

**KLASIFIKASI TEKS PADA *TWITTER* MENGGUNAKAN  
*METODE K-NEAREST NEIGHBOR* DAN SELEKSI FITUR  
*INFORMATION GAIN***

**(Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua  
DPR Ibu Puan Maharani di *Twitter* )**



Alfa Yunitra

1810651042

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

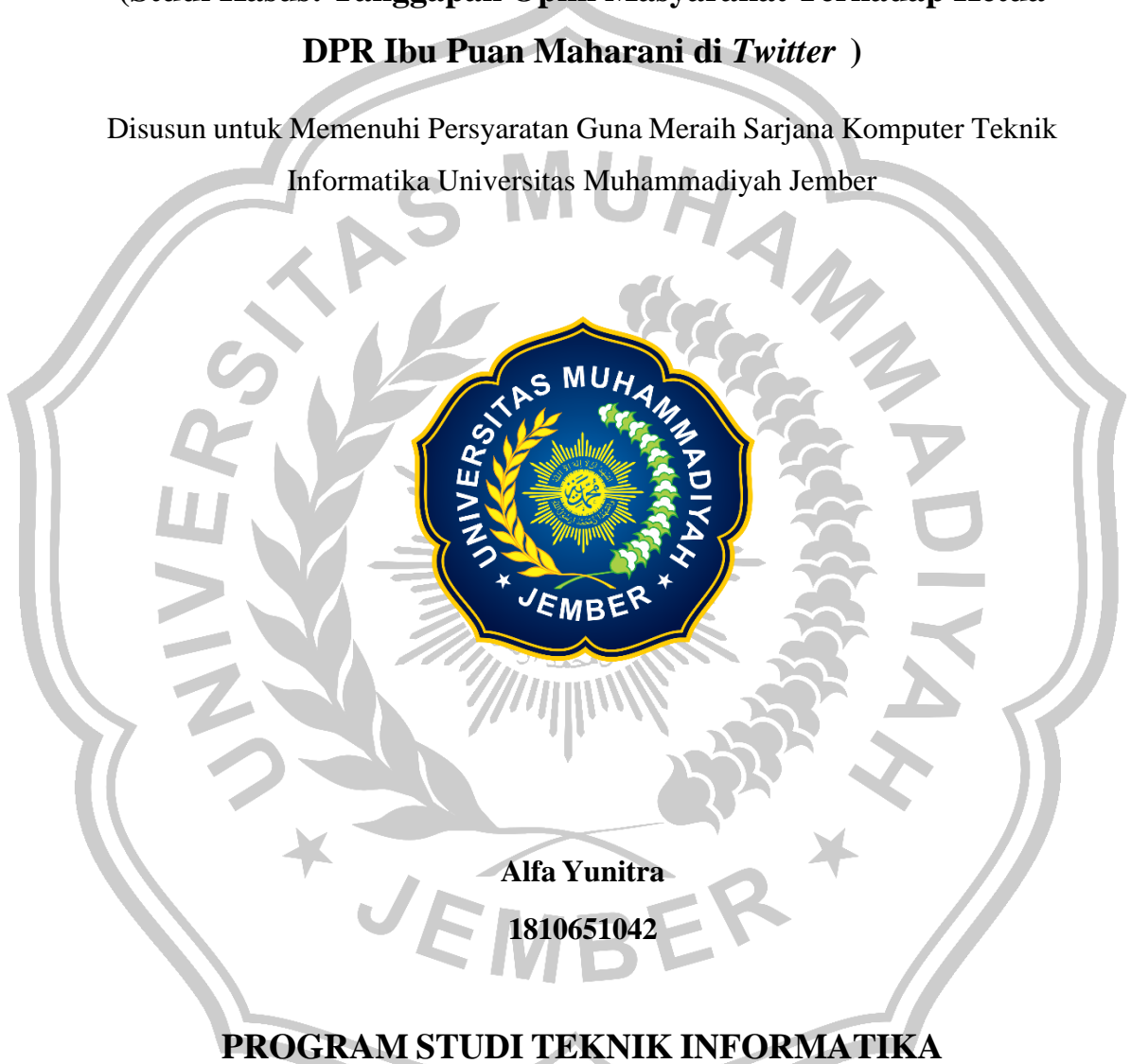
**2024**

**TUGAS AKHIR**

**KLASIFIKASI TEKS PADA *TWITTER* MENGGUNAKAN  
*METODE K-NEAREST NEIGHBOR* DAN SELEKSI FITUR  
*INFORMATION GAIN***

**(Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua  
DPR Ibu Puan Maharani di *Twitter* )**

Disusun untuk Memenuhi Persyaratan Guna Meraih Sarjana Komputer Teknik  
Informatika Universitas Muhammadiyah Jember



