family; Adam, Rapi, Dimas, Dapa, Ravi, Adit, Alip, Deni, Willy, Sun, Ferhard, yang sudah mau menerima penulis dan karakternya di dalam sirkel kalian, terima kasih selalu berhasil membuat penulis merasa bahagia.
20. Dan diriku sendiri. Terima kasih sudah berjuang.

Palembang, Januari 2022

Nanda Tirana Mahesa

DAFTAR ISI

Halaman

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR	vi
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	vi
HALAMAN PERNYATAAN	iv
MOTTO DAN PESEMBAHAN.....	vi
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xix
DAFTAR LAMPIRAN.....	xxi
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-5
1.5 Manfaat Penelitian	I-5
1.6 Batasan Masalah.....	I-6
1.7 Sistematika Penulisan	I-6
1.8 Kesimpulan	I-8
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 <i>Clickbait</i>	II-1
2.2.2 <i>Text Mining</i>	II-2
2.2.3 <i>Text Preprocessing</i>	II-4
2.2.4 <i>Naive Bayes Classifier</i>	II-7

2.2.5	Seleksi Fitur <i>Chi-Square</i>	II-9
2.2.6	<i>Confusion Matrix</i>	II-12
2.2.7	<i>Rational Unified Process</i>	II-14
2.3	Penelitian Lain Yang Relevan.....	II-16
2.4	Kesimpulan	II-20
BAB III METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis dan Sumber Data	III-1
3.2.2	Metode Pengumpulan Data	III-1
3.3	Tahapan Penelitian	III-2
3.3.1	Kerangka Kerja	III-2
3.3.2	Kriteria Pengujian	III-6
3.3.3	Format Data Pengujian.....	III-6
3.3.4	Alat yang Digunakan dalam Pelaksanaan Penelitian	III-7
3.3.5	Pengujian Penelitian.....	III-8
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-8
3.4	Metode Pengembangan Perangkat Lunak	III-9
3.4.1	Fase Insepsi	III-9
3.4.2	Fase Elaborasi	III-10
3.4.3	Fase Konstruksi	III-10
3.4.4	Fase Transisi	III-11
3.5	Manajemen Proyek Penelitian	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 Kebutuhan Perangkat Lunak	IV-42
4.2.3.2	Analisis Data	IV-5
4.2.3.3	Analisis Pra-pemrosesan Data.....	IV-5
4.2.3.4	Analisis Seleksi Fitur	IV-9

	4.2.3.5 Analisis <i>TF-DF</i>	IV-14
	4.2.3.6 Analisis <i>Naïve Bayes Classifier</i>	IV-16
	4.2.3.7 Analisis <i>Confusion Matrix</i>	IV-20
	4.2.3.8 Desain Perangkat Lunak.....	IV-21
4.3	Fase Elaborasi	IV-29
	4.3.2 Pemodelan Bisnis	IV-29
	4.3.3 Perancangan Data.....	IV-29
	4.3.4 Perancangan Antarmuka	IV-30
	4.3.5 Kebutuhan Sistem	IV-33
	4.3.6 Diagram Aktivitas	IV-34
	4.3.7 Diagram <i>Sequence</i>	IV-38
4.4	Fase Konstruksi.....	IV-41
	4.4.1 Kebutuhan Sistem	IV-41
	4.4.2 Diagram Kelas.....	IV-41
	4.4.3 Implementasi	IV-42
	4.4.3.1 Implementasi Kelas	IV-42
	4.4.3.2 Implementasi Antarmuka	IV-45
4.5	Fase Transisi	IV-48
	4.5.1 Pemodelan Bisnis.....	IV-48
	4.5.2 Kebutuhan Sistem	IV-48
	4.5.3 Rencana Pengujian.....	IV-49
	4.5.3.1 Skenario Pengujian <i>Use Case</i> Memilih Data Masukan	IV-50
	4.5.3.2 Skenario Pengujian <i>Use Case</i> Melatih Data dan Seleksi Fitur Menggunakan <i>Chi Square</i>	IV-50
	4.5.3.3 Skenario Pengujian <i>Use Case</i> Menguji Sistem Klasifikasi Judul Berita	IV-51
	4.5.4 Implementasi	IV-51
	4.5.4.1 Pengujian <i>Use Case</i> Memilih Data Masukan.	IV-52
	4.5.4.2 Pengujian <i>Use Case</i> Melatih Data dan Seleksi Fitur Menggunakan <i>Chi Square</i>	IV-53
	4.5.4.3 Pengujian <i>Use Case</i> Menguji Sistem Klasifikasi Judul Berita.....	IV-54
4.6	Kesimpulan	IV-55

BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Penelitian.....	V-1
5.2.1 Konfigurasi Percobaan	V-1
5.2.2 Data Hasil Konfigurasi.....	V-2
5.3 Analisis Hasil Penelitian	V-8
5.4 Kesimpulan	V-16
 BAB VI KESIMPULAN DAN SARAN	 VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran.....	VI-3

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Halaman

Tabel II-1.	Contoh Proses <i>Case Folding</i>	II-5
Tabel II-2.	Contoh Proses <i>Tokenizing</i>	II-5
Tabel II-3.	Proses <i>Stopwords Removal</i>	II-6
Tabel II-4.	Contoh Proses <i>Stemming</i>	II-7
Tabel II-5.	Tabel Kontingensi	II-11
Tabel II-6.	Tabel <i>Confusion Matrix</i> (Jurafsky dan Martin, 2020).....	II-12
Tabel III-1.	Rancangan Tabel <i>Confusion Matrix</i>	III-7
Tabel III-2.	Hasil Pengujian <i>Confusion Matrix</i>	III-9
Tabel III-3.	Tabel Manajemen Proyek Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS).....	III-12
Tabel IV-1.	Kebutuhan Fungsional.....	IV-3
Tabel IV-2.	Kebutuhan Non-fungsional	IV-4
Tabel IV-3.	Contoh Data Judul Berita Berbahasa Indonesia	IV-6
Tabel IV-4.	Hasil Proses <i>Case Folding</i>	IV-7
Tabel IV-5.	Hasil Proses Penghapusan <i>Stopwords</i>	IV-8
Tabel IV-6.	Hasil Proses <i>Stemming</i>	IV-8
Tabel IV-7.	Hasil Proses Tokenisasi	IV-9
Tabel IV-8.	Tabel Kontingensi Kata ‘wow’	IV-10
Tabel IV-9.	Penjumlahan Nilai Baris dan Kolom Tabel Kontingensi	IV-11
Tabel IV-10.	Perhitungan Frekuensi Harapan Kata ‘Wow’	IV-11
Tabel IV-11.	Seleksi Fitur Berdasarkan Nilai <i>Chi-Square</i> Dengan Rasio 75%.	IV-12
Tabel IV-12.	Hasil Seleksi Fitur	IV-14

