

**ANALISIS SENTIMEN TERHADAP KEMENTRIAN KOMUNIKASI
DAN INFORMATIKA PADA MEDIA SOSIAL INSTAGRAM
MENGUNAKAN METODE *NAÏVE BAYES* DAN
*K-NEAREST NEIGHBOR***

SKRIPSI

Oleh :

Muhammad Anwar Lirboyo

201910225319



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2023**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Skripsi : Analisis Sentimen Terhadap Kementrian
Komunikasi dan Informatika Pada Media
Sosial Instagram Menggunakan Metode *Naïve*
Bayes Dan K-Nearest Neighbor

Nama Mahasiswa : Muhammad Anwar Lirboyo

Nomor Pokok Mahasiswa : 201910225319

Program Studi/Fakultas : Informatika / Ilmu Komputer

Pembimbing I

Pembimbing II

Adi Muhajirin, S.Kom., M.Kom., M.M.

Prima Dina Atika, S.Kom., M.Kom.

NIDN. 0318038501

NIDN. 0311037107

LEMBAR PENGESAHAN

Judul Skripsi : Analisis Sentimen Terhadap Kementrian Komunikasi
Dan Informatika Pada Media Sosial Instagram
Menggunakan Metode Naïve Bayes Dan K-Nearest
Neighbor

Nama Mahasiswa : Muhammad Anwar Lirboyo

Nomor Pokok Mahasiswa : 201910225319

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 31 Juli 2023

Jakarta, 09 Agustus 2023

MENGESAHKAN,

Ketua Tim Penguji : Dr. Mujiono Sadikin, MT., CISA., CGEIT.

NIDN: 0406127002

Penguji I : Khairunnisa Fadhillah Ramdhanisa, S.Si., M.Si.

NIDN: 0328039201

Penguji II : Adi Muhajirin, S.Kom., M.Kom., M.M.

NIDN: 0318038501

MENGETAHUI,

Ketua

Dekan

Program Studi Informatika

Fakultas Ilmu Komputer

Ahmad Fathurrozi, S.E., M.M.S.I.

NIP. 2012486

Dr. Dra. Tyastuti Sri Lestari, M.M.

NIP. 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini:

Nama : Muhammad Anwar Lirboyoy
NPM : 201910225219
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Analisis Sentimen Terhadap Kementrian
Komunikasi Dan Informatika Pada Media Sosial
Instagram Menggunakan Metode *Naïve Bayes* Dan
K-Nearest Neighbor

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Jakarta, 10 Agustus 2023

Penulis



Muhammad Anwar Lirboyoy

ABSTRAK

Muhammad Anwar Lirboyo. 201910225319. analisis sentimen terhadap kementrian komunikasi dan informatika pada media sosial instagram menggunakan metode *naïve bayes* dan *k-nearest neighbor* program studi informatika universitas bhayangkara jakarta raya. Bekasi: Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya. 2023.

Masalah Penelitian ini adalah Belum diketahui skor akurasi untuk analisis sentimen terhadap komentar pada akun Kemenkominfo menggunakan metode *naïve bayes* dan *k-nearest neighbor*, Belum adanya teknik klasifikasi untuk mengelompokkan sentimen positif, negatif dan netral yang diberikan masyarakat pada akun Kemenkominfo. Tujuan Penelitian ini adalah Penerapan algoritma *naïve bayes* dan *k-nearest neighbor* untuk melakukan klasifikasi analisis sentimen terhadap tanggapan atau komentar masyarakat di media sosial Instagram pada akun Kementerian Komunikasi dan Informatika.

Naïve bayes merupakan metode pengklasifikasian probabilistik sederhana. Metode ini akan menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan sedangkan *K-nearest neighbor* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Dataset komentar yang didapatkan dari instagram disimpan dalam bentuk file CSV yang selanjutnya diklasifikasikan menjadi 3 macam sentimen diantaranya sentimen positif, negatif dan netral. Data Analisis Sentimen positif berjumlah 45,4%, negatif berjumlah 40,2%, netral berjumlah 14,4%. *Naïve Bayes* dengan *accuracy score* 0,99% dan *K-Nearest Neighbor* dengan *accuracy score* 100%. hasil analisis sentimen *K-Nearest Neighbor* sedikit lebih baik daripada *Naïve Bayes*.

Kata Kunci : Kemenkominfo, Analisis Sentimen, *Naïve Bayes*, *K-Nearest Neighbor*, *Accuracy Score*.