**Alfa Yunitra**

**1810651042**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

**2024**

**HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR**

**KLASIFIKASI TEKS PADA *TWITTER* MENGGUNAKAN  
*METODE K-NEAREST NEIGHBOR* DAN SELEKSI FITUR  
*INFORMATION GAIN***

**(Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua  
DPR Ibu Puan Maharani di *Twitter* )**

Oleh

Alfa Yunitra

1810651042

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada  
sidang Tugas Akhir sebagai salah satu syarat kelulusan dan  
mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Menyetujui

Pembimbing 1

Pembimbing 2



Agung Nilogiri, S.T., M.Kom

NIDN.0030037701



Deni Arifianto, M.Kom

NIDN.0718068103

HALAMAN PENGESAHAN

KLASIFIKASI TEKS PADA *TWITTER* MENGGUNAKAN  
*METODE K-NEAREST NEIGHBOR* DAN SELEKSI FITUR  
*INFORMATION GAIN*

(Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua  
DPR Ibu Puan Maharani di *Twitter* )

Oleh

Alfa Yunitra

1810651042

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir Tanggal 20 Bulan Juli Tahun 2024 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui Oleh

Dosen Penguji :

Penguji 1

Dr. Bagus Setya Rintyarna, ST., M.Kom

NIDN.0729017904

Penguji 2

Habibatul Azizah Al Faruq, M.Pd

NIDN.0718128901

Mengesahkan, Dekan Fakultas Teknik

Dr. Ir. Muhtar, S.T., M.T., IPM

NIDN. 0010067301

Dosen Pembimbing :

Pembimbing 1

Agung Nilogiri, S.T., M.Kom

NIDN0030037701

Pembimbing 2

Deni Arifianto, M.Kom

NIDN.0718068103

Mengetahui, Ketua Program  
Studi Teknik Informatika

Rosita Yumarti, S.Kom., M. Cs

NIDN. 0629018601

# KLASIFIKASI TEKS PADA TWITTER MENGGUNAKAN METODE *K-NEAREST NEIGHBORD* DAN SELEKSI FITUR *INFORMATION GAIN*

(Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua  
DPR Ibu Puan Maharani di *Twitter* )

Alfa Yunitra<sup>1</sup>, Agung Nilogiri<sup>2</sup>, Deni Arifianto<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email: <sup>1</sup>Alfayunitra12@gmail.com

## ABSTRAK

Indonesia adalah salah satu negara dengan pemakai internet yang paling cepat berkembang dengan sekitar 196 juta penduduk atau 73,3% dari populasi terjangkau internet. Perkembangan teknologi pada tahun 2017 mencapai 3,4 miliar atau 45% dari total populasi di dunia, dan pada tahun 2022 perkembangan teknologi juga diproyeksikan akan menyentuh 4,8 miliar atau sekitar 60% dari populasi di dunia. Oleh karena itu dalam penelitian ini akan menentukan perhitungan presisi dan akurasi tingkat terbaik untuk melakukan sebuah klasifikasi komentar pengguna di twitter ke dalam kelas sentimen positif juga, sentiment negatif. Metode yang digunakan yaitu *K-Nearest Neighbord* dan seleksi fitur *Information Gain*. Tujuannya untuk menganalisis tingkat perbandingan akurasi dan performa presisi *Information Gain* dan *K-Nearest Neighbor* dalam mengelompokkan komentar pengguna twitter ke dalam kelas sentimen positif dan sentiment negatif. Setelah melakukan proses perhitungan dengan metode *K-Nearest Neighbor* dan seleksi fitur *Information Gain* dengan uji coba beberapa fold antara lain 3,5, dan 9 dimana nilai  $k$  pada knn 3,5,9. Dan telah didapatkan hasil akurasi terbaik tanpa fitur 61.6% pada fold 5 dengan nilai  $k=3$ . Sedangkan hasil akurasi terbaik dengan menggunakan fitur *selection* sebesar 71.3% pada fold 9 dengan nilai  $k=9$ .

