



**KLASIFIKASI SENTIMEN PADA DATA TWITTER MENGGUNAKAN  
ALGORITMA KNN (STUDI KASUS: MAGANG MERDEKA BELAJAR)**

**TUGAS AKHIR**

**HOLLYWRIT TRAVAGANZ NAINGGOLAN**

**1810511053**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2022**



**KLASIFIKASI SENTIMEN PADA DATA TWITTER MENGGUNAKAN  
ALGORITMA KNN (STUDI KASUS: MAGANG MERDEKA BELAJAR)**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Sarjana Komputer**

**HOLLYWRIT TRAVAGANZ NAINGGOLAN**

**1810511053**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2022**

## PERNYATAAN ORISINALITAS

### PERNYATAAN ORISINALITAS

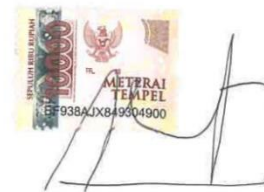
Skripsi ini adalah karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Hollywrit Travaganz Nainggolan  
NIM : 1810511053  
Tanggal: 24 Juni 2021

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan berikut, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 24 Juni 2022

Yang menyatakan

A handwritten signature in black ink is written over a yellow and green 5000 Rupiah stamp. The stamp features the Garuda Pancasila emblem and the text 'REPUBLIK INDONESIA', '5000', and 'METRAI TEPAPEL'. The serial number 'BF938AJX849004900' is visible at the bottom of the stamp.

(Hollywrit Travaganz Nainggolan)

# PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

## PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

---

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Hollywrit Travaganz Nainggolan  
NIM : 1810511053  
Fakultas : Ilmu Komputer  
Program Studi : Informatika

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

### **KLASIFIKASI SENTIMEN MAGANG MERDEKA BELAJAR MENGUNAKAN ALGORITMA KNN**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan menerbitkan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 24 Juni 2022

Yang Menyatakan,



Hollywrit Travaganz Nainggolan

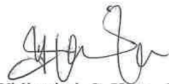
## LEMBAR PENGESAHAN

### PENGESAHAN

Skripsi yang diajukan oleh:

Nama : Hollywrit Travaganz Nainggolan  
NIM : 1810511053  
Program Studi : Informatika/S1  
Judul Skripsi : Klasifikasi Sentimen Magang Merdeka  
Belajar Menggunakan Algoritma KNN

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Yuni Widiastwi, S. Kom., M. Kom.

Ketua Penguji



Mayanda Mega Santoni, S. Kom., M. Kom.

Penguji



Bayu Hananto, S. Kom., M. Kom.

Pembimbing I



Bambang Tri Wahyono, S. Kom., M. Si.

Pembimbing II



Dr. Ermatita, M. Kom.

Dekan



Desta Sandya Prasvita, S. Kom., M. Kom.

Ketua Program Studi

Ditetapkan di  
Pada Tanggal

: Jakarta  
: 27 Juli 2022



***KLASIFIKASI SENTIMEN MAGANG MERDEKA BELAJAR  
MENGUNAKAN ALGORITMA KNN***

**Hollywrit Travaganz Nainggolan**

**ABSTRAK**

Pada tahun 2020, Kementerian Pendidikan dan Kebudayaan meluncurkan program kerja Merdeka Belajar : Kampus Merdeka, di mana program kerja ini meliputi Studi Independen, Pertukaran Mahasiswa, dan Magang. Namun, program kerja ini menuai polemik, di mana beberapa peserta program magang menceritakan beberapa pengalaman mereka, baik itu positif maupun negatif.

Salah satu cara untuk melakukan klasifikasi adalah menggunakan metode KNN. Alasan peneliti menggunakan algoritma KNN adalah kemudahan dalam implementasi dan dapat diandalkan untuk data yang cukup *noisy*.

Tujuan penelitian ini adalah untuk melakukan klasifikasi sentimen pada program kerja Kemendikbud, apakah program ini dapat diterima oleh masyarakat luas.

