

**Klasifikasi Komentar Spam pada Instagram Menggunakan
Algoritma *Mutual Information* dan *Modified K-Nearest Neighbor***

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



Oleh :

Desi Indriani
NIM : 09021181621023

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN SKRIPSI

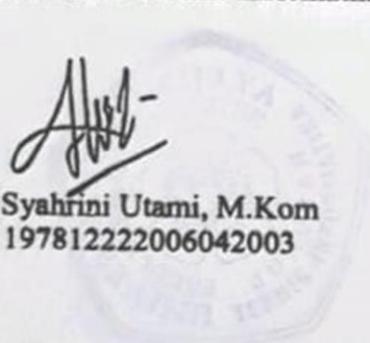
KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM MENGGUNAKAN ALGORITMA *MUTUAL INFORMATION* DAN *MODIFIED K-NEAREST NEIGHBOR*

Oleh :

Desi Indriani
NIM : 09021181621023

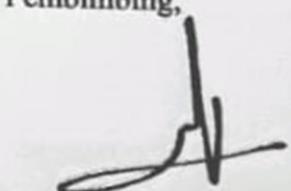
Palembang, 19 Agustus 2021

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

Pembimbing,



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

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Sabtu tanggal 31 Juli 2021 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Desi Indriani
NIM : 09021181621023
Judul : Klasifikasi Komentar Spam pada Instagram Menggunakan Algoritma *Mutual Information* dan *Modified K-Nearest Neighbor*

1. Pembimbing I

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



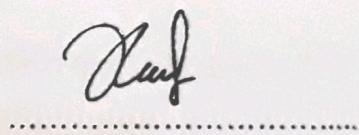
2. Pengaji I

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003



3. Pengaji II

Kanda Januar M, M.T
NIP. 199001092019031012



Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Desi Indriani

NIM : 09021181621023

Program Studi : Teknik Informatika

Judul Skripsi : Klasifikasi Komentar Spam pada Instagram

Menggunakan Algoritma *Mutual Information* dan *Modified K-Nearest Neighbor*

Hasil Pengecekan Software *iThenticate/Turnitin* : 4%

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, 24 Agustus 2021



Desi Indriani

NIM. 09021181621023

- “*Siapa yang bersungguh-sungguh dia akan berhasil.*”

Kupersembahkan Karya Tulis ini kepada :

- **Orang Tuaku**
- **Kakak-kakakku**
- **Dosen Pembimbing**
- **Teman Seperjuanganku IF REG A 2016**
- **Fakultas Ilmu Komputer, Universitas Sriwijaya**

ABSTRACT

Spam on Instagram is defined as comments that are unrelated to the photo or video that is being commented on. Spam comments disrupt the flow of the conversation, making it difficult for users to find information quickly and precisely. Some ways that can be done to solve the problem of spam comments are to classify comments based on the category of spam comments or not spam. One method that can be used for classification is Modified K-Nearest Neighbor (MKNN). However, the MKNN method has a weakness in processing high-dimensional data, so a feature selection method is required to reduce the document's number of features. The method used is Mutual Information (MI). The purpose of this study was to determine the performance of the MKNN classification with MI and MKNN without using MI. The results of the test will be compared and evaluated to see the effect of the MI feature selection method in improving MKNN performance. The test is carried out with the input k values of 3, 5, 7 and 9 and the threshold values of 0.018, 0.01 and 0.008. The test results that have been obtained show that the MKNN method with MI has better performance than the MKNN without MI. MI feature selection is able to reduce the number of features in the data set thereby improving classification performance. The difference in the increase in highest accuracy occurs when the value of k = 3 with a threshold value of 0.018, which is 22%.

Key words: Instagram, Comments, Modified K-Nearest Neighbor, Mutual Information, Spam.

