

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP RENCANA
PENGESAHAN RKUHP 2019 MENGGUNAKAN METODE
K-NEAREST NEIGHBOR**

SKRIPSI



disusun oleh

Faishal Fadhlulloh

16.11.0836

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP RENCANA
PENGESAHAN RKUHP 2019 MENGGUNAKAN METODE
K-NEAREST NEIGHBOR**

SKRIPSI

untuk memenuhi sebagai persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Faishal Fadhlulloh

16.11.0836

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP RENCANA PENGESAHAN RKUHP 2019 MENGGUNAKAN METODE

K-NEAREST NEIGHBOR

yang dipersiapkan dan disusun oleh

Faishal Fadhlulloh

16.11.0836

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Desember 2019

Dosen Pembimbing,

Kusnawi, S.Kom, M. Eng.

NIK : 190302112

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP RENCANA
PENGESAHAN RKUHP 2019 MENGGUNAKAN METODE
K-NEAREST NEIGHBOR**

yang dipersiapkan dan disusun oleh

Faishal Fadhlulloh

16.11.0836

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 September 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto, M.Kom

NIK : 190302163

Yuli Astuti, M.Kom

NIK : 190302146

Kusnawi, S.Kom, M. Eng.

NIK : 190302112

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Tanggal 7 Oktober 2020

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.

NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 20 Oktober 2020



Faishal Radhulloh

15.11.0836

MOTTO

"Sedikit lebih beda, lebih baik daripada sedikit lebih baik"

- Seth Godin -

"Time won't make you forget, it will make you understand things"

- X-Moon -

**"You'll never know when someone comes in
and press play on your paused life"**

- Pandji Pragiwaksono -

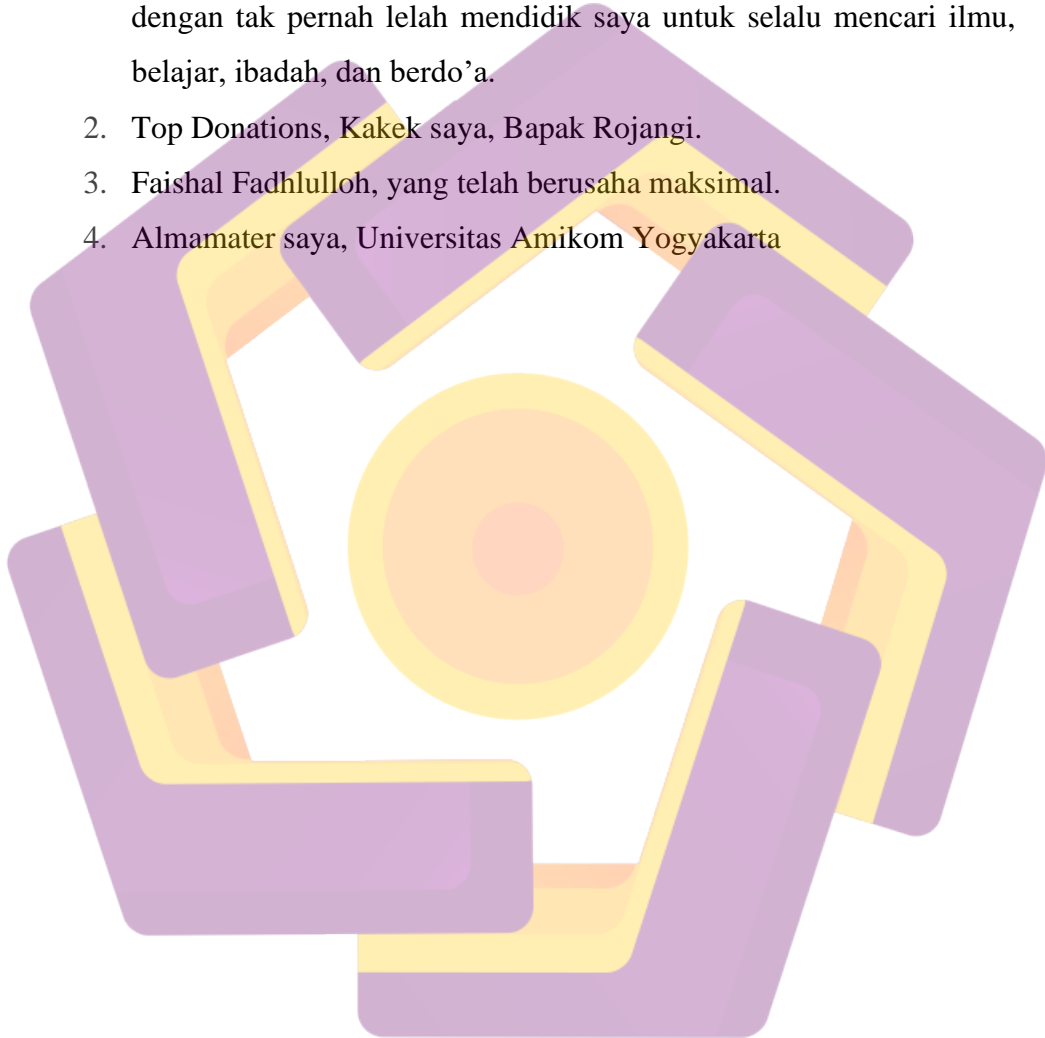
"Don't be racist, just hate everyone"