Tabel IV-13.	Hasil <i>TF-DF</i>	IV-15
Tabel IV-14.	<i>Conditional Probability</i> Setiap Kata atau <i>Term</i>	IV-17
Tabel IV-15.	Detail Hasil Klasifikasi Data Latih.....	IV-20
Tabel IV-16.	Contoh <i>confusion matrix</i>	IV-20
Tabel IV-17.	Tabel Definisi Aktor.....	IV-22
Tabel IV-18.	Definisi Diagram <i>Use Case</i>	IV-23
Tabel IV-19.	Skenario Memilih Data Masukan.....	IV-24
Tabel IV-20.	Skenario Melatih Data dan Melakukan Seleksi Fitur <i>Chi-Square</i>	IV-26
Tabel IV-21.	Skenario Pengujian	IV-27
Tabel IV-22.	Implementasi Kelas	IV-43
Tabel IV-23.	Skenario Pengujian <i>Use Case</i> Memilih Data Latih.....	IV-50
Tabel IV-24.	Skenario Pengujian <i>Use Case</i> Melatih Data.....	IV-50
Tabel IV-25.	Skenario Pengujian <i>Use Case</i> Menguji Sistem	IV-51
Tabel IV-26.	Pengujian <i>Use Case</i> Memilih Data Masukan.....	IV-52
Tabel IV-27.	Pengujian <i>Use Case</i> Melatih Data.....	IV-53
Tabel IV-28.	Pengujian <i>Use Case</i> Menguji Sistem	54
Tabel V-1.	Hasil <i>Confusion Matrix</i> Dari Pengujian Pertama Klasifikasi Menggunakan <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dan Tanpa Penghapusan <i>Stopwords</i>	V-2
Tabel V-2.	Hasil <i>Confusion Matrix</i> Dari Pengujian Kedua Klasifikasi Menggunakan <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dan Tanpa Penghapusan <i>Stopwords</i>	V-3
Tabel V-3.	Hasil <i>Confusion Matrix</i> Dari Pengujian Ketiga Klasifikasi Menggunakan <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dan Tanpa Penghapusan <i>Stopwords</i>	V-3
Tabel V-4.	Hasil <i>Confusion Matrix</i> Dari Pengujian Pertama Klasifikasi Menggunakan <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dan Dengan Penghapusan <i>Stopwords</i>	V-4

Tabel V-5.	Hasil <i>Confusion Matrix</i> Dari Pengujian Kedua Klasifikasi Menggunakan <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dan Dengan Penghapusan <i>Stopwords</i>	V-4
Tabel V-6.	Hasil <i>Confusion Matrix</i> Dari Pengujian Ketiga Klasifikasi Menggunakan <i>Naïve Bayes</i> dan Seleksi Fitur <i>Chi Square</i> dan Dengan Penghapusan <i>Stopwords</i>	V-5
Tabel V-7.	<i>Accuracy, Precision, Recall, dan F-measure</i> Klasifikasi Menggunakan <i>Naïve Bayes Classifier</i> dan Seleksi Fitur <i>Chi Square</i> Tanpa Penghapusan <i>Stopwords</i>	V-6
Tabel V-8.	<i>Accuracy, Precision, Recall, dan F-measure</i> Klasifikasi Menggunakan <i>Naïve Bayes Classifier</i> dan Seleksi Fitur <i>Chi Square</i> Dengan Penghapusan <i>Stopwords</i>	V-7

DAFTAR GAMBAR

	Halaman
Gambar II-1. Proses <i>Text Mining</i> (Purbo, 2019).....	II-3
Gambar II-2. Dimensi RUP (Liu dan Roussev, 2006).....	II-15
Gambar III-1. Kerangka Kerja Penelitian	III-3
Gambar IV-1. Contoh Perhitungan Nilai <i>Chi-Square</i> dari Kata ‘wow’	IV-12
Gambar IV-2. Contoh Perhitungan Prior Probability	IV-17
Gambar IV-3. Contoh Perhitungan Conditional Probability	IV-17
Gambar IV-4. Hasil Perhitungan Probabilitas Setiap Kategori Dari Data Latih....	IV-19
Gambar IV-5. Perhitungan <i>accuracy, precision, recall, dan f-measure</i>	IV-21
Gambar IV-6. Diagram <i>Use Case</i>	IV-22
Gambar IV-7. Rancangan Antarmuka Halaman Menu Proses <i>Training</i> Dan Memilih Data Latih	IV-30
Gambar IV-8. Rancangan Antarmuka Data Latih Telah Dipilih.....	IV-31
Gambar IV-9. Rancangan Antarmuka Proses Pelatihan Selesai	IV-31
Gambar IV-10. Rancangan Antarmuka Halaman Menu Proses <i>Training</i> Dan Memilih Data Uji	IV-32
Gambar IV-11. Rancangan Antarmuka Data Uji Telah Dipilih.....	IV-32
Gambar IV-12. Rancangan Antarmuka Proses Pengklasifikasian Selesai	IV-33
Gambar IV-13. Diagram Aktivitas Memilih Data Masukan	IV-35
Gambar IV-14. Diagram Aktivitas Melatih Data Latih dan Menghitung Nilai <i>Chi-Square</i>	36
Gambar IV-15. Diagram Aktivitas Menguji Sistem.....	IV-37
Gambar IV-16. Diagram Sequence Melatih Data dan Seleksi Fitur Menggunakan Chi Square.....	IV-39