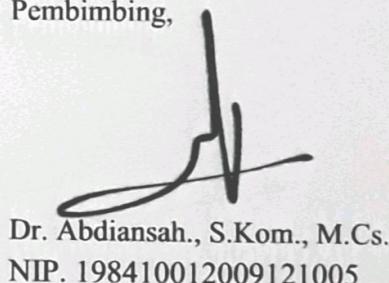
Palembang, 19 Agustus 2021

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

Pembimbing,



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

ABSTRACT

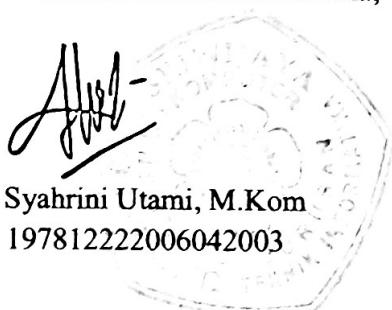
Spam yang terdapat pada Instagram berupa komentar yang tidak berhubungan dengan foto atau video yang dikomentari. Komentar spam sangat mengganggu alur diskusi sehingga membuat pengguna kesulitan dalam mencari informasi secara cepat dan tepat. Beberapa cara yang dapat dilakukan untuk mengatasi masalah komentar spam adalah dengan mengklasifikasi komentar berdasarkan kategori komentar spam atau bukan spam. Salah satu metode yang dapat digunakan adalah Modified K-Nearest Neighbor (MKNN). Namun metode MKNN memiliki kelemahan dalam memproses data yang berdimensi tinggi, sehingga dibutuhkan metode seleksi fitur untuk mengurangi jumlah fitur pada dokumen. Metode yang digunakan adalah Mutual Information (MI). Tujuan dari penelitian ini adalah untuk mengetahui kinerja yang dihasilkan klasifikasi MKNN dengan MI dan MKNN tanpa menggunakan MI. Hasil dari pengujian akan dibandingkan dan di evaluasi untuk melihat pengaruh metode seleksi fitur MI dalam meningkatkan kinerja MKNN. Pengujian dilakukan dengan nilai k masukkan 3, 5, 7 dan 9 serta nilai threshold 0,018, 0,01 dan 0,008. Hasil pengujian yang telah didapatkan menunjukkan Metode MKNN dengan MI memiliki kinerja yang lebih baik dibandingkan dengan MKNN tanpa MI. Seleksi fitur MI mampu mengurangi jumlah fitur pada dataset sehingga meningkatkan kinerja klasifikasi. Selisih peningkatan akurasi tertinggi terjadi pada saat nilai $k=3$ dengan nilai threshold 0,018 yakni sebesar 22%.

Kata Kunci : Instagram, Komentar, *Modified K-Nearest Neighbor*, *Mutual Information*, Spam.

Palembang, 19 Agustus 2021

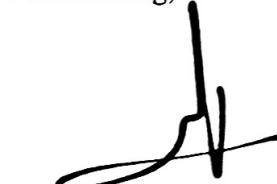
Mengetahui,

Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

Pembimbing,



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

KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala rahmat dan karunia-Nya penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir 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 material maupun moril selama proses pembuatan tugas akhir ini. Adapun penulis merujuk secara khusus pihak yang telah membantu sebagai berikut :

1. Bapak Suparjo dan ibu Partiyah, selaku orang tuaku yang tersayang. Kepada kakak-kakakku, Sunarsih, Sugio, Sunaryo, yuyun, Miron, Triastutik, Bastari Herdianto yang telah memberikan dukungan materil dan semangat hingga saya dapat menyelesaikan masa perkuliahan.
2. Bapak Jaidan Jauhari, S.pd., M.T. selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika.
4. Ibu Alvi Syahrini Utami, M.Kom selaku pembimbing akademik yang telah memberikan bimbingan serta arahan dalam urusan akademik.
5. Bapak Dr. Abdiansah., S.Kom., M.Cs selaku pembimbing I yang telah memberikan arahan, bimbingan, kritik dan saran kepada penulis untuk menyelesaikan tugas akhir.
6. Ibu Alvi Syahrini Utami, M.Kom dan bapak Kanda Januar M, M.T selaku Dosen Pengaji I dan Dosen Pengaji 2 yang telah memberikan masukan dan arahan kepada penulis dalam menyempurnakan tugas akhir ini.
7. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika yang telah memberikan ilmu kepada penulis selama masa perkuliahan.
8. Staff Jurusan Teknik Informatika beserta staff Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi dan akademik selama masa perkuliahan.

