

**SENTIMEN ANALISIS KOMENTAR INSTAGRAM PADA
AKUN MATA NAJWA MENGGUNAKAN METODE SELEKSI
FITUR *INFORMATION GAIN* DAN *CHI-SQUARE* DENGAN
*RANDOM FOREST***

Diajukan Sebagai Syarat Untuk Menyelesaikan

Pendidikan Program Strata-1

Pada Jurusan Teknik Informatika



Oleh :

M Zahran Algafaru

NIM : 09021382126123

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2025

HALAMAN PENGESAHAN

SKRIPSI

Sentimen Analisis Komentar Instagram Pada Akun Mata Najwa Menggunakan Metode Seleksi Fitur Information Gain dan Chi-Square dengan Random Forest

Sebagai salah satu syarat untuk penyelesaian studi di

Program Studi S1 Teknik Informatika

Oleh:

M ZAHRAN ALGAFARU

09021382126123

Pembimbing 1 : Alvi Syahrini Utami, S.Si., M.Kom
NIP. 197812222006042003

Mengetahui

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D
198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 11 Juli 2025 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : M ZAHARAN ALGAFARU

NIM : 09021382126123

Judul : Sentimen Analisis Komentar Instagram Pada Akun Mata Najwa Menggunakan Metode Seleksi Fitur *Information Gain* dan *Chi-Square* dengan *Random Forest*.

dan dinyatakan LULUS.

1. Ketua Penguji

Rizki Kurniati, S.Kom., M.T

NIP.199107122019032016

2. Penguji I

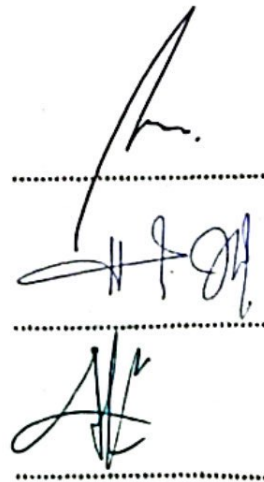
Dr. Annisa Darmawahyuni, S.Kom., M.Kom

NIP.199006302023212044

3. Pembimbing I

Alvi Syahrini Utami, S.Si., M.Kom

NIP. 197812222006042003



Mengetahui
Ketua Jurusan Teknik Informatika



HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : M Zahran Algafaru

NIM : 09021382126123

Program Studi : Teknik Informatika

Judul : Sentimen Analisis Komentar Instagram Pada Akun Mata Najwa
Menggunakan Metode Seleksi Fitur *Information Gain* dan *Chi-Square* dengan
Random Forest.

Hasil Pengecekan Software iThenticate/Turnitin : 6%

Menyatakan bahwa Laporan Penelitian saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan penelitian 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, 11 Juli 2025



M Zahran Algafaru

NIM. 09021382126123

MOTTO DAN PERSEMBAHAN

“Ilmu adalah cahaya yang menerangi jalan dan pengetahuan adalah kunci untuk membuka pintu masa depan”

Kupersembahkan karya tulis ini kepada :

- Orang Tua, Saudara/i, dan Keluargaku
- Dosen Pembimbing
- Teman-teman seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Instagram is one of the most popular social media platforms that is not only used to share pictures and videos, but also as a means of discussion on various social, political and cultural issues. One of the most active and influential accounts in Indonesia is the Mata Najwa account, which often triggers public discussion through the comments section. The large number of incoming comments creates challenges in managing and analyzing the data, especially due to the presence of comments that are ambiguous, irrelevant, or even contain hate speech. This research aims to develop software for classifying Instagram comments on the Mata Najwa account using the Random Forest algorithm. The dataset used consists of 574 comments categorized into three classes: positive, negative, and neutral. Three testing approaches were conducted: Random Forest without feature selection, Random Forest with feature selection using Information Gain, and Random Forest with feature selection using Chi-Square. The test results show that the Random Forest method without feature selection produces an average accuracy of 0.7536, precision of 0.4606, recall of 0.4163, and F-Measure of 0.4164. Meanwhile, the Random Forest method with Information Gain produced the highest F-Measure of 0.4567, and the method with Chi-Square recorded the highest accuracy of 0.7565. This research shows that the Random Forest algorithm can be effectively used in the classification of social media comments.