- FAK -

PERSEMBAHAN

Skripsi ini saya persembahkan kepada :

1. Kedua orang tua saya, Sujiman akhmad rifai dan siti kharfiah yang selalu membimbing dan memberikan do'a serta semangat buat saya dengan tak pernah lelah mendidik saya untuk selalu mencari ilmu, belajar, ibadah, dan berdo'a.
2. Top Donations, Kakek saya, Bapak Rojangi.
3. Faishal Fadhlulloh, yang telah berusaha maksimal.
4. Almamater saya, Universitas Amikom Yogyakarta



KATA PENGANTAR

Bismillahirrahmanirahim, Alhamdulillahilahirabil'alamin puji syukur penulis panjatkan kepada Allah SWT atas Rahmat dan karunia-Nya , sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Sentimen Pengguna Twitter terhadap Rencana Pengesahan Rkuhp 2019 Menggunakan Metode K-Nearest Neighbor”. Untuk memenuhi syarat akademis dalam menyelesaikan Program Studi Strata Satu (S1) pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Mengingat keterbatasan pengetahuan dan pengalaman dalam penulisan skripsi ini, Penulis banyak mendapatkan bimbingan, petunjuk, saran dan arahan dari berbagai pihak, oleh karena itu dengan kerendahan hati dan rasa hormat Penulis mengucapkan terima kasih sebesar-besarnya kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. Selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Sudarmawan, S.T., M.T. Selaku Dosen wali yang telah menjadi wali saya selama menjadi mahasiswa.
3. Bapak Kusnawi, S.Kom., M.Eng. Selaku Dosen pembimbing yang telah membantu dalam penulisan skripsi ini.
4. Bapak Anggit Dwi Hartanto, M.Kom. Selaku Dosen Penguji yang telah memberikan petunjuk, serta nasehat dalam ujian skripsi ini.
5. Ibu Yuli Astuti, M.Kom. Selaku Dosen Penguji yang telah memberikan petunjuk, serta nasehat dalam ujian skripsi ini.

6. Kedua Orang tua saya, Sujiman akhmad rifai dan siti kharfiah yang selalu support doa, biaya, dan kasih sayangnya.
7. Teman teman nyusahin yang ngajakin main terus, tapi Terimakasih sudah mau jadi teman saya.
8. S1-IF 13 (BCD Class) Universitas Amikom Yogyakarta 2016, Kelas gajelas tapi Terimakasih atas support dan bantuanya selama ini.
9. Setiap orang yang sudah berkontribusi membantu saya selama ini, maupun nyusahin saya selama ini.
10. Kamu, yang memberi semangat dan motivasi untuk menyelesaikan skripsi ini.

Penulis mendoakan untuk semua pihak yang telah membantu dalam penulisan skripsi ini semoga diberikan balasan dan berkah dari Allah SWT. Penulis menyadari masih banyak kekurangan dalam penulisan skripsi ini untuk itu saran, kritik dan perbaikan yang bersifat membangun sangat diharapkan. Akhir kata Penulis berharap semoga skripsi ini bermanfaat untuk semua pihak yang membutuhkan.

Yogyakarta, 14 Oktober 2020
Penulis,



Faishal Fadhlulloh

16.11.0836

DAFTAR ISI

JUDUL	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN	Error! Bookmark not defined.
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metode Penelitian.....	4
1.6.1 Metode Pengumpulan Data.....	4
1.6.2 Metode Analisis	4
1.6.3 Metode Perancangan	5
1.6.4 Metode Pengujian.....	5
1.7 Sistematika Penelitian	6
BAB II Landasan Teori	8
2.1 Kajian Pustaka	8
2.2 Dasar Teori	15
2.2.1 Data Mining	15
2.2.2 Pengelompokan Data Mining.....	17