9. Sahabat-sahabat ku yang telah memberikan semangat serta motivasi dalam proses penulisan tugas akhir ini.
10. Keponakan-keponakan ku yang telah memberikan semangat serta dukungannya dalam proses penulisan tugas akhir.
11. Cesil Oktavia Hersa, Nurul Hannah Djati, dan M. Abdul Aziz, sahabat seperjuangan yang selalu ada saat susah dan senang saat masa-masa perkuliahan yang tidak akan pernah terlupakan.
12. Geng Layo/ Geng Mama Mokuzai/ Geng Fasco/ Gang Wibu, sahabat seperjuangan yang selalu kompak dalam menciptakan kesenangan dan menciptakan momen-momen kebersamaan saat masa perkuliahan yang tidak akan pernah bisa dilupakan.
13. Semua pihak lainnya yang tidak dapat disebutkan satu persatu yang telah berperan bagi penulis dalam menyelesaikan tugas akhir, Terima Kasih banyak atas semuanya.

Semoga tugas akhir 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 tugas akhir ini, terdapat beberapa penjelasan yang kurang sempurna. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun agar tugas akhir ini menjadi karya tulis yang sempurna supaya terciptanya bekal pengetahuan yang baik bagi peneliti di masa depan.

Palembang, 24 Agustus 2021

Desi Indriani

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN SKRIPSI.....	ii
HALAMAN TANDA LULUS UJIAN SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
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-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Instagram	II-1
2.2.2 Spam	II-2
2.2.3 Analisis Sentimen	II-4
2.2.4 Pra-Pengolahan	II-5
2.2.5 <i>Mutual Information</i>	II-7

2.2.6	Pembobotan TF-IDF	II-8
2.2.7	<i>Modified K-Nearest Neighbor</i> (MKNN)	II-9
2.2.9	<i>Confusion Matrix</i>	II-12
2.2.10	<i>Rational Unified Process</i>	II-13
2.3	Penelitian Lain yang Relevan.....	II-14
2.4	Kesimpulan.....	II-18
BAB III METODE PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Metode Pengumpulan Data	III-1
3.2.1	Jenis dan Sumber Data.....	III-1
3.2.2	Tahap Pengumpulan Data.....	III-2
3.3	Tahapan Penelitian	III-4
3.3.1	Mengumpulkan Data	III-4
3.3.2	Menentukan Kerangka Kerja Penelitian	III-5
3.3.3	Menentukan Kriteria Pengujian	III-9
3.3.4	Menentukan Format Data Pengujian	III-10
3.3.5	Menentukan Alat Bantu Penelitian.....	III-11
3.3.6	Melakukan Pengujian Penelitian	III-11
3.3.7	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-12
3.4	Metode Pengembangan Perangkat Lunak	III-13
3.4.1	Fase Insepsi.....	III-13
3.4.2	Fase Elaborasi	III-14
3.4.3	Fase Konstruksi	III-14
3.4.4	Fase Transisi	III-15
3.4	Kesimpulan.....	III-15
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.....	IV-2
4.2.3	Analisis dan Perancangan	IV-3
4.2.2	Implementasi.....	IV-21
4.3	Fase Elaborasi	IV-27

4.3.1	Pemodelan Bisnis.....	IV-27
4.3.2	Kebutuhan.....	IV-28
4.3.3	Analisis dan Perancangan	IV-29
4.4	Fase Konstruksi.....	IV-36
4.4.1	Kebutuhan Sistem.....	IV-36
4.4.2	Kelas Diagram	IV-36
4.4.3	Implementasi.....	IV-37
4.5	Fase Transisi.....	IV-40
4.5.1	Kebutuhan.....	IV-40
4.5.2	Analisis dan Perancangan	IV-40
4.5.3	Implementasi.....	IV-42
4.6	Kesimpulan.....	IV-45
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
5.3.1	Analisis Nilai k	V-10
5.3.2	Analisis jumlah Fitur	V-17
5.4	Kesimpulan.....	V-18
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA		ix
LAMPIRAN		L-1

