



LAPORAN SKRIPSI

**ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK
MENGUNAKAN METODE K-NEAREST NEIGHBOR
(KNN)**

**FAIZA RIZQI IRAWAN
NIM. 201851234**

**DOSEN PEMBIMBING
Ahmad Jazuli, S.Kom., M.Kom
Tutik Khotimah, S.Kom., M.Kom**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MURIA KUDUS
APRIL 2022**

HALAMAN PERSETUJUAN

ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGUNAKAN METODE K-NEAREST NEIGHBOR (KNN)

FAIZA RIZQI IRAWAN

NIM. 201851234

Kudus, 21 januari 2022

Menyetujui,

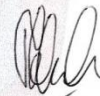
Pembimbing Utama,

Pembimbing Pendamping,



Ahmad Jazuli, S.Kom., M.Kom.

NIDN. 0406107004



Tutik Khotimah, S.Kom., M.Kom.

NIDN. 0608068502

HALAMAN PENGESAHAN

**ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK
MENGUNAKAN METODE K-NEAREST NEIGHBOR (KNN)**

FAIZA RIZQI IRAWAN

NIM. 201851234

Kudus, 29 April 2022

Menyetujui,


Ketua Penguji,


Rina Fiati, S.T., M.Cs.
NIDN. 0604047401

Anggota Penguji I,


Alif Catur Murti S.Kom., M.Kom.
NIDN. 0610129001

Anggota Penguji II,



Ahmad Jazuli, S.Kom., M.Kom.
NIDN. 0406107004

Mengetahui

Dekan Fakultas Teknik

Mohammad Danlan, S.T., M.T.
NIP. 0610701000001141

Ketua Program Studi Teknik Informatika


Mukhamad Nurkamid, S.Kom., M.Cs.
NIS. 0610701000001212

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini :

Nama : Faiza Rizqi Irawan
NIM : 201851234
Tempat & Tanggal Lahir : Pati, 21 Januari 2000
Judul Skripsi/Tugas Akhir* : Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbor (KNN)

Menyatakan dengan sebenarnya bahwa penulisan Tugas Akhir ini berdasarkan hasil penelitian, pemikiran dan pemaparan asli dari saya sendiri, baik untuk naskah laporan maupun kegiatan lain yang tercantum sebagai bagian dari Skripsi ini. Seluruh ide, pendapat, atau materi dari sumber lain telah dikutip dalam Skripsi dengan cara penulisan referensi yang sesuai.

Demikian pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar dan sanksi lain sesuai dengan peraturan yang berlaku di Universitas Muria Kudus.

Demikian pernyataan ini saya buat dalam keadaan sadar tanpa paksaan dari pihak manapun.

Kudus, 21 Januari 2022

Yang memberi pernyataan,



Faiza Rizqi Irawan
NIM. 201851234

KATA PENGANTAR

Puji syukur kehadiran Allah atas rahmat, hidayah, dan rahmatnya sehingga penulis dapat menyelesaikan tugas akhir ini dengan sukses dan lancar. Dalam proses penyelesaian tugas akhir ini, banyak narasumber yang telah memberikan dukungan baik bimbingan maupun semangat. Atas banyaknya bantuan, penulis ingin menyampaikan doa dan ucapan terima kasih tulus kepada:

1. Allah SWT yang telah memberikan segala rahmat dan Hidayahnya.
2. Bapak Prof. Dr. Ir. Darsono, M.Sig selaku Rektor Universitas Muria Kudus.
3. Bapak Mohammad Dahlan, ST., MT selaku Dekan Fakultas Teknik Universitas Muria Kudus.
4. Bapak Mukhamad Nurkamid, S.Kom, M.Cs sebagai Ketua Program Studi Teknik Informatika Universitas Muria Kudus.
5. Ibu Ratih Nindyasari, M.Kom selaku Koordinator Skripsi Program Studi Teknik Informatika Universitas Muria Kudus.
6. Bapak Ahmad Jazuli, S.Kom., M.Kom selaku dosen pembimbing utama, dan Ibu Tutik Khotimah, S.Kom., M.Kom., selaku pembimbing pendamping yang telah memberikan bimbingan, ilmu, saran, motivasi, nasihat, pengarahan, dan bimbingan selama penyusunan skripsi ini.
7. Keluarga yang senantiasa memberikan segala dukungan kepada penulis dalam menyelesaikan Skripsi ini serta do'a yang selalu menyertai penulis.

