



LAPORAN SKRIPSI

**ANALISIS SENTIMEN TERHADAP KARTU PRAKERJA
DI TENGAH PANDEMI COVID-19 PADA *TWITTER*
MENGUNAKAN ALGORITMA *K-NEAREST
NEIGHBOR* (KNN)**

MUHAMMAD ALFIAN

NIM. 201851040

DOSEN PEMBIMBING

**Esti Wijayanti, S.Kom., M.Kom.
Ahmad Abdul Chamid, S.Kom., M.Kom.**

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

2022

HALAMAN PERSETUJUAN

ANALISIS SENTIMEN TERHADAP KARTU PRAKERJA DI TENGAH PANDEMI COVID-19 PADA *TWITTER* MENGUNAKAN ALGORITMA *K-NEA REST NEIGHBOR* (KNN)



MUHAMMAD ALFIAN

NIM. 201851040

Kudus, 13 Juli 2022

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,

A handwritten signature in black ink, appearing to be 'Esti' followed by a stylized flourish.

Esti Wijayanti, S.Kom., M.Kom.

NIDN. 0605098901

A handwritten signature in black ink, appearing to be 'Ahmad' followed by a stylized flourish.

Ahmad Abdul Chamid, S.Kom., M.Kom.

NIDN. 0616109101

HALAMAN PENGESAHAN

ANALISIS SENTIMEN TERHADAP KARTU PRAKERJA DI TENGAH PANDEMI COVID-19 PADA *TWITTER* MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR* (KNN)

MUHAMMAD ALFIAN
NIM. 201851040

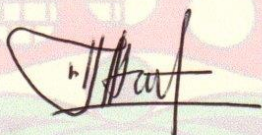
Kudus, 16 Agustus 2022

Menyetujui,

Ketua Penguji,

Anggota Penguji I,

Anggota Penguji II,



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

Ratih Nindyasari, S.Kom., M.Kom.
NIDN. 0625028501

Esti Wijayanti, S.Kom., M.Kom.
NIDN. 0605098901


Mengetahui

Dekan Fakultas Teknik

Ketua Program Studi Teknik Informatika



Mohammad Dahlan, S.T., M.T.
NIS. 0610701000001141



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

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Alfian
NIM : 201851040
Tempat & Tanggal Lahir : Kudus, 28 Agustus 2000
Judul Skripsi/Tugas Akhir* : Analisis sentimen terhadap kartu prakerja di tengah pandemi covid-19 pada twitter menggunakan algoritma k-nearest neighbor (KNN).

Menyatakan dengan sebenarnya bahwa penulisan Skripsi/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, 13 juli 2022

Yang memberi pernyataan,



Muhammad Alfian
NIM. 201851040

KATA PENGANTAR

Puji syukur kehadiran Allah SWT, karena berkat Rahmat serta Hidayah - Nya, penulis berhasil menyelesaikan Skripsi berjudul “Analisis sentimen terhadap kartu prakerja di tengah pandemi covid-19 pada *twitter* menggunakan algoritma *k-nearest neighbor* (KNN)”.

Penyusunan Skripsi/Tugas Akhir ini ditujukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer (S.Kom.). Pelaksanaan Skripsi ini tak lepas dari bantuan dan dukungan beberapa pihak, untuk itu penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Darsono, M.Si selaku Rektor Universitas Muria Kudus
2. Bapak Mohammad Dahlan, ST., MT selaku Dekan Fakultas Teknik Universitas Muria Kudus.
3. Bapak Mukhamad Nurkamid S.Kom., M.Cs. selaku Ketua Program Studi Teknik Informatika Universitas Muria Kudus.
4. Ibu Ratih Nindyasari, S.Kom., M. Kom selaku Koordinator Skripsi Teknik Informatika Universitas Muria Kudus.
5. Ibu Esti Wijayanti S.Kom., M.Kom. selaku dosen pembimbing I yang telah banyak memberi masukan selama penyusunan Skripsi ini.
6. Bapak Ahmad Abdul Chamid S.Kom., M.Kom. selaku dosen pembimbing II yang telah banyak memberi masukan selama penyusunan Skripsi ini.
7. Kedua orangtua dan keluarga besar yang senantiasa memberikan dukungan baik secara moril maupun materil.

Penulis menyadari adanya kekurangan dan ketidaksempurnaan dalam penulisan Skripsi ini, karena itu penulis menerima kritik, saran dan masukan dari pembaca sehingga penulis dapat lebih baik di masa yang akan datang. Akhirnya penulis berharap semoga buku tesis ini bisa bermanfaat khususnya bagi penulis dan umumnya bagi para pembaca.

