

**SKRIPSI**  
**ANALISIS SENTIMEN MASYARAKAT TERHADAP SYARAT**  
**LULUS PROGRAM SARJANA TIDAK WAJIB SKRIPSI PADA**  
***TWITTER* DENGAN METODE *K-NEAREST NEIGHBOR***



**NURIMA MASBAIT**

**NIM : 205610029**

**PROGRAM STUDI SISTEM INFORMASI**  
**PROGRAM SARJANA FAKULTAS TEKNOLOGI INFOMASI**  
**UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA**  
**YOGYAKARTA**

**2024**

**SKRIPSI**  
**ANALISIS SENTIMEN MASYARAKAT TERHADAP SYARAT LULUS**  
**PROGRAM SARJANA TIDAK WAJIB SKRIPSI PADA *TWITTER***  
**DENGAN METODE *K-NEAREST NEIGHBOR***

**Diajukan sebagai salah satu syarat untuk menyelesaikan studi**

**Program Sarjana**

**Program Studi Sistem Informasi**

**Fakultas Teknologi Informasi**

**Universitas Teknologi Digital Indonesia**

**Yogyakarta**

**Disusun Oleh**

**NURIMA MASBAIT**

**NIM : 205610029**

**PROGRAM STUDI SISTEM INFORMASI**  
**PROGRAM SARJANA**  
**FAKULTAS TEKNOLOGI INFORMASI**  
**UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA**  
**YOGYAKARTA**

**2024**

## HALAMAN PERSETUJUAN

### UJIAN SKRIPSI/TUGAS AKHIR

**Judul** : Analisis Sentimen Masyarakat Terhadap Syarat Lulus Program Sarjana Tidak Wajib Skripsi Pada *Twitter* Dengan Metode *K-Nearest Neighbor*

**Nama** : Nurima Masbait

**NIM** : 205610029

**Program Studi** : Sistem Informasi

**Program** : Sarjana

**Semester** : Gasal

**Tahun Akademik** : 2023/2024

Telah diperiksa dan disetujui untuk dijadikan di hadapan Dewan Penguji Skripsi

Yogyakarta, 05 Februari 2024

Dosen Pembimbing



Ir. Totok Suprawoto, MM, MT

NIDN : 0514125801

## HALAMAN PENGESAHAN

### SKRIPSI

# ANALISIS SENTIMEN MASYARAKAT TERHADAP SYARAT LULUS PROGRAM SARJANA TIDAK WAJIB SKRIPSI PADA *TWITTER DENGAN METODE K-NEAREST NEIGHBOR*

Telah dipertahankan di depan Dewan Penguji Skripsi dan dinyatakan diterima untuk memenuhi sebagian persyaratan guna memperoleh Gelar Sarjana Komputer



Program Studi Sistem Informasi

Fakultas Teknologi Informasi

Universitas Teknologi Digital Indonesia

Yogyakarta

Yogyakarta, ..... 05 Februari ..... 2024

	Dewan Penguji	NIDN	Tandatangan
1	Deborah Kurniawati, S.Kom., M.Cs. (Ketua)	0511107301	
2	Ir. Totok Suprawoto, MM, MT (Sekertaris)	0514125801	

Mengetahui  
Ketua Program Studi Sistem Infomasi

  
Deborah Kurniawati, S.Kom., M.Cs.

NPP: 051149

05 FEB 2024

## **PERNYATAAN KEASLIAN SKRIPSI**

Dengan ini saya menyatakan menyatakan bahwa naska skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naska ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 05 Februari ..... 2024



Nurima Masbait  
NIM: 205610029

## HALAMAN PERSEMBAHAN

Puji syukur Alhamdulillah saya panjatkan kepada Allah SWT, yang telah memberikan karunia kepada saya sehingga dapat menyelesaikan skripsi ini yang saya persembahkan kepada:

1. Kedua orang tua saya, Bapak Sulaeman Masbait dan Ibu Santi Buton yang telah membesarkan saya dengan penuh kasih sayang dan cinta yang tidak mungkin dapat saya balas, dan selalu memberikan dukungan atas semua keputusan dan pilihan dalam hidup saya serta tidak pernah putus mendoakan saya.
2. Adik-adikku tercinta yang selalu memberikan dukungan agar cepat selesai skripsi.
3. Segenap keluarga besar saya yang telah memberikan nasehat dan bimbingan agar saya dapat menjadi lebih baik dan mandiri serta dapat membanggakan mereka.
4. Sahabatku Yaya Karepesina, Ramlah Lakolite dan Wahyu Biloro yang telah menjadi kawan terbaik selama ini dan selalu memberikan semangat dan nasehat serta selalu menemani saya dalam suka ataupun duka.
5. Dan tidak lupa juga untuk semua pembaca skripsi ini, semoga dengan tulisan saya ini dapat membantu mendapatkan referensi bermanfaat.