**Kata kunci:** *Klasifikasi, K-nearest neighbord, Information Gain.*

**TEXT CLASSIFICATION ON TWITTER USING THE K-  
NEAREST NEIGHBORD METHOD AND INFORMATION GAIN  
FEATURE SELECTION**

**(Case Study: Public Opinion Response to the Chair of the DPR,  
Mrs. Puan Maharani, on Twitter)**

Alfa Yunitra<sup>1</sup>, Agung Nilogiri<sup>2</sup>, Deni Arifianto<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email: [1Alfayunitra12@gmail.com](mailto:1Alfayunitra12@gmail.com)

**ABSTRACT**

*Indonesia is one of the countries with the fastest growing internet users with around 196 million residents or 73.3% of the population with internet access. Technological developments in 2017 reached 3.4 billion or 45% of the world's total population, and in 2022 technological developments are also projected to reach 4.8 billion or around 60% of the world's population. Therefore, this research will determine the best level of precision and accuracy calculations for classifying user comments on Twitter into positive sentiment classes as well as negative sentiment. The method used is K-Nearest Neighbor and Information Gain feature selection. The aim is to analyze the comparative level of accuracy and precision performance of Information Gain and K-Nearest Neighbor in grouping Twitter user comments into positive sentiment and negative sentiment classes. After carrying out the calculation process using the K-Nearest Neighbor method and selecting the Information Gain feature by testing several folds, including 3, 5, and 9, where the k value in knn is 3, 5, 9. And the best accuracy results without features were obtained, 61.6% on fold 5 with a value of k=3. Meanwhile, the best accuracy results using the selection feature were 71.3% on fold 9 with a value of k=9.*

**Keywords:** *Classification, K-nearest neighbor, Information Gain.*

## HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Alfa Yunitra

Nim : 1810651042

Program Studi : Teknik Informatika

Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul “KLASIFIKASI TEKS PADA TWITTER MENGGUNAKAN METODE *K-NEAREST NEIGHBORD* DAN SELEKSI FITUR *INFORMATION GAIN* (Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua DPR Ibu Puan Maharani di *Twitter* ) ”bukan merupakan Tugas Akhir orang lain baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan yang telah disebutkan sumbernya. Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar penulis bersedia mendapatkan sanksi akademik.

Jember, 26 Juli 2024

Yang Menyatakan,



Alfa Yunitra

NIM. 1810651042

## MOTTO

Takdir setiap manusia memang telah ditentukan sejak mereka lahir, tetapi dengan kerja keras kita dapat mengalahkan takdir."

( Naruto Uzumaki )





## PERSEMBAHAN

Alhamdulillah dengan penuh rasa syukur, saya panjatkan puji syukur kehadirat Allah SWT atas rahmat dan karunia-Nya yang memungkinkan saya menyelesaikan tugas akhir dan tanggung jawab sebagai mahasiswa. Semoga Allah meridhoi dan memberikan hikmah dari ilmu yang telah diperoleh. Sholawat serta salam semoga senantiasa terlimpah kepada Nabi Muhammad SAW. Sebagai ucapan terimakasih penulis persembahkan kepada:

1. Kepada Allah SWT yang telah memberikan kemudahan dalam menyelesaikan skripsi ini.
2. Kepada kedua orang tua saya tercinta, Bapak Muswar dan Ibu Ruslina yang telah senantiasa memberikan cinta, dukungan, motivasi, dan doa terbaik sehingga saya bisa menyelesaikan masa studi saya.
3. Bapak Agung Nilogiri, S.T., M.Kom. selaku dosen pembimbing 1, dan Bapak Deni Arifianto, M.Kom selaku dosen pembimbing 2 yang sudah dengan sabar membimbing dan memberi masukan serta saran selama ini sehingga saya dapat menyelesaikan skripsi tugas akhir ini.
4. Kepada adik-adik tercinta saya, Ifalntri Mesar dan Alfis Mesar yang telah memberikan semangat, dukungan serta doa terbaik.
5. Teman serta sahabat saya Oktavia Indah, Verenita, Ajeng, Salsabilla, Amada, Dani Nurpauji, Rangga Nichola, yang telah menemani, membantu dan senantiasa memberikan arahan, dukungan serta motivasi sehingga tugas akhir ini dapat terselesaikan.
6. Seluruh teman seangkatan saya Teknik Informatika 2018.