**Kata Kunci:** *KNN, Kampus Merdeka, Klasifikasi Sentimen*

***SENTIMENT CLASSIFICATION OF MERDEKA BELAJAR'S  
INTERNSHIP PROGRAMME USING KNN ALGORITHM***

**Hollywrith Travaganz Nainggolan**

**ABSTRACT**

On 2020, Indonesian Ministry of Education and Culture launch a program called *Merdeka Belajar: Kampus Merdeka*, this program consists of internships, independent study, and student exchange. However, in case of internships, this program sparks controversy since the participants share their experiences during the internship program by Magang Merdeka.

Author are interested to do research on this sentiment by classifying it as either positive or negative sentiment. One of the preferred method is by using KNN algorithm. Author is using KNN since it is the easy-to-use and reliable algorithm out there. The final purpose is to classifying the sentiment on Kemendikbud's program, whether it's more received or not.

**Keyword:** *KNN, Kampus Merdeka, Sentiment Classification*

## **KATA PENGANTAR**

Segala puji dan syukur peneliti panjatkan kehadiran Tuhan Yang Mahakuasa atas limpahan rahmat dan karunia-Nya sehingga peneliti dapat menyelesaikan Tugas Akhir. Tugas Akhir ini disusun sebagai syarat Tugas Akhir Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Dalam penyelesaian karya tulis ilmiah ini tidak lepas dari bantuan banyak pihak yang telah memberikan masukan kepada peneliti. Untuk itu peneliti mengucapkan terima kasih kepada:

1. Dr. Ermatita, M.Kom., selaku dekan Fakultas Ilmu Komputer
2. Desta Sandya P., S. Kom., M. Kom. selaku Ketua Program Studi Sarjana Jurusan S1 Informatika.
3. Bayu Hananto, S. Kom., M. Kom. selaku dosen pembimbing 1 dari pihak jurusan.
4. Bambang Tri Wahyono, S. Kom, M. Si., selaku dosen pembimbing 2 dari pihak jurusan.
5. Iiswatun Hasanah
6. Central Pahhk
7. Para mentor, sahabat, dan keluarga besar peneliti
8. Seluruh pihak yang terlibat dalam kelancaran pembuatan makalah karya ilmiah ini dan yang belum disebutkan di atas, peneliti ucapkan terima kasih.

Peneliti menyadari bahwa masih banyak kekurangan dari laporan ini, baik dari materi maupun cara penyampaian. Oleh karena itu, kritik dan saran yang membangun peneliti harapkan.

Jakarta, 24 Juni 2022

Peneliti



## DAFTAR ISI

<b>PERNYATAAN ORISINALITAS</b> .....	iv
<b>PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS</b> .....	v
<b>LEMBAR PENGESAHAN</b> .....	ii
<b>ABSTRAK</b> .....	iii
<b>ABSTRACT</b> .....	iv
<b>KATA PENGANTAR</b> .....	v
<b>DAFTAR ISI</b> .....	vi
<b>DAFTAR TABEL</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	x
<b>DAFTAR SIMBOL</b> .....	xi
<b>BAB 1 PENDAHULUAN</b> .....	1
<b>1.1. Latar Belakang Masalah</b> .....	1
<b>1.2. Perumusan Masalah</b> .....	2
<b>1.3. Ruang Lingkup</b> .....	2
<b>1.4. Tujuan</b> .....	2
<b>1.5. Manfaat</b> .....	2
<b>1.6. Luaran yang Diharapkan</b> .....	2
<b>1.7. Sistematika Penulisan</b> .....	3
<b>BAB 2 TINJAUAN PUSTAKA</b> .....	4
<b>2.1. Merdeka Belajar</b> .....	4
<b>2.2. Klasifikasi Sentimen</b> .....	4
<b>2.3. <i>Text Mining</i></b> .....	4
<b>2.4. Kappa Value</b> .....	5