ABSTRACT

Muhammad Anwar Lirboyo. 201910225319. *sentiment analysis towards the ministry of communication and information on instagram social media using naïve bayes and k-nearest neighbor methods, informatics study program, bhayangkara university, jakarta raya. Bekasi: Faculty of Computer Science, Bhayangkara University, Jakarta Raya. 2023.*

The problem of this study is that the score of accuracy for the sentiment analysis for the ministry of communication and information on social instagram using the naïve bayes and k-nearest neighbor's methods, and yet no classification techniques for classify positive, negative and neutral sentiments that people provide on the ministry account. The aim of the study is the file application of the naïve bayes and k-nearest neighbor algorithm to file a sentiment analysis on social media and instagram accounts on the ministry of communication and information.

Naïve bayes was the simple probability rating method. This method would calculate aset of probabilities by adding up the frequency and value combination that the dataset provided while k-nearest neighbor was a method for classifying objects based on learning data that are closest to the object.

Datasset comments obtained from instagram are kept in the form of CSV files which are subsequently classified into 3 kinds of sentiments among those of positive, negative and neutral sentiments. Positive sentiment analysis data is 45.4%, negative is 40.2%, neutral, 14.4 percent. Naive bayes with a 0.99% drop-score and a k-nearest neighbor with a 100-score. K-nearest neighbor's sentiments analysis was a bit better than naïve bayes.

Keywords : *Kemenkominfo, Sentiment Analysis, Naïve Bayes, K-Nearest Neighbor, Accuracy Score.*

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Muhammad Anwar Lirboyo
NPM : 201910225319
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty-Free Right)**, atas karya ilmiah saya yang berjudul :

ANALISIS SENTIMEN TERHADAP KEMENTERIAN KOMUNIKASI DAN INFORMATIKA PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN METODE *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR*

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 09 Agustus 2023
Yang Menyatakan



Muhammad Anwar Lirboyo

ABSTRAK

Muhammad Anwar Lirboyo. 201910225319. analisis sentimen terhadap kementrian komunikasi dan informatika pada media sosial instagram menggunakan metode *naïve bayes* dan *k-nearest neighbor* program studi informatika universitas bhayangkara jakarta raya. Bekasi: Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya. 2023.

Masalah Penelitian ini adalah Belum diketahui skor akurasi untuk analisis sentimen terhadap komentar pada akun Kemenkominfo menggunakan metode *naïve bayes* dan *k-nearest neighbor*, Belum adanya teknik klasifikasi untuk mengelompokkan sentimen positif, negatif dan netral yang diberikan masyarakat pada akun Kemenkominfo. Tujuan Penelitian ini adalah Penerapan algoritma *naïve bayes* dan *k-nearest neighbor* untuk melakukan klasifikasi analisis sentimen terhadap tanggapan atau komentar masyarakat di media sosial Instagram pada akun Kementerian Komunikasi dan Informatika.

Naïve bayes merupakan metode pengklasifikasian probabilistik sederhana. Metode ini akan menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan sedangkan *K-nearest neighbor* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Dataset komentar yang didapatkan dari instagram disimpan dalam bentuk file CSV yang selanjutnya diklasifikasikan menjadi 3 macam sentimen diantaranya sentimen positif, negatif dan netral. Data Analisis Sentimen positif berjumlah 45,4%, negatif berjumlah 40,2%, netral berjumlah 14,4%. *Naïve Bayes* dengan *accuracy score* 0,99% dan *K-Nearest Neighbor* dengan *accuracy score* 100%. hasil analisis sentimen *K-Nearest Neighbor* sedikit lebih baik daripada *Naïve Bayes*.

Kata Kunci : Kemenkominfo, Analisis Sentimen, *Naïve Bayes*, *K-Nearest Neighbor*, *Accuracy Score*.

ABSTRACT

Muhammad Anwar Lirboyo. 201910225319. *sentiment analysis towards the ministry of communication and information on instagram social media using naïve bayes and k-nearest neighbor methods, informatics study program, bhayangkara university, jakarta raya. Bekasi: Faculty of Computer Science, Bhayangkara University, Jakarta Raya. 2023.*

The problem of this study is that the score of accuracy for the sentiment analysis for the ministry of communication and information on social instagram using the naïve bayes and k-nearest neighbor's methods, and yet no classification techniques for classify positive, negative and neutral sentiments that people provide on the ministry account. The aim of the study is the file application of the naïve bayes and k-nearest neighbor algorithm to file a sentiment analysis on social media and instagram accounts on the ministry of communication and information.