Gambar IV-17.	Diagram Sequence Proses Klasifikasi Menggunakan Metode Naïve Bayes Classifier dan Seleksi Fitur Chi Square.....	IV-40
Gambar IV-18.	Diagram Kelas Klasifikasi Judul Berita <i>Clickbait</i>	IV-42
Gambar IV-19.	Implementasi Antarmuka Halaman Menu Proses <i>Training</i> Dan Memilih Data Latih	IV-45
Gambar IV-20.	Implementasi Antarmuka Data Latih Telah Dipilih.....	IV-46
Gambar IV-21.	Implementasi Antarmuka Proses Pelatihan Selesai	IV-46
Gambar IV-22.	Implementasi Antarmuka Halaman Menu Proses <i>Training</i> Dan Memilih Data Uji	IV-47
Gambar IV-23.	Implementasi Antarmuka Data Uji Telah Dipilih.....	IV-47
Gambar IV-24.	Implementasi Antarmuka Proses Pengklasifikasian Selesai ..	IV-48
Gambar V-1.	Grafik Perbandingan Data Yang Tidak Bisa Terklasifikasi Antara Pengklasifikasian Tanpa Proses Penghapusan <i>Stopwords</i> Dan Dengan Penghapusan <i>Stopwords</i> Pada Pengujian Pertama	V-9
Gambar V-2.	Grafik Perbandingan Data Yang Tidak Bisa Terklasifikasi Antara Pengklasifikasian Tanpa Proses Penghapusan <i>Stopwords</i> Dan Dengan Penghapusan <i>Stopwords</i> Pada Pengujian Kedua	V-9
Gambar V-3.	Grafik Perbandingan Data Yang Tidak Bisa Terklasifikasi Antara Pengklasifikasian Tanpa Proses Penghapusan <i>Stopwords</i> Dan Dengan Penghapusan <i>Stopwords</i> Pada Pengujian Ketiga.....	V-10
Gambar V-4.	Grafik Perbandingan Nilai <i>Accuracy</i> Dari Setiap Rasio Seleksi Fitur Hasil Klasifikasi Tanpa Penghapusan <i>Stopwords</i> dan Dengan Penghapusan <i>Stopwords</i>	V-11
Gambar V-5.	Grafik Perbandingan Nilai <i>Precision</i> Dari Setiap Rasio Seleksi Fitur Hasil Klasifikasi Tanpa Penghapusan <i>Stopwords</i> dan Dengan Penghapusan <i>Stopwords</i>	V-12
Gambar V-6.	Grafik Perbandingan Nilai <i>Recall</i> Dari Setiap Rasio Seleksi Fitur Hasil Klasifikasi Tanpa Penghapusan <i>Stopwords</i> dan Dengan Penghapusan <i>Stopwords</i>	V-13
Gambar V-7.	Grafik Perbandingan Nilai <i>F-measure</i> Dari Setiap Rasio Seleksi Fitur Hasil Klasifikasi Tanpa Penghapusan <i>Stopwords</i> dan Dengan Penghapusan <i>Stopwords</i>	V-14

DAFTAR LAMPIRAN

1. Dokumentasi *Source Code*

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini memberikan penjelasan umum mengenai keseluruhan penelitian berdasarkan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan serta kesimpulan.