Kudus, 13 Juli 2022

Penulis

**ANALISIS SENTIMEN TERHADAP KARTU PRAKERJA
DI TENGAH PANDEMI COVID-19 PADA *TWITTER* MENGGUNAKAN
ALGORITMA *K-NEAREST NEIGHBOR* (KNN)**

Nama mahasiswa : Muhammad Alfian

NIM : 201851040

Pembimbing :

1. Esti Wijayanti, S.Kom., M.Kom
2. Ahmad Abdul Chamid, S.Kom., M.Kom

RINGKASAN

Kartu prakerja di tujukan untuk masyarakat yang sedang mencari pekerjaan seperti buruh, karyawan, dan pegawai, namun peluncuran program ini berbarengan dengan indonesia yang sedang menghadapi wabah COVID-19. Banyak pro dan kontra terkait program yang diluncurkan pada masa pandemi ini terlebih lagi pelatihan yang awalnya direncanakan offline, di tengah masa pandemic ini dengan terpaksa dilakukan secara online. Hal tersebut mengakibatkan banyak masyarakat yang membahas dan mendebatkan program kartu prakerja ini, salah satunya pada media sosial twitter. Analisis sentimen atau *sentiment analysis* merupakan proses yang bertujuan untuk mengidentifikasi dataset berbentuk teks untuk mendapatkan informasi sentimen yang bersifat positif, negatif, atau netral. Penelitian ini bertujuan untuk melakukan analisis sentimen dengan mengidentifikasi tanggapan masyarakat terhadap program kartu prakerja pada twitter dan mengklasifikasi menggunakan metode *K-nearest neighbor* (KNN).

Kata kunci : Kartu Prakerja, Twitter, *sentiment analysis*, *K-nearest neighbor*.

**SENTIMENT ANALYSIS OF PRACTICE CARD
IN THE MIDDLE OF THE COVID-19 PANDEMIC ON TWITTER USING
THE K-NEAREST NEIGHBOR (KNN) ALGORITHM**

Student Name : Muhammad Alfian

Student Identity Number : 201851040

Supervisor :

1. Esti Wijayanti, S.Kom., M.Kom
2. Ahmad Abdul Chamid, S.Kom., M.Kom

ABSTRACT

The pre-employment card is intended for people who are looking for work such as laborers, employees, and employees, but the launch of this program coincides with Indonesia which is facing the COVID-19 outbreak. There are many pros and cons related to programs launched during this pandemic, especially training that was originally planned to be offline, in the midst of this pandemic, it was forced to be done online. This has resulted in many people discussing and debating this pre-employment card program, one of which is on social media twitter. Sentiment analysis or sentiment analysis is a process that aims to identify datasets in the form of text to obtain positive, negative, or neutral sentiment information. This study aims to analyze sentiment by identifying public responses to the pre-employment card program on Twitter and classifying it using the K-nearest neighbor (KNN) method.

Keywords : Pre-Employment Card, Twitter, sentiment analysis, K-nearest neighbor.

DAFTAR ISI

HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN.....	iv
KATA PENGANTAR.....	v
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.....	3
1.5. Sistematika penulisan	3
BAB II TINJAUAN PUSTAKA.....	5
2.1. Penelitian Terkait	5
2.2. Landasan Teori	6
2.2.1 Program Kartu Prakerja.....	6
2.2.2 Media Sosial	7
2.2.3 <i>Pre-processing</i>	7
2.2.4 Pembobotan Kata.....	8
2.2.5 Analisis Sentimen	9
2.2.6 <i>Text Mining</i>	9
2.2.7 <i>Twitter</i>	9
2.2.8 <i>Scraping</i>	10
2.2.9 Pembobotan TF-IDF.....	10
2.2.10 <i>Klasifikasi K-Nearest Neighbor</i>	11
2.2.11 Evaluasi	12

BAB III METODOLOGI.....	15
3.1. Metode Penelitian.....	15
3.2. Studi Literatur	15
3.2.1 <i>Python</i>	15
3.2.2 NLTK	15
3.2.3 <i>Scikit-Learn</i>	15
3.2.4 <i>Flask</i>	16
3.3. Metode Pengumpulan Data	16
3.4. Skenario Sistem.....	16
3.4.1 Gambaran Umum	16
3.4.2 Tahapan <i>Scraping</i>	17
3.4.3 Tahapan <i>Preprocessing</i>	18
3.4.4 Tahapan <i>Labeling</i>	24
3.4.5 Tahapan Klasifikasi <i>K-nearest neighbor</i> (KNN).....	25
3.4.6 Tahapan Evaluasi.....	26
3.4.7 Tahapan Akurasi.....	26
3.5. Analisis Kebutuhan Sistem	26
3.6. Kerangka Pemikiran	27
BAB IV HASIL DAN PEMBAHASAN	29
4.1. Implementasi	29
4.1.1 Tahap Pengumpulan <i>Dataset</i>	29
4.1.2 <i>Preprocessing</i>	30
4.1.3 <i>Labeling</i> Menggunakan <i>TextBlood</i>	33
4.1.4 Klasifikasi <i>K-Nearest Neighbor</i> (KNN).....	34
4.1.5 Pengujian	35
4.2. Hasil dan Pembahasan.....	36
4.2.1 Hasil <i>Scraping</i>	36
4.2.2 Hasil <i>Preprocessing</i>	39
4.2.3 Hasil <i>Labeling</i>	42
4.2.4 Hasil Klasifikasi <i>K-Nearest Neighbor</i> (KNN).....	43
4.2.5 Perhitungan manual <i>K-Nearest Neighbor</i> (KNN)	45
4.2.6 Perhitungan <i>Confusion Matrix</i>	50
4.2.7 Visualisasi.....	52
BAB V PENUTUP.....	53
5.1. Kesimpulan.....	53
5.2. Saran.....	53