DAFTAR TABEL

Tabel II- 1. Truth Table Fitur Terhadap Kelas.....	II-8
Tabel II- 2. <i>Confusion Matrix</i>	II-12
Tabel III-1. Contoh data yang telah dikumpulkan	III-1
Tabel III-2. Daftar 5 akun artis/aktor dengan pengikut di atas 10 juta.	III-2
Tabel III- 3. Rancangan tabel confusion matrix hasil klasifikasi.....	III-10
Tabel III-4. Rancangan Tabel Confusion Matrix Hasil Klasifikasi dengan seleksi fitur.....	III-10
Tabel III- 5. Rancangan Tabel hasil pengujian	III-10
Tabel III- 6. Rancangan Tabel Hasil Pengujian	III-11
Tabel III- 7. Tabel Hasil Analisis Klasifikasi	III-13
Tabel IV- 1. Kebutuhan Fungsional Perangkat Lunak.....	IV-2
Tabel IV- 2. Kebutuhan Non-Fungsional Perangkat Lunak	IV-2
Tabel IV- 3. Contoh Data Komentar.....	IV-5
Tabel IV- 4. Hasil Cleaning Data Komentar.....	IV-5
Tabel IV- 5. Hasil Case Folding Data Komentar.....	IV-6
Tabel IV- 6. Hasil Normalisasi Data Komentar.....	IV-6
Tabel IV- 7. Hasil Tokenizing Data Komentar	IV-7
Tabel IV- 8. Hasil Stopword Removal Data Komentar	IV-7
Tabel IV- 9. Hasil Stemming Data Komentar.....	IV-8
Tabel IV- 10. Perhitungan Nilai TF dan IDF dari Data Hasil Pra-pengolahan..	IV-9
Tabel IV- 11. Perhitungan Nilai TF-IDF dari Data Hasil Pra-pengolahan	IV-9
Tabel IV- 12. Tabel data latih yang telah dilakukan prapengolahan	IV-10
Tabel IV- 13. Data Latih	IV-11
Tabel IV- 14. Tabel Jarak Latih	IV-12
Tabel IV- 15. Tabel Tingkat Kesamaan Kelas.....	IV-12
Tabel IV- 16. Tabel Nilai Validity.....	IV-13
Tabel IV- 17. Tabel Jarak Data Latih dengan Data Uji	IV-13
Tabel IV- 18. Tabel Nilai Weight Voting	IV-14
Tabel IV- 19. Tabel 3 Nilai Tertinggi Dari <i>Weight Voting</i> data Uji dan Latih	IV-14
Tabel IV- 20. Tabel Hasil Kelas Prediksi Data Uji.....	IV-15
Tabel IV- 21. Perhitungan Nilai Mutual Information	IV-15
Tabel IV- 22. Data hasil seleksi fitur	IV-16
Tabel IV- 23. Perhitungan nilai TF dan IDF fitur hasil dari seleksi fitur	IV-16
Tabel IV- 24. Perhitungan nilai TF_IDF fitur hasil dari seleksi fitur	IV-17
Tabel IV- 25. Jarak Antar Data Latih Hasil Seleksi Fitur.....	IV-18
Tabel IV- 26. Tabel Tingkat Kesamaan Data Hasil Seleksi	IV-18
Tabel IV- 27. Nilai Validitas Data Hasil seleksi Fitur	IV-19
Tabel IV- 28. Jarak Data Uji dengan Data Latih Hasil Seleksi Fitur.....	IV-19
Tabel IV- 29. Nilai <i>Weight Voting</i>	IV-20
Tabel IV- 30. Tiga Nilai Tertinggi <i>Weight Voting</i> Data Uji Seleksi Fitur.....	IV-20
Tabel IV- 31. Kelas Prediksi Data Uji Data Seleksi Fitur	IV-21
Tabel IV- 32. Tabel Definisi Aktor.....	IV-22
Tabel IV- 33. Tabel Definisi <i>Use Case</i>	IV-23