1.2 Latar Belakang

Dengan meledaknya pertumbuhan informasi dari internet, dibutuhkan pengklasifikasian yang tepat guna mengorganisasikan informasi tersebut sehingga dapat dipahami oleh para pencari informasi lainnya (Februariyanti dan Zuliarso, 2012). Namun, pengklasifikasian secara manual oleh manusia mempunyai tantangan dimana membutuhkan waktu dan tenaga yang ekstra (Ting, Ip, dan Tsang, 2011), mengingat informasi yang ada jumlahnya sangat besar dan tidak terorganisir (Februariyanti dan Zuliarso, 2012). Oleh karena itu, dibutuhkan strategi pengelompokan dokumen-dokumen atau informasi-informasi tersebut secara otomatis (Februariyanti dan Zuliarso, 2012).

Ada banyak pendekatan yang dikembangkan dalam penelitian *text mining* khususnya pengklasifikasian dokumen secara otomatis, salah satunya adalah *Naive Bayes Classifier*. Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Yavi, 2018). *Naive Bayes Classifier* adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam

pengklasifikasian teks (Routray *et al.*, 2013). Menurut Narayanan, Arora, dan Bhatia (2013), algoritma *Naive Bayes Classifier* dapat meningkatkan hasil akurasi klasifikasi dan menghilangkan *noise* dengan cara memilih jenis fitur yang tepat dan sesuai.

Seleksi fitur bisa menjadikan alat klasifikasi lebih baik dan lebih efisien, serta efektif dengan cara menyeleksi jumlah fitur data latih, serta menentukan fitur yang cocok untuk dipertimbangkan dalam proses pembelajaran (Utami dan Wahono, 2015). Atribut-atribut kurang relevan yang menyebabkan besarnya jumlah ruang fitur dapat dikurangkan dengan mengeleminasi atribut tersebut menggunakan seleksi fitur (Koncz dan Paralic, 2011). Penyebab klasifikasi menjadi tidak benar juga salah satunya karena adanya fitur *noise*, dan fitur *noise* tersebut menurut Rasywir dan Purwarianti (2015) dapat dihilangkan dengan penerapan seleksi fitur.

Chi-Square adalah salah satu contoh metode seleksi fitur. Metode ini merupakan metode pengujian data hipotesis diskrit secara statistik, yang mengevaluasi korelasi antar dua variabel, dan menentukan apakah dua variabel tersebut memiliki hubungan atau tidak (Snedecor dan Cochran, 1989). Dalam subjek klasifikasi teks, dua variabel dalam *Chi-Square Testing* tersebut adalah kata dan kategori (Thabtah *et al.*, 2009).

Kegiatan penelitian yang dilakukan di bidang klasifikasi dokumen ada banyak, contohnya yaitu pengklasifikasian berita *clickbait*. *Clickbait* mengacu pada jenis iklan konten web tertentu yang dirancang untuk memikat pembaca agar mengklik tautan (Potthast *et al.*, 2016). Tidak sedikit berita yang menyalahgunakan *clickbait* dengan memberikan konten berita yang tidak sesuai dengan judulnya

bahkan yang bersifat palsu atau dibuat-buat (Yavi, 2018). Berdasarkan studi (Chen, Conroy, dan Rubin, 2015) *clickbait* berpotensi menjadi berita bohong. Semakin maraknya penyalahgunaan *clickbait* pada artikel berita *online* menyebabkan banyak pengguna atau pembaca berita *online* merasa tertipu (Maulidi et al., 2018). Berita *online* menyebar dengan sangat cepat, artinya kebenaran yang simpang siur dapat menyebabkan kebingungan dan perpecahan bagi orang banyak (Chen, Conroy, dan Rubin, 2015).

Penelitian dalam bidang klasifikasi dengan studi kasus *clickbait* salah satunya dilakukan oleh (Yavi, 2018) yang membuat aplikasi untuk mengklasifikasikan judul berita berbahasa Indonesia menjadi *clickbait* atau *non-clickbait* dengan menggunakan algoritma *Naive Bayes Classifier*. Penelitian ini memperoleh nilai *recall* sebesar 79%.