Penulis mengakui kekurangan serta kesalahan dalam penyusunan tugas akhir ini dan sekaligus menerima kritik, saran dan kontribusi pembaca untuk perbaikan di masa yang akan datang. Akhir kata, penulis berharap semoga buku-buku dalam tugas akhir ini dapat bermanfaat bagi penulis dan pembaca pada umumnya.

Kudus, 21 Januari 2022

Penulis

ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGUNAKAN METODE K-NEAREST NEIGHBOR (KNN)

Nama mahasiswa : Faiza Rizqi Irawan

NIM : 201851234

Pembimbing :

1. Ahmad Jazuli, S.Kom., M.Kom
2. Tutik Khotimah, S.Kom., M.Kom

RINGKASAN

Zaman modern menjadikan internet sebagai hal wajar, masyarakat sering menggunakan media sosial sebagai *microblogging*. Masyarakat menggunakannya untuk mengekspresikan diri dan mengemukakan pendapat. Salah satu media sosial yang digunakan yaitu twitter. Informasi yang dapat dibagikan pada twitter dinamakan *tweet*. *Tweet* memiliki batas yang dulunya 140 karakter menjadi 280 karakter (Haqqizar dan Larasyanti, 2019). *Tweet* dapat berupa teks, foto, video, dan sebagainya, melalui *tweet* pengguna dapat berinteraksi satu sama lain. Mereka dapat membagikan informasi, memberikan pendapat, menjelaskan suatu kejadian dan sebagainya.

Gojek merupakan *startup* yang bergerak dibidang *ride hailing*. Perusahaan Gojek sudah berdiri dari tahun 2010 oleh Nadiem Makarim, hampir seluruh kota di Indonesia sudah terjangkau Gojek. Setiap tahunnya dibutuhkan evaluasi dan juga penilaian terhadap layanan Gojek untuk mengetahui respon pengguna Gojek. Penelitian ini membahas klasifikasi sentimen positif, negatif, atau netral terhadap respon pengguna Gojek di media sosial twitter dengan menerapkan metode *K-Nearest Neighbor* (KNN). Hasil klasifikasi menggunakan metode *K-Nearest Neighbor* (KNN) mampu mengklasifikasikan respon dari pengguna twitter dan dapat digunakan perusahaan Gojek sebagai bahan evaluasi dan penilaian terhadap layanan Gojek. Hasil pengujian metode *K-Nearest Neighbor* (KNN) menggunakan *confusion matrix* dengan data sebanyak 1409 mendapatkan tingkat akurasi sebesar 79,43% dengan nilai $k=15$.

Kata Kunci : Analisis Sentimen, *confusion matrix*, Gojek, Klasifikasi, *K-Nearest Neighbor*, KNN, Twitter.

SENTIMENT ANALYSIS OF GOJEK USERS USING K-NEAREST NEIGHBOR (KNN) METHOD

Student Name : Faiza Rizqi Irawan

Student Identity Number : 201851234

Supervisor :

1. Ahmad Jazuli, S.Kom., M.Kom
2. Tutik Khotimah, S.Kom., M.Kom

ABSTRACT

Modern times make the internet a natural thing, people often use social media as microblogging. People use it to express themselves and express opinions. One of the social media used is Twitter. Information that can be shared on Twitter is called a tweet. Tweets have a limit from 140 characters to 280 characters (Haqqizar and Larasyanti, 2019). Tweets can be text, photos, videos, and so on, through tweets users can interact with each other. They can share information, give opinions, explain an incident and so on.

Gojek is a startup engaged in ride-hailing. The Gojek company has been established since 2010 by Nadiem Makarim, almost all cities in Indonesia have been reached by Gojek. Every year an evaluation and assessment of Gojek's services is needed to find out the response of Gojek users. This study discusses the classification of positive, negative, or neutral sentiments on the response of Gojek users on Twitter social media by applying the K-Nearest Neighbor (KNN) method. The results of the classification using the K-Nearest Neighbor (KNN) method are able to classify responses from twitter users and can be used by the Gojek company as an evaluation and assessment material for Gojek services. The results of testing the K-Nearest Neighbor (KNN) method using a confusion matrix with 1409 data get an accuracy rate of 79.43% with a value of $k=15$.