## **HALAMAN MOTTO**

“Kalau pendekar terjatuh dia tidak akan sedih, dia tidak akan kecewa, dia tidak akan menangis, dia akan berdiri lagi, di jatuhkan lagi, berdiri lagi” (Prabowo Subianto”)

“ Menyesali nasib tidak akan mengubah keadaan. Terus berkarya dan bekerjalah yang membuat kita berharga” (Abduraahman Wahid)

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT, karena berkat rahmat dan hidayah-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Masyarakat Terhadap Program Lulus Sarjana Tidak Wajib Skripsi Pada *Twitter* Dengan Metode K-Nearest Neighbor” dengan baik . Penulisan skripsi ini bertujuan memenuhi salah satu syarat dalam menyelesaikan program studi S1 pada Fakultas Teknologi Informasi di Universitas Teknologi Digital Indonesia.

Penulis menyadari sepenuhnya, telah banyak mendapatkan dukungan, bimbingan dan dorongan dari berbagai pihak yang telah menyumbangkan pikiran, waktu, tenaga dan sebagainya. Oleh karena itu, pada kesempatan ini dengan setulus hati penulis mengucapkan terima kasi kepada:

1. Ibu Sri Redjeki, S.Si., M.Kom., PhD, selaku Rektor Universitas Teknologi Digital Indonesia
2. Deborah Kurniawati, S.Kom., M.Cs, selaku Ketua Program Studi Sistem Informasi
3. Bapak Ir. Totok Suprawoto, M.M., M.T, selaku Dosen Pembimbing Skripsi yang telah banyak memberikan arahan bimbingan dan motivasi pada saat penulisan skripsi ini. Semoga kebaikan beliau menjadi ladang pahala Aamiin.
4. Bapak dan Ibu Dosen Sistem Informasi Universitas Teknologi Digital Indonesia Yogyakarta yang telah memberikan ilmu dari awal perkuliahan.
5. Kedua orang tua saya, Bapak Sulaeman Masbait dan Ibu Santi Buton yang telah memberikan semangat dan nasehat serta doa yang selalu menyertai sehingga dapat menyelesaikan skripsi ini.



6. Semua pihak yang tidak dapat disebutkan satu persatu yang telah banyak membantu dalam menyelesaikan skripsi ini secara langsung maupun tidak langsung. Terimakasih telah memberikan semangat dan motivasi. Terimakasih atas semuanya, terutama kebersamaan dan rasa kekeluargaan yang begitu indah selama ini.

Dalam penulisan skripsi ini, penulis sudah berusaha sebaik mungkin. Namun, masih banyak kesalahan dan kekurangan baik sengaja ataupun yang tidak disengaja. Penulis menerima kritik serta saran yang membangun sebagai perbaikan kedepannya dan semoga skripsi ini membawa manfaat bagi semua.

Yogyakarta, 05 Februari 2024



Penulis

## DAFTAR ISI

SKRIPSI.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN .....	v
HALAMAN MOTTO .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	2
1.3 Ruang Lingkup .....	2
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	6
2.1 Tinjauan Pustaka .....	6
2.2 Landasan Teori .....	9
2.2.1 Analisis Sentimen .....	9
2.2.2 <i>Tweet Harvest</i> (Twitter Crawler) .....	11
2.2.3 <i>Python</i> .....	11
2.2.4 <i>Text Preprocessing</i> .....	12
2.2.5 Pembobotan Kata .....	15
2.2.6 <i>K-Nearest Neighbor</i> .....	16
2.2.7 <i>Confusion Matrix</i> .....	17
BAB III METODE PENELITIAN.....	19
3.1 Bahan/Data .....	19