Tabel IV- 34. Skenario Use Case Memuat Data	IV-24
Tabel IV- 35. Skenario Melakukan Klasifikasi Mennggunakan Algoritma MK-NN	IV-25
Tabel IV- 36. Skenario Klasifikasi Menggunakan Algoritma MKNN+MI.....	IV-26
Tabel IV- 37. Tabel Implementasi Kelas	IV-38
Tabel IV- 38. Rencana Pengujian Use Case Memuat Data	IV-41
Tabel IV- 39. Rencana Pengujian Use Case Klsifikasi Menggunakan Modified K-Nearest Neighbor.	IV-41
Tabel IV- 40. Rencana pengujian Use Case Melakukan Klasifikasi Menggunakan Modified K-Nearest Neighbor den Seleksi Fitur Mutual Information	IV-41
Tabel IV- 41. Pengujian Use Case Memproses Data.....	IV-44
Tabel IV- 42. Pengujian Use Case Melakukan Klasifikasi Menggunakan MKNN	IV-45
Tabel IV- 43. Pengujian Use Case Melakukan Klasifikasi Menggunakan MKNN dan Seleksi Fitur MI.....	IV-46
Tabel V- 1. Tabel <i>Confusion Matrix</i> hasil klasifikasi menggunakan MKNN+MI	V-2
Tabel V- 2. Data Hasil Evaluasi Klasifikasi MKNN tanpa seleksi fitur.....	V-2
Tabel V- 3. Tabel <i>Confusion Matrix</i> hasil klasifikasi menggunakan MKNN dan seleksi fitur MI dengan treshold = 0,018	V-4
Tabel V- 4. Data Hasil Evaluasi Klasifikasi MKNN dan Seleksi Fitur MI dengan Threshold = 0,018	V-4
Tabel V- 5. Tabel <i>Confusion Matrix</i> Hasil Klasifikasi Menggunakan MKNN dan Seleksi Fitur MI dengan Threshold = 0,01	V-6
Tabel V- 6. Data Hasil Evaluasi Klasifikasi MKNN+MI dengan Threshold = 0,01	V-6
Tabel V- 7. Tabel <i>Confusion Matrix</i> Hasil Klasifikasi Menggunakan MKNN dan Seleksi Fitur MI dengan Threshold = 0,008	V-8
Tabel V- 8. Data Hasil Evaluasi Klasifikasi MKNN dan Seleksi Fitur MI dengan Threshold = 0,008	V-8
Tabel V- 9. Perbandingan Hasil Klasifikasi Menggunakan MKNN dan MKNN dengan Seleksi Fitur MI	V-10

DAFTAR GAMBAR

Gambar II-1. Persentasi umur pengguna media sosial di Indonesia	II-2
Gambar II- 2. Contoh Komentar Spam di Instagram.....	II-3
Gambar II- 3 Asitektur Rational Unified Process	II-13
Gambar III- 1. Komentar yang telah diberi label.....	III-3
Gambar III-2. Tahapan penelitian	III-4
Gambar III-3. Diagram Alur Proses Umum Perangkat Lunak	III-6
Gambar IV- 1. Diagram <i>Use Case</i>	IV-22
Gambar IV- 2. Rancangan Antarmuka Perangkat Lunak	IV-28
Gambar IV- 3. Diagram Aktivitas Memproses Data Masukan.....	IV-30
Gambar IV- 4. Diagram Aktivitas Melakukan Klasifikasi Menggunakan Algoritma Modified K-Nearest Neighbor.....	IV-31
Gambar IV- 5. Diagram Aktivitas Melakukan Klasifikasi Menggunakan Algoritma Modified K-Nearest Neighbor dan seleksi fitur Mutual Information	IV-32
Gambar IV-6. Sequence Diagram Memuat Data	IV-33
Gambar IV-7. <i>Sequence Diagram</i> Klasifikasi Menggunakan MKNN	IV-34
Gambar IV-8. Sequence Diagram Klasifikasi Menggunakan MKNN+MI	IV-35
Gambar IV-9. Diagram Kelas Perngakat Lunak	IV-37
Gambar IV-10.Tampilan Antarmuka Halaman Utama Perangkat Lunak.....	IV-39
Gambar IV-11.Tampilan Antarmuka dengan Hasil Klasifikasi dari Metode yang dipilih.....	IV-39
Gambar V- 1. Grafik Hasil Pengujian Setiap Nilai k.....	V-3
Gambar V- 2. Grafik Hasil Evaluasi Denangan Nilai Threshold 0,018.....	V-5
Gambar V- 3. Grafik Evaluasi Klasifikasi MKNN+MI dengan Threshold 0,01	V-7
Gambar V- 4. Grafik Evaluasi Klasifikasi MKNN+MI dengan Threshold 0,008	V-9
Gambar V- 5. Grafik Perbandingan Hasil Evaluasi Klasifikasi k=3	V-11
Gambar V- 6. Grafik Perbandingan Hasil Evaluasi Klasifikasi k=5	V-12
Gambar V- 7. Grafik Perbandingan Hasil Evaluasi Klasifikasi k=7	V-14
Gambar V- 8. Grafik Perbandingan Hasil Evaluasi Klasifikasi k=9	V-15
Gambar V- 9. Grafik Perbandingan Jumlah Fitur	V-17

DAFTAR LAMPIRAN

Lampiran 1. Kode Program.....L-1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini berisi tentang penjelasan poin penting yang menjadi dasar masalah pembuatan tugas akhir. Poin penting ini meliputi latar belakang, rumusan masalah, tujuan dan manfaat penelitian serta batasan masalah penelitian.