2.2.3	Klasifikasi	19
2.2.4	Analisis Sentimen	20
2.2.5	Twitter	21
2.2.6	Analisis Sentimen Twitter.....	23
2.2.7	Twitter API	24
2.2.8	Text Mining.....	25
2.2.9	Text Preprocessing	26
2.2.10	Algoritma Nazief & Adriani	28
2.2.11	K-Nearest Neighbors (K-NN).....	30
2.2.12	Term Frequency Inverse Document Frequency (TF-IDF).....	32
2.2.13	Cosine Similarity.....	34
2.2.14	Confusion Matrix	35
2.2.15	Confusion Matrix	36
BAB III METODE PENELITIAN		39
3.1	Mengidentifikasi Masalah	40
3.2	Merumuskan Masalah	40
3.3	Penentuan Tujuan	40
3.4	Studi Literatur.....	40
3.5	Analisis dan Perancangan.....	41
3.5.1	Analisis Kebutuhan Sistem	41
3.5.1.1	Kebutuhan Fungsional	41
3.5.1.2	Kebutuhan Nonfungsional	41
3.5.2	Analisis Data	42
3.5.3	Penentuan Query	44
3.5.4	Preprocessing	45
3.5.4.1	Cleaning	45
3.5.4.2	Case Folding	46
3.5.4.3	Tokenizing.....	47
3.5.4.5	Normalization.....	47
3.5.4.6	Filtering dan Stemming.....	48
3.5.5	Analisis menggunakan Algoritma K-NN.....	52

3.5.5.1	Pembangunan index dan pembobotan.....	53
3.5.5.2	Preprocessing query.	67
3.5.5.3	Klasifikasi query.	70
3.5.6	Perancangan Sistem	74
3.5.6.1	Diagram Alir Proses Data	75
3.5.7	Perancangan Antarmuka	75
3.5.8	Pengujian Akurasi	77
BAB IV	HASIL DAN PEMBAHASAN	78
4.1	Implementasi	78
4.1.1	Data	78
4.1.2	Preprocessing	80
4.1.2.1	Cleaning	80
4.1.2.2	Case Folding	83
4.1.2.3	Tokenizing.....	84
4.1.2.4	Normalization.....	85
4.1.2.5	Filtering dan Stemming.....	87
4.1.3	Algoritma K-NN	90
4.3	Hasil dan pembahasan	91
4.3.1	Hasil Akurasi.....	91
4.3.2	Hasil Sentimen	101
BAB V	KESIMPULAN DAN SARAN	106
5.1	Kesimpulan.....	106
5.2	Saran	107
DAFTAR PUSTAKA		108
LAMPIRAN.....		112

DAFTAR TABEL

Tabel 2. 1 Perbandingan Metode K-NN untuk Klasifikasi data	11
Tabel 2. 2 Perbandingan Metode K-NN dengan Metode Lainnya.....	13
Tabel 2. 3 Confusion matrix.....	35
Tabel 3. 1 Contoh data latih	54
Tabel 3. 2 Data hasil cleaning	56
Tabel 3. 3 Data hasil case folding	57
Tabel 3. 4 Data hasil cleaning	58
Tabel 3. 5 Data hasil filtering/stopword dan stemming	60
Tabel 3. 6 Data hasil normalisasi.....	61
Tabel 3. 7 Data hasil indexing.....	62
Tabel 3. 8 Data hasil pembobotan Term Frequency (TF)	65
Tabel 3. 9 Data hasil pembobotan TF-IDF	66
Tabel 3. 10 Tabel Sampel data uji.....	71
Tabel 3. 11 Hasil Cosine Similarity (Langkah 1).....	71
Tabel 3. 12 Hasil Cosine Similarity (Langkah 2).....	72
Tabel 3. 13 Hasil Cosine Similarity (Langkah 3).....	72
Tabel 4. 1 Sampel data latih	80
Tabel 4. 2 Hasil output program cleaning	82
Tabel 4. 3 Hasil output program case folding	83
Tabel 4. 4 Hasil output program tokenizing	85
Tabel 4. 5 Hasil output program normalizing	86
Tabel 4. 6 Hasil output program filtering	88
Tabel 4. 7 Hasil output program stemming	90
Tabel 4. 8 Hasil output program tahap K-NN	91
Tabel 4. 9 Confusion matrix.....	92
Tabel 4. 10 Confuse Matrix tanpa preprocessing	93
Tabel 4. 11 Metrics Classification Report tanpa preprocessing.....	95
Tabel 4. 12 Confuse Matrix K-179	97

Tabel 4. 13 Metrics Classification Report K-179100
Tabel 4. 14 Hasil pengujian sentimen101
Tabel 4. 15 Hasil Sentimen103