Keywords: *Instagram, Random Forest, Information Gain, Chi-Square*

ABSTRAK

Instagram merupakan salah satu platform media sosial paling populer yang tidak hanya digunakan untuk berbagi gambar dan video, tetapi juga sebagai sarana diskusi mengenai berbagai isu sosial, politik, dan budaya. Salah satu akun yang aktif dan berpengaruh di Indonesia adalah akun *Mata Najwa*, yang kerap memicu diskusi publik melalui kolom komentar. Banyaknya komentar yang masuk menciptakan tantangan dalam mengelola dan menganalisis data tersebut, khususnya karena adanya komentar yang ambigu, tidak relevan, atau bahkan mengandung ujaran kebencian. Penelitian ini bertujuan untuk mengembangkan perangkat lunak sentimen analisis komentar Instagram pada akun Mata Najwa menggunakan algoritma *Random Forest*. Dataset yang digunakan terdiri dari 574 komentar yang dikategorikan ke dalam tiga kelas: positif, negatif, dan netral. Tiga pendekatan pengujian dilakukan: *Random Forest* tanpa seleksi fitur, *Random Forest* dengan seleksi fitur menggunakan *Information Gain*, dan *Random Forest* dengan seleksi fitur menggunakan *Chi-Square*. Hasil pengujian menunjukkan bahwa metode *Random Forest* tanpa seleksi fitur menghasilkan akurasi rata-rata sebesar 0,7536, presisi 0,4606, recall 0,4163, dan F-Measure 0,4164. Sementara itu, metode *Random Forest* dengan *Information Gain* menghasilkan F-Measure tertinggi yaitu 0,4567, dan metode dengan *Chi-Square* mencatat akurasi tertinggi sebesar 0,7565. Penelitian ini menunjukkan bahwa algoritma *Random Forest* dapat digunakan secara efektif dalam klasifikasi komentar media sosial, serta bahwa pemilihan metode seleksi fitur memiliki pengaruh terhadap performa model.

Kata Kunci : Instagram, Random Forest, Information Gain, Chi-Square

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas segala rahmat, hidayah, dan karunia-Nya yang tak terhingga, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Sentimen Analisis Komentar Instagram Pada Akun Mata Najwa Menggunakan Metode Seleksi Fitur *Information Gain* dan *Chi-Square* dengan *Random Forest*”. Tugas akhir ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan pada program studi strata-1 di Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Penulis ingin menyampaikan rasa terima kasih yang mendalam kepada semua pihak yang telah memberikan dukungan dan bantuan, baik secara materil maupun moril, selama proses penyusunan tugas akhir ini. Secara khusus penulis ingin mengapresiasi pihak-pihak yang telah berkontribusi sebagai berikut :

1. Bapak Alm. Najamudin dan Ibu Sutinawati, selaku orang tua yang sangat saya cintai, serta seluruh anggota keluarga besar yang senantiasa memberikan dukungan.
2. Bapak Prof. DR. Erwin, S.Si., M.Si., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Hadipurnawan Satria. M.Sc., Ph.D, selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
4. Ibu Alvi Syahrini Utami, S.Si., M.Kom., selaku Dosen Pembimbing yang telah memberikan arahan, bimbingan, serta masukan yang berharga kepada penulis, sehingga penulis dapat menyelesaikan tugas akhir ini.
5. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika yang telah berbagi ilmu dan pengetahuan kepada penulis selama perkuliahan.
6. Staf Jurusan Teknik Informatika beserta Staf Fakultas Ilmu Komputer yang telah memberikan bantuan dalam hal administrasi dan akademik selama penulis menempuh pendidikan.