3.2	Peralatan Penelitian .....	20
3.3	Prosedur Kerja dan Pengumpulan Data.....	20
3.3.1	Analisis.....	21
3.3.2	Identifikasi Masalah .....	21
3.3.3	<i>Crawling Data Twitter</i> .....	21
3.3.4	Pre – Processing .....	23
3.3.5	Labeling.....	26
3.3.6	Pembobotan Kata .....	28
3.3.7	Implementasi <i>K-Nearest Neighbor</i> .....	28
3.3.8	<i>Confusion Matrix</i> .....	29
BAB IV IMPLEMENTASI DAN PEMBAHASAN .....		30
4.1	Implementasi dan Pembahasan .....	30
4.1.1	<i>Crawling Data Twitter</i> .....	30
4.1.2	Pre – Processing .....	34
4.1.3	Labeling.....	41
4.1.4	Pembobotan Kata .....	43
4.1.5	<i>K – Nearest Neighbor</i> .....	45
4.1.6	<i>Confusion Matrix</i> .....	48
4.1.7	Pengujian Evaluasi Label Manual.....	54
BAB V PENUTUP.....		56
5.1	Kesimpulan.....	56
5.2	Saran .....	57
DAFTAR PUSTAKA .....		58
LAMPIRAN.....		60

## DAFTAR GAMBAR

Gambar 2.1 Contoh case folding (Pramayasa et al., 2023, p. 93).....	12
Gambar 2.2 Contoh cleaning (Pramayasa et al., 2023, p. 93).....	13
Gambar 2.3 Contoh <i>tokenizing</i> (Pramayasa et al., 2023, p. 93).....	13
Gambar 2.4 Contoh stopword (Pramayasa et al., 2023, p. 93) .....	14
Gambar 2.5 Contoh stemming (Pramayasa et al., 2023, p. 93).....	14
Gambar 2.6 Tabel confusion matrix oleh Nugroho (Hikmawan et al., 2020) .....	17
Gambar 3.1 Alur penelitian.....	21
Gambar 3.2 Alur crawling data.....	22
Gambar 3.3 Alur pre-processing.....	23
Gambar 3.4 Tahap labeling.....	26
Gambar 4.1 Aut_token twitter.....	31
Gambar 4.2 Program aut_token twitter.....	31
Gambar 4.3 Install library python dan node.js.....	31
Gambar 4.4 Hasil research keyword dan taggal.....	32
Gambar 4.5 Program crawling dengan setting keyword dan tanggal .....	32
Gambar 4.6 Tampilan postingan yang di crawling pada komentar .....	33
Gambar 4.7 Program crawling kolom komentar pada sebuah postingan.....	33
Gambar 4.8 Hasil crawling tweet harvest .....	34
Gambar 4.9 Program case folding.....	34
Gambar 4.10 Hasil case folding .....	35
Gambar 4.11 Program <i>tokenizing</i> .....	36
Gambar 4.12 Hasil <i>tokenizing</i> .....	37
Gambar 4.13 Code program normalization.....	37
Gambar 4.14 Hasil normalization .....	38
Gambar 4.15 Code program stopwords .....	39
Gambar 4.16 Hasil stopwords removal .....	39
Gambar 4.17 Code program stemming .....	40
Gambar 4.18 Hasil stemming.....	41
Gambar 4.19 Code program hapus duplikat .....	41
Gambar 4.20 Code program pelabelan.....	42
Gambar 4.21 Hasil pelabelan .....	42
Gambar 4.22 Code program perhitungan jumlah polarity .....	42
Gambar 4.23 Hasil perhitungan polarity.....	43
Gambar 4.24 Visualisasi data label dalam plot.....	43
Gambar 4.25 Pembobotan TF-IDF dengan rapidminer .....	44
Gambar 4.26 Hasil TF-IDF rapidminer .....	44
Gambar 4.27 Frekuensi kemunculan kata.....	45
Gambar 4.28 Penerapan algoritma dengan split data.....	46
Gambar 4.29 Confusion matrix skema pertama.....	49
Gambar 4.30 Confusion matrix skema kedua.....	51
Gambar 4.31 Confusion matrix skema ketiga.....	52