Keywords: Sentiment Analysis, confusion matrix, Gojek, Classification, K-Nearest Neighbor, KNN, Twitter.

DAFTAR ISI

HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN	iv
KATA PENGANTAR	iv
RINGKASAN	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR SIMBOL	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR ISTILAH DAN SINGKATAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan	2
1.5 Sistematika penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terkait	5
2.2 Landasan Teori	7
2.2.1 Analisis Sentimen	7
2.2.2 Confusion matrix	8
2.2.3 Gojek	10
2.2.4 Metode K-Nearest Neighbor	10
2.2.5 Python	11
2.2.6 Text Mining	11
2.2.7 Twitter	12
2.2.8 Twitter API	12
2.2.9 Pre-processing	13

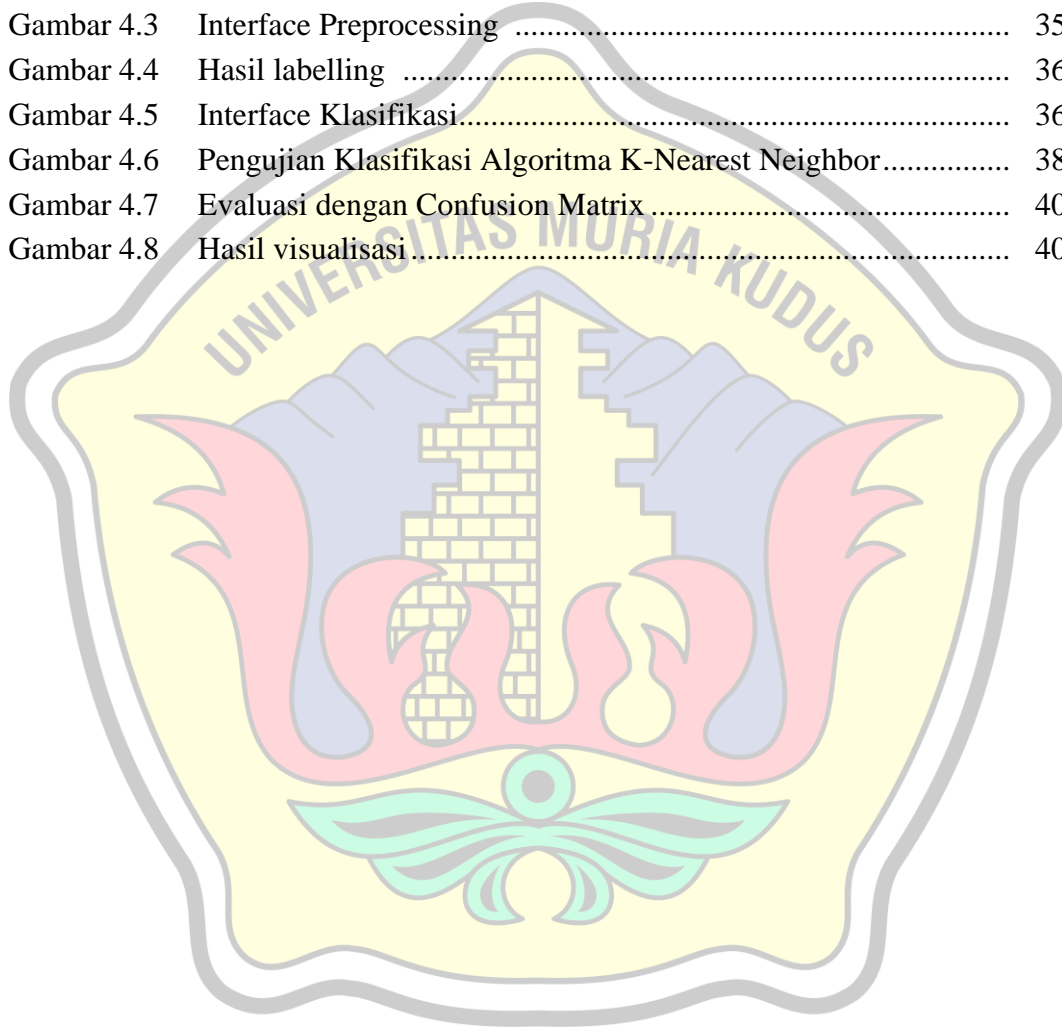
2.2.9.1 Cleaning	13
2.2.9.2 Case Folding	13
2.2.9.3 Tokenizing.....	13
2.2.9.4 Stopword	13
2.2.9.5 Stemming	13
2.2.10 Pembobotan Kata.....	13
BAB III METODOLOGI	15
3.1 Gambaran Penelitian	15
3.2 Studi Literatur	16
3.3 Metode Pengumpulan Data	16
3.4 Analisa Kebutuhan Sistem	16
3.5 Skenario Sistem.....	16
3.5.1 Crawling.....	16
3.5.2 Pre-processing.....	17
3.5.3 Klasifikasi K-Nearest Neighbor.....	18
3.6 Tahap Evaluasi Uji Perhitungan.....	19
BAB IV HASIL DAN PEMBAHASAN	21
4.1. Pengumpulan Dataset.....	21
4.2. Tahap Preprocessing Data.....	24
4.2.1. Hasil <i>Cleaning</i>	24
4.2.2. Hasil <i>Case Folding</i>	26
4.2.3. Hasil <i>Tokenizing</i>	28
4.2.4. Hasil <i>Stopword</i>	30
4.2.5. Hasil <i>Stemming</i>	32
4.2.6. Hasil Preprocessing.....	34
4.3. Pembobotan Kata	35
4.4. Klasifikasi Algoritma <i>K-Nearest Neighbor</i>	36
4.5. Evaluasi dengan Confusion Matrix	38
4.6. Visualisasi	40
5.1. Kesimpulan.....	41
5.2. Saran.....	42
DAFTAR PUSTAKA	43
LAMPIRAN 1.....	47