7. Teman-teman Informatika Bilingual B angkatan 2021 terutama, Rafi Wisesa, Achmad Rausyan Fiker, Rifqi Edrial, Ksatria, Farhan Nugraha, dan Julian Gabriel yang selalu berbagi informasi dan senantiasa menemani penulis selama masa perkuliahan yang pastinya akan selalu diingat.
8. Aliyah Dynda Shafira, Terima Kasih atas kasih sayang, dukungan, dan semangat yang tak pernah putus selama proses penyelesaian skripsi ini. Kehadiranmu selalu menjadi motivasi yang sangat berarti bagi saya.
9. Semua pihak lainnya yang tidak dapat saya sebutkan satu persatu yang telah berkontribusi dalam penyelesaian tugas akhir ini, penulis mengucapkan terima kasih yang sebesar-besarnya atas segala bantuan dan dukungannya.

Semoga tugas akhir ini dapat memberikan manfaat, baik bagi penulis sendiri maupun bagi para pembaca secara umum, serta dapat dijadikan referensi dan acuan untuk hal-hal yang berguna. Penulis menyadari bahwa dalam proses penyusunan tugas akhir ini terdapat beberapa penjelasan yang mungkin belum sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang konstruktif agar karya ini dapat menjadi lebih baik dan memberikan bekal pengetahuan yang bermanfaat bagi peneliti di masa mendatang.

Palembang, 11 Juli 2025

M Zahran Algafaru

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv

BAB I PENDAHULUAN

1.1	Latar Belakang	I-1
1.2	Rumusan Masalah	I-3
1.3	Tujuan Penelitian.....	I-4
1.4	Manfaat Penelitian.....	I-4
1.5	Batasan Masalah.....	I-4
1.6	Sistematika Penulisan.....	I-5
1.7	Kesimpulan.....	I-6

BAB II TINJAUAN PUSTAKA

2.1	Landasan Teori	II-1
2.1.1	Instagram.....	II-1
2.1.2	<i>Text Pre Processing</i>	II-2
2.1.3	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	II-3
2.1.4	<i>Random Forest</i>	II-5
2.1.5	<i>Chi-Square</i>	II-6
2.1.6	<i>Information Gain</i>	II-7
2.1.7	<i>Confusion Matrix</i>	II-8
2.1.8	<i>Rational Unified Process</i>	II-11

2.2	Penelitian Lain yang Relevan.....	II-13
2.3	Kesimpulan.....	II-15

BAB III METODOLOGI PENELITIAN

3.1	Pengumpulan Data	III-1
3.1.1	Jenis dan Sumber Data	III-1
3.2	Tahap Penelitian	III-3
3.2.1	Mengumpulkan Data	III-4
3.2.2	Menentukan Kerangka Kerja Penelitian	III-5
3.2.3	Pembagian Data Uji dan Data Latih.....	III-5
3.2.4	Menentukan Alat Bantu Penelitian	III-6
3.2.5	Menentukan Kriteria Penelitian	III-6
3.2.6	Menentukan Format Data Penelitian.....	III-7
3.2.7	Melakukan Uji Penelitian.....	III-9
3.2.8	Melakukan Analisis Hasil Pengujian dan Kesimpulan	III-10
3.3	Metode Pengembangan Perangkat Lunak	III-10
3.3.1	<i>Inception</i>	III-11
3.3.2	<i>Elaboration</i>	III-11
3.3.3	<i>Construction</i>	III-11
3.3.4	<i>Transition</i>	III-11
3.4	Manajemen Proyek Penelitian.....	III-12
3.5	Kesimpulan.....	III-12

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Fase Insepsi	IV-1
4.1.1	Pemodelan Bisnis	IV-1
4.1.2	Kebutuhan Sistem	IV-3
4.1.3	Analisis dan Sistem	IV-4
4.1.4	Implementasi	IV-16
4.2	Fase Elaborasi.....	IV-23
4.2.1	Pemodelan Bisnis	IV-23
4.2.1.1	Perancangan Data	IV-23
4.2.1.2	Perancangan Antarmuka Pengguna.....	IV-24