Naïve bayes was the simple probability rating method. This method would calculate aset of probabilities by adding up the frequency and value combination that the dataset provided while k-nearest neighbor was a method for classifying objects based on learning data that are closest to the object.

Datasset comments obtained from instagram are kept in the form of CSV files which are subsequently classified into 3 kinds of sentiments among those of positive, negative and neutral sentiments. Positive sentiment analysis data is 45.4%, negative is 40.2%, neutral, 14.4 percent. Naive bayes with a 0.99% drop-score and a k-nearest neighbor with a 100-score. K-nearest neighbor's sentiments analysis was a bit better than naïve bayes.

Keywords : *Kemenkominfo, Sentiment Analysis, Naïve Bayes, K-Nearest Neighbor, Accuracy Score.*

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xv
DAFTAR RUMUS	xvi
DAFTAR LAMPIRAN	xvii
Bab I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Identifikasi Masalah	6
1.3 Rumusan Masalah	6
1.4 Batasan Masalah.....	7
1.5 Tujuan Penelitian.....	7
1.6 Hasil/Dampak Yang Dirasakan	7
1.7 Sistematika Penulisan.....	8
BAB II LANDASAN TEORI	9
2.1 Tinjauan Pustaka	9
2.2 Teori Penghubung	12
2.2.1 Sentiment Analysis	12

2.2.3	<i>Naïve Bayes</i>	13
2.2.4	<i>Text Preprocessing</i>	13
2.2.5	<i>Instagram</i>	15
2.2.6	<i>Confusion Matrix</i>	16
2.2.7	<i>Web Scraping</i>	17
2.2.8	<i>Machine Learning</i>	17
2.2.9	<i>Data Mining</i>	18
2.2.10	<i>Flowchart</i>	18
2.2.11	<i>Fishbone Diagram</i>	20
2.2.12	<i>Python</i>	20
2.2.13	<i>Google Colaboratory</i>	21
2.2.14	<i>Slovin</i>	21
2.2.15	<i>K-Nearest Neighbor</i>	22
BAB III METODOLOGI PENELITIAN		23
3.1	<i>Objek Penelitian</i>	23
3.2	<i>Tempat dan Waktu Penelitian</i>	23
3.3	<i>Kerangka Pemikiran</i>	23
3.4	<i>Pembuatan Model Machine Learning</i>	24
3.5	<i>Metode Pengumpulan Data</i>	25
3.6	<i>Sampel Data</i>	25
3.7	<i>Preprocessing</i>	26
3.7.1	<i>Cleansing</i>	27
3.7.2	<i>Case Folding</i>	28
3.7.3	<i>Tokenizing</i>	28
3.7.4	<i>Removing Stopword</i>	28
3.7.5	<i>Stemming KBBI</i>	29

3.8	Pelabelan Data	29
3.9	Vektorisasi	29
3.10	Klasifikasi Dengan <i>Naïve Bayes</i> dan <i>KNN</i>	30
3.11	Analisis Kebutuhan Perangkat	30
BAB IV HASIL DAN PEMBAHASAN		32
4.1	Pengambilan Data	32
4.2	Analisis Deskriptif	34
4.3	<i>Preprocessing</i>	34
4.3.1	<i>Data Cleansing</i>	34
4.3.2	<i>Case Folding</i>	35
4.3.3	<i>Tokenizing</i>	36
4.3.4	<i>Removing Stopword</i>	36
4.4	Pelabelan Data	39
4.5	Visualisasi <i>Wordcloud</i>	41
4.6	Vektorisasi	45
4.7	Pembagian Data Latih dan Uji	46
4.8	Klasifikasi dengan Metode <i>Naïve Bayes</i>	47
4.9	Klasifikasi dengan Metode <i>K-Nearest Neighbor</i>	49
4.10	Infografis Analisis Sentimen Akun Kemenkominfo	53
BAB V PENUTUP		55
5.1	Kesimpulan	55
5.2	Saran	56
DAFTAR PUSTAKA		57
LAMPIRAN		60