LAMPIRAN 2	53
LAMPIRAN 3	57
BIODATA PENULIS	64



DAFTAR GAMBAR

Gambar 3.1	Gambaran Penelitian	15
Gambar 3.2	Diagram Alir <i>Crawling</i>	17
Gambar 3.3	Diagram Alir Preprocessing	17
Gambar 3.4	Diagram Alir Klasifikasi	18
Gambar 4.1	API Twitter	21
Gambar 4.2.	Crawling Data	24
Gambar 4.3	Interface Preprocessing	35
Gambar 4.4	Hasil labelling	36
Gambar 4.5	Interface Klasifikasi.....	36
Gambar 4.6	Pengujian Klasifikasi Algoritma K-Nearest Neighbor.....	38
Gambar 4.7	Evaluasi dengan Confusion Matrix	40
Gambar 4.8	Hasil visualisasi	40



DAFTAR TABEL

Tabel 2.1	Analisis Sentimen	8
Tabel 2.2	Confusion matrix	8
Tabel 4.1	Hasil Data <i>Crawling</i>	22
Tabel 4.2	Hasil <i>Cleaning</i>	24
Tabel 4.3	Hasil <i>Case Folding</i>	26
Tabel 4.4.	Hasil <i>Tokenizing</i>	28
Tabel 4.5	Hasil <i>Stopword</i>	30
Tabel 4.6	Hasil <i>Stemming</i>	32
Tabel 4.7	Hasil Klasifikasi <i>K-Nearest Neighbor</i>	37



DAFTAR SIMBOL

Simbol	Keterangan	Nomor Persamaan
d	Jumlah dokumen	2.6
d_i	Jarak euclidean ke-i	2.5
df	Banyak dokumen yang mengandung kata (term)	2.6
IDF	Kemunculan kata pada kumpulan dokumen	2.6
I	Baris ke-i	2.5
n	Banyak data training	2.5
tf	Kemunculan kata pada dokumen	2.6
X_{2i}	Jarak training ke-i	2.5
X_{1i}	Jarak testing ke-i	2.5

DAFTAR LAMPIRAN

Lampiran 1	47
Lampiran 2	53
Lampiran 3	57



DAFTAR ISTILAH DAN SINGKATAN

TP	: <i>True Positive</i>
FP	: <i>Flase Positive</i>
FN	: <i>False negative</i>
KNN	: <i>K-Nearest Neighbor</i>
NLP	: <i>Natural Language Processing</i>
NLTK	: <i>Natural Language Toolkit</i>
TN	: <i>True negatif</i>