4.2.2	Kebutuhan Sistem	IV-25
4.2.3	<i>Activity Diagram</i>	IV-26
4.2.4	<i>Sequence Diagram</i>	IV-30
4.3	Fase Kontruksi.....	IV-33
4.3.1	<i>Class Diagram</i>	IV-34
4.3.2	Implementasi	IV-34
4.4	Fase Transisi	IV-37
4.4.1	Pemodelan Bisnis	IV-37
4.4.2	Rencana Pengujian	IV-37
4.4.3	Implementasi	IV-38
4.5	Kesimpulan.....	IV-40
BAB V HASIL DAN ANALISIS		
5.1	Data Hasil Penelitian	V-1
5.1.1	Konfigurasi Percobaan	V-1
5.1.2	Hasil Pengujian	V-2
5.2	Analisis Hasil Pengujian Secara Keseluruhan.....	V-7
5.3	Kesimpulan.....	V-14
BAB VI KESIMPULAN DAN SARAN		
6.1	Kesimpulan.....	VI-1
6.2	Saran	VI-2
DAFTAR PUSTAKA		xv
DAFTAR LAMPIRAN		xvii

DAFTAR GAMBAR

	Halaman
Gambar II- 1 Arsitektur RUP.....	II-12
Gambar III- 1 Tahap Penelitian	III-4
Gambar III- 2 Kerangka Kerja Penelitian.....	III-5
Gambar IV- 1 Diagram Use Case.....	IV-17
Gambar IV- 2 Rancangan Unggah Dokumen	IV-24
Gambar IV- 3 Rancangan Halaman Unggah Dokumen dan Random Forest.....	IV-24
Gambar IV- 4 Rancangan Halaman Unggah Dokumen Random Forest dan Information Gain.....	IV-25
Gambar IV- 5 Rancangan Halaman Unggah Dokumen Random Forest dan Chi-Square.....	IV-25
Gambar IV- 6 Activity Diagram Unggah Dokumen.....	IV-26
Gambar IV- 7 Activity Diagram Tabel Hasil Klasifikasi Random Forest.....	IV-27
Gambar IV- 8 Activity Diagram Tabel Hasil Klasifikasi Random Forest dan Information Gain.....	IV-28
Gambar IV- 9 Activity Diagram Tabel Hasil Klasifikasi Random Forest dan Chi-Square.....	IV-29
Gambar IV- 10 Sequence Diagram Unggah Dokumen.....	IV-30
Gambar IV- 11 Sequence Diagram Hasil Random Forest	IV-31
Gambar IV- 12 Sequence Diagram Hasil Random Forest dan Information Gain	IV-32
Gambar IV- 13 Sequence Diagram Hasil Random Forest dan Chi-Square ...	IV-33
Gambar IV- 14 Class Diagram.....	IV-34
Gambar IV- 15 Antarmuka Unggah Dokumen	IV-35
Gambar IV- 16 Antarmuka Hasil Random Forest\	IV-36
Gambar IV- 17 Antarmuka Hasil Random Forest dan Information Gain.....	IV-36
Gambar IV- 18 Antarmuka Hasil Random Forest dan Chi-Square.....	IV-36
Gambar V- 1 Perbandingan Accuracy	V-8
Gambar V- 2 Perbandingan Rata-Rata Accuracy	V-9
Gambar V- 3 Perbandingan Precision	V-10
Gambar V- 4 Perbandingan Rata-Rata Precison	V-10
Gambar V- 5 Perbandingan Recall.....	V-11
Gambar V- 6 Perbandingan Rata-Rata Recall.....	V-11
Gambar V- 7 Perbandingan F-Measure.....	V-12
Gambar V- 8 Perbandingan Rata-Rata F-Measure.....	V-13