Penelitian yang dilakukan Potthast *et al.* (2016) adalah mencoba mendeteksi *clickbait* dari judul berita berbahasa Inggris dengan membandingkan tiga algoritma, yaitu *Logistic Regression*, *Naive Bayes*, dan *Random Forest* serta menerapkan seleksi fitur *Chi-Square* dengan memilih 10, 100, dan 1000 fitur teratas dan menghasilkan metode *Random Forest* unggul dengan nilai ROC-AUC 0.79.

Pada penelitian yang dilakukan oleh Ling, Kencana, dan Oka (2014), dibangun sistem untuk analisis sentimen yaitu mengelompokkan sebuah dokumen menjadi positif atau negatif dengan menggunakan metode *Naive Bayes Classifier* dan seleksi fitur *Chi Square* yang memperoleh akurasi sebesar 83%. Fitur yang digunakan setelah diseleksi adalah sebanyak 25% dari keseluruhan. Ling, Kencana, dan Oka (2014) menyimpulkan kemunculan frekuensi fitur pada kategori yang

diharapkan dan kategori yang tidak diharapkan memiliki peran penting dalam seleksi fitur *Chi-Square*.

Berdasarkan dari hasil uraian tersebut, maka penelitian ini akan mengklasifikasikan judul berita berbahasa Indonesia menjadi *clickbait* atau *non-clickbait* menggunakan metode *Naive Bayes Classifier* dan seleksi fitur *Chi-Square*.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana tingkat kinerja metode *Naive Bayes Classifier* dan seleksi fitur *Chi-Square* dalam mengklasifikasikan dokumen. Untuk menyelesaikan permasalahan tersebut, maka penelitian ini dibagi dalam beberapa *research question* antara lain:

1. Bagaimana mekanisme metode *Naive Bayes Classifier* dan seleksi fitur *Chi Square* dalam mengklasifikasikan judul berita berbahasa Indonesia?
2. Bagaimana kinerja metode *Naive Bayes Classifier* dan metode seleksi fitur *Chi-Square* tanpa penghapusan *stopwords* dalam mengklasifikasikan judul berita berbahasa Indonesia?
3. Bagaimana kinerja metode *Naive Bayes Classifier* dan metode seleksi fitur *Chi-Square* dengan penghapusan *stopwords* dalam mengklasifikasikan judul berita berbahasa Indonesia?
4. Bagaimana pengaruh pembagian rasio dalam penggunaan fitur berdasarkan seleksi fitur *Chi-Square* terhadap proses pengklasifikasian dan kinerja sistem?
5. Bagaimana pengaruh *stopwords removal* terhadap proses pengklasifikasian dan kinerja sistem?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Membangun sistem klasifikasi judul berita berbahasa Indonesia menjadi *clickbait* dan bukan *clickbait* menggunakan metode *Naïve Bayes Classifier* dan seleksi fitur *Chi Square*.
2. Mengetahui kinerja metode *Naive Bayes Classifier* dan metode seleksi fitur *Chi-Square* tanpa penghapusan *stopwords* dalam mengklasifikasikan judul berita berbahasa Indonesia. Tingkat kinerja metode diukur berdasarkan nilai *accuracy, precision, recall* dan *f-measure*.
3. Mengetahui kinerja metode *Naive Bayes Classifier* dan metode seleksi fitur *Chi-Square* dengan penghapusan *stopwords* dalam mengklasifikasikan judul berita berbahasa Indonesia. Tingkat kinerja metode diukur berdasarkan nilai *accuracy, precision, recall* dan *f-measure*.
4. Mengetahui pengaruh pembagian rasio dalam penggunaan fitur berdasarkan seleksi fitur *Chi-Square* terhadap proses pengklasifikasian dan kinerja sistem.
5. Mengetahui pengaruh *stopwords removal* terhadap proses pengklasifikasian dan kinerja sistem.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Sistem dari hasil penelitian ini bisa digunakan untuk mengklasifikasikan judul berita berbahasa Indonesia dengan kategori *clickbait* atau bukan *clickbait*.