1.2 Latar Belakang

Instagram merupakan salah satu media sosial yang digunakan sebagai sarana untuk mengunggah foto/gambar. Salah satu pengguna Instagram adalah para artis/aktor Indonesia. Semakin populer artis/aktor maka semakin banyak pengikutnya, maka semakin banyak pula komentar spam yang didapat. Banyaknya komentar spam pada sebuah unggahan akan mengganggu alur diskusi yang membuat pengguna kesulitan mencari informasi secara cepat dan tepat (Chrismanto dan Lukito, 2017).

Salah satu yang dapat dilakukan untuk masalah komentar spam pada Instagram adalah dengan pengklasifikasian komentar untuk di golongkan berdasarkan kategori spam atau bukan spam. Metode yang dapat digunakan dalam proses klasifikasi teks mining yaitu *Naive Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbor* (Ipmawati et al., 2017).

Chrismanto dan Lukito (2017) melakukan penelitian dengan membandingkan antara metode *K-Nearest Neighbor* dengan *Naive Bayes* untuk

klasifikasi komentar spam di Instagram. Hasil penelitian didapatkan bahwa metode *K-Nearest Neighbor* mendapatkan akurasi tertinggi pada saat nilai k=3 dengan rata-rata nilai akurasi sebesar 87,07% dan memiliki tingkat akurasi lebih tinggi dibandingkan dengan metode *Naive Bayes* dengan selisih nilai akurasi sebesar 12,67%.

Namun dari algoritma *K-Nearest Neighbor* memiliki kekurangan yaitu nilai bias k, komputasi yang kompleks, keterbatasan memori, dan rentan terhadap variabel non-informatif (Mutrofin et al., 2015). Salah satu metode perbaikan dari *K-Nearest Neighbor* adalah metode *Modified K-Nearest Neighbor*. Metode *Modified K-Nearest Neighbor* diperbaiki untuk meningkatkan hasil akurasi dari metode *K-Nearest Neighbor*.

Rozi dan Sulistyawati (2019) melakukan penelitian dengan menggunakan metode *Modified K-Nearest Neighbor* untuk mengklasifikasi berita *hoax* pilpres. Pada penelitian tersebut dilakukan proses pembobotan menggunakan TF-IDF . Hasil dari penelitian didapatkan nilai akurasi sebesar 92,30%. Pada penelitian Nur et al. (2019) dilakukan perbandingan pada beberapa algoritma perbaikan dari *K-Nearest Neighbor* untuk klasifikasi tafsir Qur'an. Didapatkan hasil penelitian bahwa algoritma *Modified K-Nearest Neighbor* menghasilkan akurasi yang lebih baik dibandingkan dengan algoritma *Fuzzy K-Nearest Neighbor* dan *K-Nearest Neighbor* yaitu sebesar 98,12%. Sedangkan untuk algoritma *Fuzzy K-Nearest Neighbor* dan *K-Nearest Neighbor* nilai akurasi yang didapat sebesar 88,3% dan 98,01%.

Namun dari metode klasifikasi dengan *Modified K-Nearest Neighbor* memiliki kekurangan dimana proses klasifikasi menggunakan semua fitur yang ada. Hal ini dapat menyebabkan kesalahan deteksi karena adanya fitur-fitur yang tidak relevan terhadap objek. Untuk mengatasi masalah tersebut dibutuhkan metode seleksi fitur untuk mengurangi jumlah kata yang dianggap tidak relevan terhadap objek. Salah satu metode seleksi fitur yang bisa digunakan adalah *Mutual Information*.