DAFTAR TABEL

	Halaman
Tabel II- 1 Contoh Case Folding	II-2
Tabel II- 2 Contoh Tokenizing	II-3
Tabel II- 3 Contoh Stopword Removal.....	II-3
Tabel II- 4 Contoh Stemming	II-3
Tabel II- 5 Tabel Relevansi Antara Kata dan Kelas Kata.....	II-6
Tabel II- 6 Tabel Confusion Matrix Tiga Kelas	II-9
Tabel III- 1 Contoh Data Komentar.....	III-1
Tabel III- 2 Tabel Dataset.....	III-3
Tabel III- 3 Rancangan Tabel Confusion Matrix Hasil Klasifikasi.....	III-7
Tabel III- 4 Rancangan Tabel Hasil Pengujian.....	III-9
Tabel III- 5 Rancangan Tabel Hasil Analisis Klasifikasi	III-10
Tabel IV- 1 Kebutuhan Fungsional.....	IV-4
Tabel IV- 2 Kebutuhan Non-Fungsional.....	IV-4
Tabel IV- 3 Contoh Data.....	IV-5
Tabel IV- 4 Hasil Casefolding	IV-5
Tabel IV- 5 Hasil Cleaning	IV-6
Tabel IV- 6 Hasil Normalisasi	IV-6
Tabel IV- 7 Hasil Stopword Removal.....	IV-7
Tabel IV- 8 Hasil Stemming	IV-7
Tabel IV- 9 Hasil Tokenizing	IV-8
Tabel IV- 10 Transformasi TF-IDF	IV-9
Tabel IV- 11 Hasil Bootstrap Sampling.....	IV-10
Tabel IV- 12 Membangun Pohon dari Bootstrap Sampling.....	IV-11
Tabel IV- 13 Membuat Pohon Keputusan.....	IV-11
Tabel IV- 14 Penggabungan Dari Tree	IV-12
Tabel IV- 15 Nilai Information Gain	IV-12
Tabel IV- 16 TF-IDF Setelah Seleksi Fitur.....	IV-12
Tabel IV- 17 Bootstrap Sampling Random Forest + Information Gain.....	IV-13
Tabel IV- 18 Root Node dan Pembuatan Pohon	IV-13
Tabel IV- 19 Voting Mayoritas	IV-14
Tabel IV- 20 Seleksi Fitur Berdasarkan Chi-Square.....	IV-14
Tabel IV- 21 TF-IDF Setelah Chi-Square.....	IV-14
Tabel IV- 22 Bootstrap Sampling	IV-15
Tabel IV- 23 Pembuatan Phon Keputusan	IV-15
Tabel IV- 24 Voting Mayoritas	IV-16
Tabel IV- 25 Definisi Aktor	IV-17
Tabel IV- 26 Definisi Use Case	IV-18
Tabel IV- 27 Skenario Use Case	IV-18

Tabel IV- 28 Perhitungan dan Menampilkan Analisis Metode Random Forest Tanpa Metode Seleksi Fitur Information Gain dan Chi-Square	IV-19
Tabel IV- 29 Perhitungan dan Menampilkan Hasil Analisis Metode Random Forest Menggunakan Metode Seleksi Fitur Information Gain	IV-20
Tabel IV- 30 Perhitungan dan Menampilkan Analisis Metode Random Forest Menggunakan Metode Seleksi Fitur Chi-Square	IV-22
Tabel IV- 31 Deskripsi Kelas Dokumen	IV-34
Tabel IV- 32 Daftar Dokumen Program Interface	IV-35
Tabel IV- 33 Perencanaan Use Case Unggah Dokumen.....	IV-37
Tabel IV- 34 Perencanaan Use Case Hasil Klasifikasi Random Forest Tanpa Metode Seleksi Fitur Information Gain dan Chi-Square	IV-37
Tabel IV- 35 Perencanaan Use Case Hasil Klasifikasi Random Forest dengan Information Gain.....	IV-38
Tabel IV- 36 Perencanaan Use Case Hasil Klasifikasi Random Forest dengan Chi-Square	IV-38
Tabel IV- 37 Evaluasi Use Case dari Unggah Dokumen.....	IV-38
Tabel IV- 38 Evaluasi Use Case dari Hasil Klasifikasi Random Forest.....	IV-38
Tabel IV- 39 Evaluasi Use Case dari Hasil Klasifikasi Random Forest dengan Information Gain.....	IV-39
Tabel IV- 40 Evaluasi Use Case dari Hasil Klasifikasi Random Forest dengan Chi-Square	IV-39
Tabel V- 1 Confusion Matrix Random Forest Tanpa Metode Seleksi Fitur Information Gain dan Chi-Square	V-3
Tabel V- 2 Hasil Evaluasi Klasifikasi Random Forest Tanpa Metode Seleksi Fitur Information Gain dan Chi-Square	V-4
Tabel V- 3 Confusion Matrix Random Forest dengan Metode Seleksi Fitur Information Gain.....	V-4
Tabel V- 4 Hasil Evaluasi Random Forest dengan Metode Seleksi Fitur Information Gain	V-5
Tabel V- 5 Confusion Matrix Random Forest dengan Metode Seleksi Fitur Chi-Square.....	V-6
Tabel V- 6 Hasil Evaluasi Random Forest dengan Metode Seleksi Fitur Chi-Square.....	V-7