DAFTAR GAMBAR

Gambar 1.1 Dataset <i>Scraping</i> Instagram	4
Gambar 2.1 Logo Instagram	15
Gambar 2.2 <i>Google Colaboratory</i>	221
Gambar 3.1 Kerangka Penelitian	24
Gambar 3.2 Pembuatan Model <i>Machine Learning</i>	25
Gambar 3.3 <i>Text Preprocessing</i>	27
Gambar 4.1 Ekstensi Instagram <i>Scraper Apify</i>	33
Gambar 4.2 Dataset Instagram <i>Scraper</i>	33
Gambar 4.3 <i>Cleansing</i> Data	34
Gambar 4.4 Kode Program Pada Tahapan <i>Case Folding</i>	35
Gambar 4.5 Hasil Tahapan <i>Case Folding</i>	35
Gambar 4.6 Kode Program Pada Tahapan <i>Tokenizing</i>	36
Gambar 4.7 Hasil Tahapan <i>Tokenizing</i>	36
Gambar 4.8 Kode Program Pada Tahapan <i>Removing Stopword</i>	37
Gambar 4.9 Hasil Tahapan <i>Removing Stopword</i>	37
Gambar 4.10 Kode Program Pada Tahapan <i>Stemming</i>	38
Gambar 4.11 Hasil Tahapan <i>Stemming</i>	38
Gambar 4.12 Kode Program Pada Tahapan <i>Lexicon</i>	39
Gambar 4.13 Kode Program Pada Tahapan <i>Lexicon</i>	39
Gambar 4.14 Hasil Tahapan <i>Lexicon</i>	39
Gambar 4.15 Kode Program Pada Tahapan Memberikan Label Pada Nilai <i>Sentiment</i>	40
Gambar 4.16 Hasil Tahapan Memberikan Label Pada Nilai <i>Sentiment</i>	41
Gambar 4.17 Hasil Tahapan <i>Wordcloud</i> Seluruh Komentar	42
Gambar 4.18 Hasil Tahapan <i>Wordcloud</i> Komentar Positif	43
Gambar 4.19 Hasil Tahapan <i>Wordcloud</i> Komentar Negatif	44
Gambar 4.20 Hasil Tahapan <i>Wordcloud</i> Komentar Netral	45
Gambar 4.21 Hasil Tahapan <i>Vektorisasi</i>	45
Gambar 4.22 Kode Pemrograman <i>Classification Report</i> dan <i>Confusion Matrix</i>	47

Gambar 4.23 Hasil Tahapan <i>Classification Report</i> dan <i>Accuracy Score Naïve Bayes</i>	47
Gambar 4.24 Hasil Tahapan <i>Confusion Matrix</i>	48
Gambar 4.25 Kode Pemrograman <i>Classification Report</i> dan <i>Confusion Matrix</i>	49
Gambar 4.26 Hasil Tahapan <i>Classification Report</i> dan <i>Accuracy Score K-Nearest Neighbor</i>	50
Gambar 4.27 Hasil Tahapan <i>Confusion Matrix</i>	50
Gambar 4.28 Infografis Analisis Sentimen Akun Instagram Kemenkominfo	53



DAFTAR TABEL

Tabel 2.1 <i>State Of The Art</i>	9
Tabel 2.2 <i>Confusion Matrix</i>	17
Tabel 2.3 Simbol-Simbol <i>Flowchart</i>	17
Tabel 4.1 Perbandingan Data Latih dan Data Uji	49



DAFTAR RUMUS

Rumus 2.1 <i>Naïve Bayes</i>	13
Rumus 2.2 <i>Confusion Matrix</i>	17
Rumus 2.3 <i>Slovin</i>	17
Rumus 2.4 <i>K-Nearest Neighbor</i>	17
Rumus 3.1 Perhitungan <i>Slovin</i> Pada Sampal Data.....	26
Rumus 4.1 Perhitungan <i>Confusion Matrix</i> Akurasi Pada Klasifikasi <i>Naïve Bayes</i>	49
Rumus 4.2 Perhitungan <i>Confusion Matrix</i> Presisi Pada Klasifikasi <i>Naïve Bayes</i>	49
Rumus 4.3 Perhitungan <i>Confusion Matrix Recall</i> Pada Klasifikasi <i>Naïve Bayes</i>	49
Rumus 4.4 Perhitungan <i>Confusion Matrix</i> Akurasi Pada Klasifikasi <i>K-Nearest Neighbor</i>	51
Rumus 4.5 Perhitungan <i>Confusion Matrix</i> Presisi Pada Klasifikasi <i>K-Nearest Neighbor</i>	51
Rumus 4.6 Perhitungan <i>Confusion Matrix Recall</i> Pada Klasifikasi <i>K-Nearest Neighbor</i>	51

DAFTAR LAMPIRAN

1. Lampiran Plagiarisme
2. Lampiran Kartu Bimbingan Tugas Akhir
3. Lampiran Surat Rekomendasi
4. Biodata Mahasiswa