2. Dapat digunakan sebagai salah satu rujukan atau referensi penelitian dalam pengklasifikasian dokumen menggunakan *Naive Bayes Classifier* dan seleksi fitur *Chi-Square*.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah :

1. Dokumen yang digunakan merupakan judul-judul berita *online* berbahasa Indonesia.
2. Data latih dan data uji yang akan digunakan berupa dokumen dengan format .xlsx.
3. Kategori pengklasifikasian terdiri dari 2 kategori yaitu *clickbait* dan bukan *clickbait*.
4. Pilihan rasio seleksi fitur berdasarkan nilai *Chi-Square* sebesar 25%, 50%, 75%, dan 100%.

1.7 Sistematika Penulisan

Sistematika penulisan proposal skripsi ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Bab I menguraikan latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan.

BAB II. KAJIAN LITERATUR

Bab II berisi landasan teori yang digunakan dalam penelitian ini. Pembahasan dalam bab ini meliputi penjelasan mengenai *Clickbait*, *text*

preprocessing, *Naive Bayes Classifier*, Seleksi Fitur *Chi Square*, dan *Confusion Matrix*. Selain itu, pada bab ini juga membahas penelitian-penelitian lain yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab III berisi pembahasan mengenai tahapan yang dilaksanakan pada penelitian ini. Rencana tahapan akan dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini juga akan dijabarkan perancangan manajemen proyek untuk pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab IV membahas tahapan yang dilakukan selama proses pengembangan perangkat lunak berdasarkan metode RUP (*Rational Unified Process*).

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab V menguraikan hasil pengujian sesuai dengan rancangan yang telah dibangun sebelumnya dan membahas hasil analisis dari pengujian yang telah dilakukan.

BAB VI. KESIMPULAN DAN SARAN

Bab VI berisi kesimpulan yang didapatkan dari penelitian serta membahas saran yang dapat diberikan oleh penulis untuk penelitian selanjutnya.

1.8 Kesimpulan

Pada penelitian ini, akan membahas bagaimana cara menerapkan metode *Naive Bayes Classifier* dengan pengimplementasian seleksi fitur *Chi-Square*. Selain itu, penelitian ini juga akan menguji pengaruh banyaknya penggunaan berdasarkan rasio seleksi fitur dari nilai *Chi-Square* pada pengklasifikasian dokumen. Selanjutnya, dilakukan pengujian tingkat kinerja hasil klasifikasi dari masing-masing rasio fitur untuk bisa ditentukan rasio mana yang paling terbaik yang dicapai dalam penelitian ini.

DAFTAR PUSTAKA

- Adiwijaya, I. (2006). Text Mining dan Knowledge Discovery. *Kolokium Bersama Komunitas Datamining Indonesia & Soft-Computing Indonesia*, 1–9. [http://web.ipb.ac.id/~ir-lab/pdf/tm \(text summarization\).pdf](http://web.ipb.ac.id/~ir-lab/pdf/tm%20(text%20summarization).pdf)
- Aggarwal, C. C. (2018). *Machine learning for text*. Cham: Springer International Publishing.
- Anand, A., Chakraborty, T., & Park, N. (2017, April). We used neural networks to detect clickbaits: You won't believe what happened next!. In *European Conference on Information Retrieval* (pp. 541-547). Springer, Cham.
- Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma dan Implementasi*. Penerbit Andi.
- Biyani, P., Tsioutsoulis, K., & Blackmer, J. (2016). Detecting Clickbaits in News Streams Using Article Informality. *Thirtieth AAAI Conference on Artificial Intelligence*, 94–100.
- Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop Clickbait: Detecting and preventing clickbaits in online news media. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 9–16. <https://doi.org/10.1109/ASONAM.2016.7752207>
- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). Misleading online content: Recognizing clickbait as “false news.” *WMDD 2015 - Proceedings of the ACM Workshop on Multimodal Deception Detection, Co-Located with ICMI 2015*, 15–19. <https://doi.org/10.1145/2823465.2823467>
- Februariyanti, Herny. Zuliarso, E. (2012). Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi. *Teknologi Informasi DINAMIK*, 17(1), 14–23. <http://www.unisbank.ac.id/ojs/index.php/fti1/article/view/1612/594>
- Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A brief survey of text mining. In *Ldv Forum* (Vol. 20, No. 1, pp. 19-62).
- Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J., Nithya, M., Kannan, S., & Gurusamy, V. (2015). Preprocessing Techniques for Text Mining. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.
- Koncz, P., & Paralic, J. (2011). An approach to feature selection for sentiment analysis. *INES 2011 - 15th International Conference on Intelligent Engineering Systems, Proceedings*, 357–362. <https://doi.org/10.1109/INES.2011.5954773>