BAB I

PENDAHULUAN

Bab ini mendeskripsikan latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, dan kesimpulan penelitian yang merupakan gagasan utama yang menjadi dasar penelitian.

1.1 Latar Belakang

Instagram merupakan salah satu platform media sosial yang paling populer di dunia, Instagram juga termasuk platform untuk berbagi gambar dan video (Nurhasna Ayutiani & Primadani Satria Putri, 2018). Platform ini tidak hanya digunakan untuk berbagi gambar dan video, tetapi juga sebagai sarana untuk berinteraksi dan berdiskusi mengenai berbagai isu, termasuk isu-isu sosial, politik, dan budaya. Salah satu akun yang cukup berpengaruh di Indonesia adalah akun Mata Najwa, yang sering kali memposting konten yang memicu diskusi dan komentar dari pengikutnya. Dengan banyaknya komentar yang diterima penting untuk menganalisis dan mengklasifikasikan komentar tersebut agar dapat memahami sentimen publik terhadap isu yang diangkat.

Namun dengan volume komentar yang sangat besar terdapat tantangan dalam mengelola dan menganalisis data tersebut. Banyak komentar yang bersifat ambigu, tidak relevan, atau bahkan mengandung ujaran kebencian. Hal ini membuat proses klasifikasi komentar menjadi kompleks dan memerlukan pendekatan yang tepat untuk mendapatkan hasil yang akurat. Penelitian ini bertujuan untuk mengatasi

masalah tersebut dengan menggunakan metode klasifikasi yang efektif untuk mengelompokkan komentar berdasarkan sentimen dan relevansinya. Metode seleksi fitur seperti *Chi-Square* dan *Information Gain* sangat penting dalam penelitian ini karena dapat membantu dalam memilih fitur-fitur yang paling relevan untuk klasifikasi. Dengan mengurangi dimensi data, metode ini tidak hanya meningkatkan akurasi model, tetapi juga mengurangi waktu pemrosesan (Ruhyana & Rosiyadi, 2020). Hal ini sangat krusial dalam konteks analisis data besar dimana kecepatan dan efisiensi menjadi faktor utama dalam pengambilan keputusan.

Metode seleksi fitur *Chi-Square* dan *Information Gain* berfungsi untuk menilai relevansi fitur dalam dataset. *Chi-Square* mengukur ketergantungan antara dua variabel, sedangkan *Information gain* mengukur seberapa banyak informasi yang diperoleh tentang kelas dari fitur tertentu. Dengan menggunakan kedua metode ini, peneliti dapat mengidentifikasi fitur-fitur yang paling berkontribusi terhadap klasifikasi komentar sehingga model yang dibangun menjadi lebih efisien dan efektif.

Beberapa penelitian sebelumnya telah menunjukkan keberhasilan penggunaan *Random Forest* dalam klasifikasi teks. Penelitian oleh Masbadi Hatullah Nurnaryo dkk, (2022) menggunakan *Random Forest* untuk mendeteksi *cyberbullying* pada tweet berbahasa Indonesia dan mencatat akurasi sebesar 72,1%. Di sisi lain Salim & Sutabri, (2023) melaporkan tingkat akurasi 86% dalam klasifikasi komentar *bullying* di YouTube dengan algoritma *Random Forest*. Hal ini menunjukkan bahwa *Random Forest* adalah pilihan yang tepat untuk digunakan

dalam penelitian ini mengingat kemampuannya dalam menangani data yang besar dan kompleks.