2.5.	Metodologi Klasifikasi .....	6
2.5.1.	<i>Near Miss</i> .....	7
2.5.2.	TF-IDF .....	7
2.5.3.	K-Nearest Neighbor .....	8
2.5.4.	<i>Confusion Matrix</i> .....	9
2.6.	Penelitian Terdahulu .....	11
<b>BAB 3 METODOLOGI PENELITIAN .....</b>		<b>13</b>
3.1.	Bagan Alir .....	13
3.1.1.	Identifikasi Masalah .....	15
3.1.2.	Studi Literatur .....	15
3.1.3.	Penarikan Data .....	15
3.1.4.	Pelabelan Data .....	16
3.1.5.	Pra-pemrosesan Data .....	16
3.1.6.	Pembobotan TF-IDF .....	17
3.1.7.	Klasifikasi .....	17
3.1.8.	Hasil Klasifikasi .....	18
3.1.9.	<i>Confusion Matrix</i> .....	18
3.2.	Perangkat Pendukung .....	19
3.3.	Waktu Penelitian .....	19
<b>BAB 4 HASIL DAN PEMBAHASAN .....</b>		<b>21</b>
4.1.	Data .....	21
4.1.1.	Pelabelan Data .....	21
4.2.	Pra-proses Data .....	25
4.2.1.	Case Folding .....	25
4.2.2.	Cleaning .....	26
4.2.3.	Tokenisasi .....	27

4.2.4. Stopword removal .....	28
4.2.5. Stemming .....	29
4.2.6. Normalisasi .....	29
4.3. Pembobotan TF-IDF .....	30
4.4. <i>Data Balancing</i> .....	33
4.5. Proses Klasifikasi .....	34
4.5.1. <i>K-Nearest Neighbor</i> .....	34
4.6. Hasil Klasifikasi .....	35
4.7. Evaluasi .....	36
<b>BAB 5 KESIMPULAN DAN SARAN</b> .....	39
5.1. Kesimpulan .....	39
5.2. Saran .....	39
<b>DAFTAR PUSTAKA</b> .....	40
<b>RIWAYAT HIDUP</b> .....	42
<b>LAMPIRAN</b> .....	43






## DAFTAR TABEL

Tabel 1 <i>Kappa Value</i> .....	5
Tabel 2 <i>Confusion Matrix</i> .....	14
Tabel 3 Waktu Penelitian.....	17
Tabel 4 Hasil Pelabelan.....	19
Tabel 5 Sampel Pelabelan untuk <i>Kappa Value</i> .....	20
Tabel 6 Data Sebelum Pra-Proses.....	22
Tabel 7 Hasil <i>Case Folding</i> .....	23
Tabel 8 Hasil Pembersihan Data.....	24
Tabel 9 Hasil Tokenisasi.....	24
Tabel 10 Hasil Pembuangan <i>Stopword</i> .....	25
Tabel 11 Hasil <i>Stemming</i> .....	26
Tabel 12 Hasil Normalisasi.....	27
Tabel 13 Sampel Data Latih.....	28
Tabel 14 Pembobotan TF-IDF.....	28
Tabel 15 Hasil <i>Balancing Data</i> .....	30
Tabel 16 Evaluasi.....	32
Tabel 17 Hasil Evaluasi.....	34

## DAFTAR GAMBAR

Gambar 1 Bagan Alir.....	7
Gambar 2 Hasil <i>Crawling</i> Data.....	16

## DAFTAR SIMBOL

<b>Simbol Flowchart</b>			
NO	GAMBAR	NAMA	KETERANGAN
1		<i>Terminal</i> ( <i>start, end</i> )	Menggambarkan bagaimana kegiatan dimulai atau kegiatan berakhir.
2		<i>Flow</i> <i>Direction</i>	Menggambarkan hubungan antar simbol yang menyatakan suatu jalannya proses dalam sistem.
3		<i>Process</i>	Menggambarkan deskripsi dari proses yang dijalankan.
4		<i>Document</i>	Menggambarkan bahwa masukkan ( <i>input</i> ) berasal dari sebuah data dokumen yang dapat berupa kertas atau keluaran ( <i>output</i> ) yang dicetak ke kertas.
5		<i>Process</i>	Menggambarkan pelaksanaan dari sebuah proses atau disebut sebagai subprogram.