DAFTAR GAMBAR

Gambar 3.1 <i>Alur Metode Penelitian</i>	39
Gambar 3. 2 Jumlah sentimen negative dan positif dataset	43
Gambar 3. 3 Pembagian data sentiment twitter.....	43
Gambar 3. 4 Flowchart dari proses penentuan query	44
Gambar 3. 5 Contoh hasil cleaning	46
Gambar 3. 6 Contoh hasil case folding	46
Gambar 3. 7 Contoh hasil tokenizing	47
Gambar 3. 8 Contoh hasil Normalization.....	48
Gambar 3. 9 Algoritma Normalisasi dan Algoritma Nazief & Adriani	49
Gambar 3. 10 Contoh filtering dan stemming	52
Gambar 3. 11 flowchart diagram analis algoritma K-NN.....	53
Gambar 3. 12 Pembangunan index dan pembobotan	54
Gambar 3. 13 Preprocessing query.....	67
Gambar 3. 14 Perancangan sistem	74
Gambar 3. 15 Alir proses data.....	75
Gambar 3. 16 Perancangan sistem	76
Gambar 3. 17 Pengujian akurasi algoritma K-NN	77
Gambar 4. 1 Jumlah sentimen negatif dan positif dataset.....	79
Gambar 4. 2 Pembagian data sentiment twitter.....	79
Gambar 4. 3 Grafik Hasil Hyperparameter 1-200.....	96
Gambar 4. 4 Grafik Hasil Hyperparameter 150-200.....	97
Gambar 4. 5 Perbandingan Akurasi	100
Gambar 4. 6 Persentase sentimen.....	105

INTISARI

Angka pengguna internet yang mengakses media sosial mencapai 85,7%. Sosial media Twitter memiliki pengguna 7.05% atau sekitar 12,06 juta pengguna di Indonesia. Terdapat banyak informasi dan tweet dari para penggunanya tentang berbagai hal, salah satunya kebijakan pemerintah. Kebijakan pemerintah yang trending topic di Twitter pada eptember 2019 adalah tentang rencana pengesahan RKUHP (Rancangan Kitab Undang Undang Hukum Pidana).

Metode K-NN digunakan untuk classification atau clustering tweet berdasarkan tetangga terdekat pada data latih. Dengan begitu penelitian ini dapat digunakan untuk menentukan apakah tweet tentang rencana pengesahan RKUHP tersebut berorientasi sentimen positif atau negatif serta mengetahui presentase akurasi dari Metode K-NN dalam *classification* tweet Bahasa Indonesia.

Penelitian ini menganalisis sentimen pengguna Twitter berupa data tweet dalam Bahasa Indonesia dengan Metode K-NN (K-Nearest Neighbors). Pengujian menggunakan data tweet 3000 tweet dengan kata kunci RKUHP, dilakukan dengan pembagian 80% data latih dan 40% data uji untuk melihat tingkat akurasi dan kinerja hasil klasifikasi. Berdasarkan hasil pengujian diperoleh akurasi 72%. Penerapan tahap preprocessing steaming menggunakan algoritma Nazief & Adriani mendapatkan peningkatan akurasi dengan nilai $K=179$ menjadi 81% .

Dari penelitian ini terdapat beberapa kata yang mempengaruhi sentimen yaitu rkuhp, tolak, tunda, kpk. Hasilnya 48% pengguna twitter memilih penundaan RKUHP sedangkan 52% memilih untuk menolak RKUHP.

Kata kunci – Data mining, klasifikasi, Analisis sentimen, Algoritma K-NN

ABSTRACT

The number of internet users who access social media reaches 85.7%. Social media Twitter has 7.05% users or around 12.06 million users in Indonesia. There is a lot of information and tweets from its users about various things, one of which is government policies. The government policy that became a trending topic on Twitter in September 2019 was about the plan to ratify the RKUHP (Draft Criminal Code).

The K-NN method is used for classification or clustering of tweets based on the closest neighbors in the training data. That way, this study can determine whether the tweet about the plan to ratify the RKUHP has oriented towards positive or negative sentiments as well as an accurate indicator of the K-NN Method in the classification of Indonesian tweets.

This study analyzes the sentiment of Twitter users in the form of tweets in Indonesian with the K-NN (K-Nearest Neighbors) method. Testing using 3000 tweet data with the keyword RKUHP, was done by dividing 80% of training data and 40% of testing data to see the level of accuracy and performance of the assessment results. Based on the results of the tester, the accuracy is 72%. The application of the steaming preprocessing stage used the Nazief algorithm and the increase in the value increased the value of K = 179 to 81%.

In this study, several words affect sentiment, namely RKUHP, reject, delay, KPK. The result is that 48% of Twitter users choose to reject the RKUHP while 52% refuse to reject the RKUHP.

Keywords - Data mining, classification, sentiment analysis, K-NN Algorithm