Random Forest adalah algoritma *machine learning* berbasis *ensemble* yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi klasifikasi. Dengan menggunakan teknik *Bootstrap aggregating*, *Random Forest* dapat mengurangi *overfitting* dan meningkatkan generalisasi model. Keunggulan ini menjadikannya salah satu metode yang paling banyak digunakan dalam berbagai aplikasi klasifikasi termasuk analisis sentimen di media sosial (Salim & Sutabri, 2023).

Dalam penelitian ini, penggunaan *Chi-Square* dan *Information Gain* sebagai metode seleksi fitur akan membantu dalam meningkatkan performa *Random Forest*. Dengan memilih fitur yang paling relevan, model klasifikasi dapat lebih fokus pada informasi yang penting sehingga menghasilkan prediksi yang lebih akurat dan dapat diandalkan. Hasil dari penelitian ini diharapkan tidak hanya mampu meningkatkan performa model klasifikasi, tetapi juga membantu mempercepat proses pengolahan data secara lebih efisien, sehingga dapat di aplikasikan pada berbagai kebutuhan analisis data yang kompleks.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini sebagai berikut :

1. Bagaimana perbandingan kinerja antara metode seleksi fitur *Chi-Square* dan *Information Gain* dalam meningkatkan akurasi klasifikasi komentar pada media sosial ?
2. Bagaimana pengaruh seleksi fitur terhadap akurasi klasifikasi komentar yang dilakukan oleh model *Random Forest* ?

1.3 Tujuan Penelitian

1. Untuk mengetahui perbandingan kinerja antara metode seleksi fitur *Chi-Square* dan *Information Gain* dalam meningkatkan akurasi klasifikasi komentar pada media sosial.
2. Mengetahui pengaruh seleksi fitur terhadap akurasi klasifikasi komentar yang dilakukan oleh model *Random Forest*.

1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Penelitian ini dapat menjadi acuan yang bermanfaat bagi penelitian selanjutnya.
2. Sebagai masukan bagi pengelola akun Instagram, khususnya Mata Najwa untuk mengurangi munculnya komentar negatif di kolom komentar.

1.5 Batasan Masalah

Berikut ini adalah keterbatasan penelitian ini :

1. Data komentar yang digunakan ada 574 komentar.
2. Data komentar diambil dari akun Instagram Mata Najwa dan diperoleh melalui platform Kaggle.

3. Klasifikasi dibagi menjadi tiga kategori : Positif, Negatif, dan Netral.

1.6 Sistematika Penulisan

Prosedur penulisan yang digunakan dalam penelitian ini disusun dengan ketentuan penulisan tugas akhir yang berlaku di Fakultas Ilmu Komputer Universitas Sriwijaya, dengan langkah-langkah yang disusun secara sistematis sebagai berikut :

BAB I. PENDAHULUAN

Bab ini mencakup pembahasan tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta metode penulisan yang digunakan dalam penyusunan laporan akhir ini.

BAB II. KAJIAN LITERATUR

Bab ini menjelaskan landasan teori yang digunakan untuk mendukung penelitian. Bab ini berisi literatur dan penelitian terdahulu yang terkait dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini menguraikan langkah-langkah atau prosedur yang digunakan dalam penelitian ini, termasuk teknik pengumpulan data dan teknik perancangan perangkat lunak. Setiap fase penelitian akan diuraikan secara menyeluruh sesuai dengan kerangka kerja yang telah ditentukan sebelumnya.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini akan membahas secara menyeluruh tentang proses pengembangan perangkat lunak, dimulai dari tahap pengumpulan data yang menjadi dasar dalam merancang sistem, hingga mencapai hasil akhir berupa program yang dihasilkan.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan temuan penelitian beserta analisis mendalamnya. Dalam bab ini, data yang telah dikumpulkan dan dianalisis akan dipresentasikan secara sistematis untuk menggambarkan pencapaian dari setiap tujuan penelitian yang telah ditetapkan.