Ulfa et al. (2018) melakukan penelitian mengenai klasifikasi *tweet* menggunakan *Naive Bayes* dan seleksi fitur *Mutual Information*, mendapatkan hasil akurasi sebesar 97,9%. Untuk klasifikasi *tweet* tanpa seleksi fitur menghasilkan akurasi sebesar 96,2%. Hasil ini menunjukan bahwa penggunaan seleksi fitur *Mutual Information* dapat meningkatkan nilai akurasi.

Berdasarkan hasil penelitian yang telah dibahas di atas, maka dalam penelitian ini akan dilakukan klasifikasi untuk komentar spam pada media sosial Instagram menggunakan algoritma *Mutual Information* dan *Modified K-Nearest Neighbor*. Dengan metode yang diterapkan diharapkan dapat menyelesaikan masalah terkait komentar spam pada media sosial Instagram.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah dibahas rumusan masalah dalam penelitian ini adalah:

1. Bagaimana melakukan klasifikasi komentar spam pada Instagram menggunakan metode *Modified K-Nearest Neighbor*?
2. Bagaimana pengaruh seleksi fitur *Mutual Information* terhadap hasil klasifikasi?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun perangkat lunak yang dapat melakukan klasifikasi komentar spam pada Instagram menggunakan metode *Modified K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* dengan seleksi fitur *Mutual Information*.
2. Mengetahui hasil kinerja dari klasifikasi komentar spam pada Instagram menggunakan metode *Modified K-Nearest Neighbor* dan metode *Modified K-Nearest Neighbor* dengan seleksi fitur *Mutual Information*.
3. Mengetahui pengaruh seleksi fitur *Mutual Information* terhadap klasifikasi menggunakan metode *Modified K-Nearest Neighbor*.

1.5 Manfaat Penelitian

Manfaat yang dihasilkan dari penelitian ini adalah:

1. Hasil penelitian dapat dijadikan sebagai rujukan untuk penelitian yang relevan.

2. Memperoleh informasi mengenai hasil klasifikasi komentar spam menggunakan metode *Modified K-Nearest Neighbor* dan metode *Modified K-Nearest Neighbor* dengan seleksi fitur *Mutual Information*.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Menggunakan data dari 5 akun Instagram artis/aktor Indonesia yang memiliki pengikut lebih dari 10 juta.
2. Menggunakan 250 data komentar bahasa Indonesia.
3. Data yang dipakai merupakan komentar spam yang berupa komentar *online shop*, peninggi badan, pemutih, penambah berat badan, penurun berat badan, *link website* dan lain sebagainya.
4. Menggunakan 2 kelas yaitu spam dan bukan spam.
5. Pengujian menggunakan 3 nilai ambang batas/*threshold*.
6. Kriteria nilai k yang digunakan untuk klasifikasi MKNN adalah 3, 5, 7, 9.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir penelitian adalah sebagai berikut :

BAB I. PENDAHULUAN

Pada bab ini membahas hal utama dalam melandasi pengembangan perangkat lunak yang terdiri dari latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan bagaimana penulisan sistematika yang baik.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas landasan teori yang akan menjadi dasar dalam penelitian ini, seperti definisi Instagram, komentar spam, algoritma *Modified K-Nearest Neighbor*, *Mutual Information* dan beberapa teori yang berkaitan dengan penelitian.

BAB III. METODOLOGI PENELITIAN

Pada bab ini berisi tentang tahapan yang akan dilakukan dalam penelitian tugas akhir. Tahapan ini berisi tentang rincian metode yang akan digunakan dalam penelitian, pengumpulan dan analisa data, serta membahas rancangan sistem yang dikembangkan. Tahapan ini dibuat dengan mengacu pada suatu kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini berisi tentang analisis proses pengolahan dan perhitungan data, analisis perangkat lunak, perancangan dan pengembangan perangkat lunak, dan hasil pengujian dari perangkat lunak.

BAB V. HASIL DAN ANALISA PENELITIAN

Pada bab ini dilakukan penyajian hasil pengujian berdasarkan langkah-langkah yang telah dirancang sebelumnya. Analisis hasilnya disertakan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan keseluruhan dari penelitian ini dan juga berisi saran-saran yang diharapkan dapat berguna dalam pengembangan perangkat lunak selanjutnya.

1.8 Kesimpulan

Dari hasil pembahasan yang telah dijabarkan pada setiap pokok pikiran, penelitian akan dilakukan dengan metode yang ditentukan serta sistematika penulisan yang baik dan benar.