## KATA PENGANTAR

Alhamdulillah, Puji dan Syukur kita panjatkan kepada Allah Subhanahu Wata'ala. Dzat yang hanya kepada-Nya memohon pertolongan. Alhamdulillah atas segala pertolongan, rahmat, dan kasih sayang-Nya sehingga penulis dapat menyelesaikan skripsi tugas akhir yang berjudul: “**KLASIFIKASI TEKS PADA TWITTER MENGGUNAKAN METODE *K-NEAREST NEIGHBORD* DAN SELEKSI FITUR *INFORMATION GAIN*(Studi Kasus: Tanggapan Opini Masyarakat Terhadap Ketua DPR Ibu Puan Maharani di *Twitter*)**”. Shalawat serta salam tidak lupa kita panjatkan kepada Rasulullah Shallallahu Alaihi Wasallam dan sahabat yang senantiasa menjadi sumber inspirasi dan teladan terbaik untuk umat manusia.

Penulis menyadari banyak pihak yang memberikan dukungan dan bantuan selama menyelesaikan studi dan tugas akhir ini. Oleh karena itu, sudah sepantasnya penulis dengan penuh hormat mengucapkan terimakasih dan mendoakan semoga Allah memberikan balasan terbaik kepada:

1. Bapak Dr. Ir.Muhtar, S.T., M.T., IPM. Selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
2. Ibu Rosita Yanuarti, S.Kom., M.Cs Selaku Ketua Prodi Teknik Informatika Universitas Muhammadiyah Jember.
3. Bapak Agung Nilogiri, S.T., M.Kom. dan Bapak Deni Arifianto, M.Kom selaku dosen pembimbing penulis dalam menyelesaikan tugas akhir ini..
4. Bapak Dr. Bagus Setya Rintyarna, S.T., M.Kom dan Ibu Habibatul Azizah Alfaruq., M.Pd yang telah menjadi penguji dalam seminar proposal, seminar hasil penelitian dan sidang skripsi penulis.

Terimakasih penulis juga haturkan untuk semua pihak yang telah membantu peneliti dalam menyelesaikan skripsi ini yang tidak dapat penulis sebutkan satu persatu. Akhir kata penulis menyadari bahwa tidak ada yang sempurna, penulis masih melakukan kesalahan dalam penyusunan skripsi. Oleh karena itu, penulis meminta maaf yang sedalam-dalamnya atas kesalahan yang dilakukan penulis. Penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca dan dapat dijadikan referensi demi pengembangan ke arah yang lebih baik.

## DAFTAR ISI

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terkait.....	5
2.2 <i>Sentiment analysis</i> .....	6
2.3 <i>Twitter</i> .....	6
2.4 <i>Twitter API (Application Programming Interface)</i> .....	7
2.5 <i>Text Mining</i> .....	7
2.5.1 <i>Case Folding</i> .....	7
2.5.2 <i>Cleansing</i> .....	8
2.5.3 <i>Stopword Removal / Filtering</i> .....	9
2.5.4 <i>Stemming</i> .....	10
2.6 <b><i>Information Gain</i></b> .....	11
2.7 <i>Metode K-Nearest Neighbor (KNN)</i> .....	11
2.8 <i>Term Frequency- Inverse Document Frequency (TF-IDF)</i> .....	12
2.9 <i>K-Fold Cross Validation</i> .....	14
2.10 <i>Confusion Matrix</i> .....	14
2.11 <i>Phyton</i> .....	15
BAB III METODOLOGI PENELITIAN.....	16
3.1 Tahapan Penelitian.....	16
3.1.1 Studi Pustaka.....	16

3.1.2	Pengumpulan Data.....	16
3.1.3	<i>Klasifikasi</i> .....	17
3.1.4	Penyusunan Laporan .....	17
3.2	Alur Proses Sistem.....	17
3.3	<i>Pre-Processing</i> .....	18
3.4	Pembobotan <i>TF-IDF</i> .....	21
3.5	<i>K-Nearest Neighbord (KNN)</i> .....	22
3.6	<i>Fitur Selection Information Gain (IG)</i> .....	23
3.7	Perhitungan Akurasi dengan <i>Confusion Matrix</i> .....	24
BAB IV .....		26
HASIL DAN PEMBAHASAN .....		26
4.1	Proses <i>Preprocessing</i> .....	26
4.2	Proses <i>Klasifikasi</i> Menemukan K terbaik dengan <i>K-Fold Cross Validation</i> .....	32
4.3	Proses <i>Klasifikasi</i> Menemukan K terbaik dengan <i>K-fold Cross Validation</i> dan <i>Information Gain</i> . .....	37
4.4	Hasil Akhir Rata-rata <i>Klasifikasi</i> dengan <i>Confusion Matrix</i> .....	43
KESIMPULAN DAN SARAN.....		50
5.1	Kesimpulan .....	50
5.2	Saran .....	50
DAFTAR PUSTAKA.....		51
LAMPIRAN.....		53