BAB VI. KESIMPULAN DAN SARAN

Bab ini memberikan kesimpulan hasil penelitian dan saran terkait penelitian.

1.7 Kesimpulan

Dalam bab ini telah membahas latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang akan digunakan dalam proses penelitian.

DAFTAR PUSTAKA

- Amalia, A., Zaidiah, A., Isnainiyah, I. N., Program,), S1, S., Informasi, S., Komputer, I., Veteran, U., Jl, J. R., Fatmawati, P., & Labu, J. (n.d.). *PREDIKSI KUALITAS UDARA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR*. <https://data.jakarta.go.id/>.
- Amrullah, A. Z., Sofyan Anas, A., Adrian, M., & Hidayat, J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, 2(1). <https://doi.org/10.30812/bite.v2i1.804>
- Antasari, C., & Pratiwi, R. D. (2022). *PEMANFAATAN FITUR INSTAGRAM SEBAGAI SARANA KOMUNIKASI PEMASARAN KEDAI BABAKKEROYOKAN DI KOTA PALU* (Vol. 9, Issue 2).
- Atmoko, B. Dwi. (2012). *Instagram Handbook*.
- Candra, R. M., & Nanda Rozana, A. (2020a). Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor. *IT Journal Research and Development*, 5(1), 45–52. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).4962](https://doi.org/10.25299/itjrd.2020.vol5(1).4962)
- Candra, R. M., & Nanda Rozana, A. (2020b). Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor. *IT Journal Research and Development*, 5(1), 45–52. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).4962](https://doi.org/10.25299/itjrd.2020.vol5(1).4962)
- Indraloka dan Santosa. (2017). *Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia*.
- Jha, M. K. . (2012). *Applied mathematics in electrical & computer engineering : proceedings of the American conference on applied mathematics (AMERICAN-MATH '12) : proceedings of the 6th WSEAS International conference on circuits, systems, signal and telecommunications (CSST '12) : proceedings of the 6th WSEAS International conference on computer engineering and applications (CEA '12) : Harvard, Cambridge, USA, January 25-27, 2012*. WSEAS.
- Masbadi Hatullah Nurnaryo, R., Oktavia Suzanti, I., Abdul Fatah, D., Dwi Cahyani, A., & Ayu Mufarroha, F. (2022). *DETEKSI CYBERBULLYING PADA DATA TWEET MENGGUNAKAN METODE RANDOM FOREST DAN SELEKSI FITUR INFORMATION GAIN CYBERBULLYING DETECTION ON TWEET DATA USING RANDOM FOREST METHOD AND*

INFORMATION GAIN FEATURE SELECTION. 11(1).
<https://www.femina.co.id/trending->

Maulida, I., Suyatno, A., Rahmania Hatta, H., & Mulawarman, U. (n.d.). Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. *OKTOBER 2016 IJCCS*, 17, 1–5.

Nurhasna Ayutiani, D., & Primadani Satria Putri, B. (2018). *Penggunaan Akun Instagram sebagai Media Informasi Wisata Kuliner*. 3(1), 39–59.

Salim, A. N., & Sutabri, T. (n.d.). *Klasifikasi Cyberbullying Pada Komentar Video Youtube Menggunakan Metode Random Forest* (Vol. 11).

Santoso, H., Putri, R. A., & Sahbandi, S. (2023). Deteksi Komentar Cyberbullying pada Media Sosial Instagram Menggunakan Algoritma Random Forest. *Jurnal Manajemen Informatika (JAMIKA)*, 13(1), 62–72.
<https://doi.org/10.34010/jamika.v13i1.9303>

Wifalin, M., Komunikasi, I., Kristen, U., & Surabaya, P. (n.d.). *Efektivitas Instagram Common Grounds*.