DAFTAR PUSTAKA	54
LAMPIRAN 1 Lembar Konsultasi	56
LAMPIRAN 2 Lembar Revisi	60
BIODATA PENULIS	63



DAFTAR GAMBAR

Gambar 2. 1 <i>confusion matrix</i>	13
Gambar 3. 1 Gambaran umum sistem.....	17
Gambar 3. 2 Alur proses <i>scraping</i>	17
Gambar 3. 3 Alur proses <i>Preprocessing</i>	18
Gambar 3. 4 Alur proses <i>Labeling</i>	25
Gambar 3. 5 Alur klasifikasi <i>K-nearest neighbor</i> (KNN).....	25
Gambar 3. 6 Skema Kerangka Pemikiran	27
Gambar 4. 1 <i>Twitter Developer</i>	29
Gambar 4. 2 <i>key</i> dan token <i>API Twitter</i>	30
Gambar 4. 3 Fungsi <i>scraping</i>	30
Gambar 4. 4 Fungsi <i>cleansing</i> pada <i>tweet</i>	31
Gambar 4. 5 Fungsi <i>casefold</i>	32
Gambar 4. 6 Fungsi <i>Tokenizing</i>	32
Gambar 4. 7 Fungsi <i>stopword</i>	32
Gambar 4. 8 Fungsi <i>Stemming</i>	33
Gambar 4. 9 Fungsi <i>Labeling</i>	34
Gambar 4. 10 Proses <i>Splitting Data</i>	34
Gambar 4. 11 Proses <i>vektorisasi</i>	35
Gambar 4. 12 Proses <i>TF-IDF</i>	35
Gambar 4. 13 Proses <i>Klasifikasi K-Nearest Neighbor</i> (KNN)	35
Gambar 4. 14 <i>Source code confusion matrix</i>	36
Gambar 4. 15 <i>Interface</i> proses <i>scraping</i>	36
Gambar 4. 16 <i>Interface</i> hasil <i>scraping</i>	37
Gambar 4. 17 <i>Interface</i> Hasil <i>preprocessing</i>	40
Gambar 4. 18 <i>Interface</i> Hasil <i>Labeling</i>	42
Gambar 4. 19 <i>Interface</i> halaman klasifikasi.....	44
Gambar 4. 20 <i>Interface</i> hasil halaman klasifikasi	44
Gambar 4. 21 <i>Interface</i> hasil halaman visualisasi.....	52

DAFTAR TABEL

Tabel 3. 1 Contoh <i>Cleaning</i>	18
Tabel 3. 2 Contoh <i>Case Folding</i>	20s
Tabel 3. 3 Contoh <i>Tokenizing</i> pada <i>Tweet</i>	21
Tabel 3. 4 Contoh <i>Stop Removal</i> pada <i>Tweet</i>	22
Tabel 3. 5 Contoh <i>Stemming</i> pada <i>Tweet</i>	23
Tabel 4. 1 Contoh <i>Sraping</i> pada <i>Tweet</i>	37
Tabel 4. 2 Contoh <i>Preprocessing</i> pada <i>Tweet</i>	40
Tabel 4. 3 Contoh <i>labeling</i> pada <i>Tweet</i>	42
Tabel 4. 4 Contoh inputan text.....	45
Tabel 4. 5 Contoh menghitung <i>document frequency</i> (df)	45
Tabel 4. 6 Contoh menghitung invers <i>document frequency</i> (idf).....	46
Tabel 4. 7 Contoh menghitung bobot pada masing masing dokumen	47
Tabel 4. 8 Contoh pembobotan kata dengan perhitungan tf idf.....	47
Tabel 4. 9 Klasifikasi <i>K-Nearest Neighbor</i> (KNN)	49

DAFTAR SIMBOL

Simbol	Keterangan	Nomor Persamaan
TF	Kehadiran kata di dokumen	2.1
IDF	Kehadiran kata di kumpulan dokumen	2.1
d	jumlah dokumen	2.1
df	banyak dokumen mengandung kata(term)	2.1
D	Jarak antara dua titik x dan y	2.2
X	Data uji	2.2
Y	Sampel data	2.2
n	Dimensi data	2.2

DAFTAR LAMPIRAN

Lampiran 1	Lembar Konsultasi.....	56
Lampiran 2	Lembar Revisi	60



DAFTAR ISTILAH DAN SINGKATAN

- KNN : K-Nearest Neighbor
NLTK : NaturalaLanguage Toolkit