DAFTAR PUSTAKA

- Andrian, R., Sakethi, D., & Chairuddin, M. (2014). Pengembangan Sistem Informasi Penelitian Dan Pengabdian Dosen Jurusan Ilmu Komputer Menggunakan Metode Rational Unified Process (Rup). *Jurnal Komputasi*, Vol 2, No 2 (2014), 1–8. <http://jurnal.fmipa.unila.ac.id/index.php/komputasi/article/view/1090>
- Arifin, Y. T. (2016). Komparasi Fitur Seleksi Pada Algoritma Support Vector Machine Untuk Analisis Sentimen Review. *Jurnal Informatika (JI) UBSI*, 3(September), 191–199.
- Astagiri, B. (2010). Spamming Dalam Perspektif Hukum Pidana. *Yuridika*, 25(3), 89–102. <https://doi.org/10.20473/ydk.v25i3.257>
- Chrismanto, A. R., & Lukito, Y. (2017a). Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes. *Jurnal ULTIMATICS*, 9(1), 50–58. <https://doi.org/10.31937/ti.v9i1.564>
- Chrismanto, A. R., & Lukito, Y. (2017b). Klasifikasi Komentar Spam Pada Instagram Berbahasa Indonesia Menggunakan K-NN. *Seminar Nasional Teknologi Informasi Kesehatan (SNATIK)*, 2, 298–306.
- Hidayat, andi nurul. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Elektronik Sistim Informasi Dan Komputer (Jesik)*, 1(1), 1–7.
- Ipmawati, J., Kusrini, & Taufiq Luthfi, E. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, 6(1), 28–36.
- Irham, L. G., Adiwijaya, A., & Wisesty, U. N. (2019). Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, 3(4), 284. <https://doi.org/10.30865/mib.v3i4.1410>
- Ismanto, B., & Amalia, N. (2018). Peningkatan Akurasi Pada Modified K-NN Untuk Klasifikasi Pengajuan Kredit Koperasi Dengan Menggunakan

- Algoritma Genetika. *IC-Tech*, 3(2), 66–70. ejurnal.stmik-wp.ac.id/index.php/ictech/article/view/42/40
- Karunia, S. A., Saptono, R., & Anggrainingsih, R. (2017). Online News Classification Using Naive Bayes Classifier with Mutual Information for Feature Selection. *Jurnal Ilmiah Teknologi Dan Informasi*, 6(1), 10–15. <https://doi.org/10.20961/its.v6i1.11114>
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(11), 4704–4713.
- Mutrofin, S., Izzah, A., Kurniawardhani, A., & Masrur, M. (2015). Optimasi teknik klasifikasi modified k nearest neighbor menggunakan algoritma genetika. *Jurnal Gamma*, 10(1).
- Nugraha, B., & Akbar, M. F. (2018). *Instagram Active User Communication Behaviour*. 2(2), 95–101.
- Nur, A., Syarifandi, S., & Amin, S. (2019). *Implementation of Text Mining Classification as a Model in the Conclusion of Tafsir Bil Ma ’tsur and Bil RA ’YI Contents*. 1, 2789–2795. <https://doi.org/10.35940/ijeat.A9780.109119>
- Prasetyo, A. R., Indrianti, & Adikara, P. P. (2018). Klasifikasi Hoax Pada Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Modified K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(12), 7466–7473.
- Rozi, F. N., & Sulistyawati, D. H. (2019). KLASIFIKASI BERITA HOAX PILPRES MENGGUNAKAN METODE MODIFIED K-NEAREST NEIGHBOR DAN PEMBOBOTAN MENGGUNAKAN TF-IDF. *KONVERGENSI*, Vol 15, No 1 (2019). <http://jurnal.untagsby.ac.id/index.php/KONVERGENSI/article/view/2828>
- Ulfah, M. A., Irmawati, B., & Husodo, A. Y. (2018). Twitter Sentiment Analysis using Naive Bayes Classifier with Mutual Information Feature Selection. Ulfah, M. A., Irmawati, B., & Husodo, A. Y. (2018). Twitter Sentiment Analysis using Naive Bayes Classifier with Mutual Information Feature Selection. Journal

of Com. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, 2(2), 106–111.