## DAFTAR GAMBAR

Gambar 1. 1 Data Pengguna Internet Dalam Satuan Juta.....	1
Gambar 2. 1 <i>Flowchart Case Folding</i> .....	8
Gambar 2. 2 <i>Flowchart Cleansing</i> .....	9
Gambar 2. 3 <i>Flowchart Stopword</i> .....	10
Gambar 2. 4 <i>Flowchart Stemming</i> .....	11
Gambar 2. 5 Kerja <i>K-Fold Cross Validation</i> .....	14
Gambar 3. 1 Bagan Tahapan Penelitian .....	16
Gambar 3. 2 <i>Flowchart Proses Klasifikasi</i> .....	18
Gambar 3. 3 <i>Flowchart Pre-Processing</i> .....	19
Gambar 4. 1 Proses <i>Case Folding</i> .....	26
Gambar 4. 2 Proses <i>Cleansing</i> .....	28
Gambar 4. 3 Proses <i>Stopword Removal</i> .....	29
Gambar 4. 4 Proses <i>Stemming</i> .....	30
Gambar 4. 5 Proses <i>K-Nearest Neighbour</i> .....	32
Gambar 4. 6 Proses <i>Information Gain</i> .....	37

## DAFTAR TABEL

Tabel 2. 1 Contoh <i>Confusion Matrix</i> .....	14
Tabel 3. 1 Proses <i>Case Folding</i> .....	19
Tabel 3. 2 Proses <i>Cleansing</i> .....	20
Tabel 3. 3 Daftar Kata <i>Stopword</i> .....	21
Tabel 3. 4 Hasil <i>Stemming</i> .....	21
Tabel 3. 5 Perhitungan Tf-Idf.....	21
Tabel 3. 6 Sampel List Hasil Proses Klasifikasi.....	23
Tabel 3. 7 Perhitungan Dengan <i>Confusion Matrix</i> K=3.....	24
Tabel 4. 1 Hasil <i>Case Folding</i> .....	27
Tabel 4. 2 Hasil Proses <i>Cleansing</i> .....	29
Tabel 4. 3 Hasil Proses <i>Stopword Removal</i> .....	30
Tabel 4. 4 Hasil Proses <i>Stemming</i> .....	31
Tabel 4. 5 Hasil K=3 .....	32
Tabel 4. 6 Hasil K=5 .....	33
Tabel 4. 7 Hasil K=9 .....	33
Tabel 4. 8 Hasil K=3 .....	34
Tabel 4. 9 Hasil K=5 .....	34
Tabel 4. 10 Hasil K=9 .....	35
Tabel 4. 11 Hasil K=3 .....	35
Tabel 4. 12 Hasil K=5 .....	36
Tabel 4. 13 Hasil K=9 .....	37
Tabel 4. 14 Hasil K=3 .....	38
Tabel 4. 15 Hasil K=5 .....	38
Tabel 4. 16 Hasil K=9 .....	39
Tabel 4. 17 Hasil K=3 .....	39
Tabel 4. 18 Hasil K=5 .....	40
Tabel 4. 19 Hasil K=9 .....	40
Tabel 4. 20 Hasil K=3 .....	41
Tabel 4. 21 Hasil K=5 .....	42

Tabel 4. 22 Hasil K=9 .....	42
Tabel 4. 23 Rata-rata 3 <i>Fold</i> Tanpa Fitur .....	43
Tabel 4. 24 Rata-rata 5 <i>Fold</i> Tanpa Fitur .....	44
Tabel 4. 25 Rata-rata 9 <i>Fold</i> Tanpa Fitur .....	45
Tabel 4. 26 Rata-rata 3 <i>Fold</i> Menggunakan Fitur .....	46
Tabel 4. 27 Rata-rata 5 <i>Fold</i> Menggunakan Fitur .....	47
Tabel 4. 28 Rata-rata 9 <i>Fold</i> Menggunakan Fitur .....	48