## DAFTAR TABEL

Tabel 2.1 Tinjauan pustaka.....	8
Tabel 3.1 Contoh tweet .....	22
Tabel 3.2 Contoh mapping term.....	24
Tabel 3.3 Contoh hasil pre-processing.....	25
Tabel 3.4 Contoh kamus negatif dan positif .....	27
Tabel 3.5 Hasil pelabelan dengan kamus positif dan negatif.....	27
Tabel 3.6 Confusion matrix.....	29
Tabel 4.1 Skema pengujian data latih dan data uji.....	46
Tabel 4.2 Hasil skema pengujian pertama .....	47
Tabel 4.3 Hasil skema pengujian kedua.....	48
Tabel 4.4 Hasil skema pengujian ketiga .....	48
Tabel 4.5 Perhitungan presisi skema pertama.....	50
Tabel 4.6 Perhitungan recall skema pertama .....	50
Tabel 4.7 Perhitungan Presisi skema kedua.....	52
Tabel 4.8 Perhitungan recall skema kedua.....	52
Tabel 4.9 Perhitungan presisi skema ketiga.....	53
Tabel 4.10 Perhitungan recall skema ketiga .....	53
Tabel 4.12 Confusion matrix evaluasi manual data uji skema pertama.....	54

## INTISARI

Skripsi merupakan salah satu syarat kelulusan dalam memperoleh gelar sarjana. Namun dalam kegiatan Merdeka Belajar Episode 26 Transformasi Standar Nasional dan Akreditasi Pendidikan Tinggi yang dilaksanakan 29 Agustus 2023, menteri Pendidikan dan Kebudayaan Nadiem Anwar Makarim menyatakan bahwa skripsi tidak lagi menjadi syarat wajib lulus perguruan tinggi program sarjana, namun dapat diganti dalam bentuk lainnya seperti prototipe, proyek, atau bentuk sejenis lainnya. Dari topik tersebut tentu menjadi perhatian masyarakat yang tertuang pada sosial media. Dalam hal ini perlu dilakukan penelitian melihat seberapa besar masyarakat yang setuju dengan adanya program mahasiswa tidak wajib skripsi tersebut.

Tahapan yang dilakukan dalam penelitian ini yaitu crawling data dengan *tweet harvest*, *pre-processing* yang terdiri dari *case folding*, *tokenizing*, *normalization*, *stopword* dan *stemming*. Setelah itu dilakukan pelabelan dengan kamus lexicon bahasa indonesia yang ditentukan secara manual. Kemudian pembobotan kata, selanjutnya proses pengklasifikasian data dengan metode *K-Nearest Neighbor* dan *confusion matrix*.

Hasil penelitian ini menggunakan 1924 *tweet* yang telah melalui proses pelabelan terdapat 68% sentimen netral, 18% sentimen positif, dan 13% sentimen negatif. Dari hasil perhitungan 3 skema pengujian analisis sentimen metode *K-Nearest Neighbors* dengan menggunakan *tools rapidminer* mendapatkan hasil terbaik menggunakan data latih 90% sebanyak 1732 data dan data uji 10% sebanyak 192 dengan  $k=5$  dengan akurasi sebesar 78.12%.

**Kata Kunci:** *K-Nearest Neighbors, Analisi sentimen, tweet harvest*

## ABSTRACT

A thesis is one of the graduation requirements for obtaining a bachelor's degree. However, in the Merdeka Belajar episode 26 Transformation of National Standards and Accreditation of Higher Education activities which was held on August 29 2023, the Minister of Education and Culture, Nadiem Anwar Makarim, stated that a thesis is no longer a mandatory requirement for graduating from a university degree program, but can be replaced in another form such as a prototype, projects, or other similar forms.

The stages carried out in this research are data crawling with tweet harvest, pre-processing which consists of case folding, *tokenizing*, normalization, stopword and stemming. After that, labeling was carried out using the Indonesian lexicon dictionary which was determined manually. Then word weighting, then the data classification process using the K-Nearest Neighbor method and confusion matrix.

The results of this research used 1924 tweets that had gone through a labeling process, containing 68% neutral sentiment, 18% positive sentiment and 13% negative sentiment. From the calculation results of 3 artistic analysis test schemes using the K-Nearest Neighbors method using the Rapidminer tool, the best results were obtained using 90% training data of 1732 data and 192 10% test data with  $k=5$  with an accuracy of 78.12%.

**Keywords:** *K-Nearest Neighbors, Sentiment Analysis, tweet harvest*