- LING, J., N. KENCANA, I. P. E., & OKA, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92. <https://doi.org/10.24843/mtk.2014.v03.i03.p070>
- Jurasfky, D., & Martin, J. H. (2000). An introduction to natural language processing, computational linguistics, and speech recognition.
- Manjesh, S., Kanakagiri, T., Vaishak, P., Chettiar, V., & Shobha, G. (2018). Clickbait Pattern Detection and Classification of News Headlines Using Natural Language Processing. *2nd International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS 2017*, 1–5. <https://doi.org/10.1109/CSITSS.2017.8447715>
- Maulidi, R., Ayilillahi, M. F., Isyiriyah, L., & Palandi, J. F. (2018). Penerapan Neural Network Backpropagation untuk Klasifikasi Artikel Clickbait. *Seminar Nasional FST 2018*, 1(July), 751–757.
- Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.
- Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8206 LNCS, 194–201. https://doi.org/10.1007/978-3-642-41278-3_24
- Palau-Sampio, D. (2016). Reference press metamorphosis in the digital context: Clickbait and tabloid strategies in Elpais.com. *Communication and Society*, 29(2), 63–71. <https://doi.org/10.15581/003.29.2.63-79>
- Potthast, M., Köpsel, S., Stein, B., & Hagen, M. (2016, March). Clickbait detection. In *European Conference on Information Retrieval* (pp. 810-817). Springer, Cham.
- Purbo, O. W. (2019). Text Mining. *Yogyakarta: ANDI*.
- Rasywir, E., & Purwarianti, A. (2015). Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin. *Jurnal Cybermatika*, 3(2), 1–8. <https://www.mendeley.com/import/>
- Routray, P., Swain, C. K., & Mishra, S. P. (2013). A survey on sentiment analysis. *International Journal of Computer Applications*, 76(10).
- Shalahuddin, M. (2014). *Rekayasa Perangkat Lunak*.
- Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods, 8thEdn. *Ames: Iowa State Univ. Press Iowa*, 54, 71-82.

- Srivastava, A. N., & Sahami, M. (Eds.). (2009). *Text mining: Classification, clustering, and applications*. CRC press.
- Suharno, C. F., Fauzi, M. A., & Perdana, R. S. (2017). Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square. *Systemic: Information System and Informatics Journal*, 3(1), 25–32. <https://doi.org/10.29080/systemic.v3i1.191>
- Thabtah, F., Eljinini, M. A. H., Zamzeer, M., & Hadi, M. (2009). *Naïve Bayesian Based on Chi Square to Categorize Arabic Data - Thabtah et al. - 2009.pdf*. 10, 158–163.
- Ting, S. L., Ip, W. H., & Tsang, A. H. C. (2011). Is Naïve bayes a good classifier for document classification? *International Journal of Software Engineering and Its Applications*, 5(3), 37–46.
- Utami, Lila Dini. Wahono, R. S. (2015). Integrasi Metode Information Gain untuk Seleksi Fitur dan AdaBoost untuk Mengurangi Bias pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naive Bayes. *Journal of Intelligent Systems*, 1(2), 120–126.
- Yavi, A. F. (2018). Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi Clickbait menggunakan Metode Naïve Bayes. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Yulianto, B (2010). Teknologi *Location Based Service (Global Positioning System)* Pada Perangkat *Mobile*. Comtech (Vol. 1 No. 1 61-74).