

**PERBANDINGAN PRE-PROCESSING OPINI NETIZEN TERHADAP
RUU PKS MENGGUNAKAN ALGORITMA
NAIVE BAYES CLASSIFIER**

SKRIPSI



disusun oleh

Mia Andhara Corry

16.11.0771

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**PERBANDINGAN PRE-PROCESSING OPINI NETIZEN TERHADAP
RUU PKS MENGGUNAKAN ALGORITMA
NAIVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Mia Andhara Corry

16.11.0771

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN NETIZEN TERHADAP RUU PKS PADA
SOSIAL MEDIA TWITTER DENGAN ALGORITMA
NAIVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Mia Andhara Corry

16.11.0771

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 9 Agustus 2021

Dosen Pembimbing,

Dina Maulina, M.Kom.

NIK. 90302250

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN NETIZEN TERHADAP RUU PKS PADA SOSIAL
MEDIA TWITTER DENGAN ALGORITMA
NAIVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Mia Andhara Corry

16.11.0771

telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Agustus 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mardhiya Hayaty, S.T., M.Kom.
NIK. 190302108

Supriatin, M.Kom.
NIK. 190302239

Dina Maulina, M.Kom.
NIK. 190302250

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Agustus 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom.
NIK. 190302096

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, 8 September 2021



Mia Andhara Corry

NIM. 16.11.0771

MOTTO

”I dont belive in Nationalism, I just belive in Humanity”

”Risk! Risk anything! Care no more for the opinions of others, for those voices.
Do the hardest thing on earth for you. Act for yourself. Face the truth”

”I feel like the possibility of all these possibilities being possible it’s just another
possibility that could possibly happen”



PERSEMBAHAN

Segala puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala karunia dan kesempatan yang telah diberikan, sehingga penulis dapat menyelesaikan skripsi ini. Segala cobaan yang telah disampaikan melalui semesta telah membuktikan bahwa penulis hanya seorang manusia yang memiliki banyak kekurangan dan masih diizinkan untuk jatuh dan belajar lebih banyak lagi. Skripsi ini penulis persembahkan dengan rasa terima kasih yang besar kepada:

1. Kepada orang tua penulis terutama ibu penulis yang sudah mendukung dan merelakan segala kesenangan dan hobinya agar penulis tetap bisa melanjutkan jenjang pendidikan perguruan tinggi.
2. Ibu Dina Maulina, M.Kom selaku dosen pembimbing yang sudah penulis anggap sebagai orang tua yang telah membimbing dan memberikan arahan kepada penulis selama pengerjaan skripsi ini.
3. Keluarga dan teman-teman dekat penulis yang masih tetap sabar ketika penulis sedang dalam keadaan terburuk, terima kasih karena masih tetap sabar dan mau menunggu penulis untuk bangkit dan berdiri kembali meskipun penulis selalu hilang tanpa kabar, terima kasih untuk segala pengertiannya.
4. Kepada orang-orang yang telah membuktikan penulis bahwa penulis belum cukup mampu untuk berdiri di dunia ini. Terima kasih karena sudah memperlihatkan kerasnya dunia dan terima kasih untuk segala rasa sakit sehingga penulis bisa menapakkan kaki di dunia ini.
5. Terakhir kepada NCT Dream terutama Na Jaemin, terima kasih sudah menemani setiap tetes air mata penulis, terima kasih karena telah membuktikan kepada penulis bahwa setiap perjalanan hidup membutuhkan perjuangan dan pengorbanan, rasa sakit dan jatuh, segala senyuman dapat dengan mudah ditampakkan namun segala keterpurukan hanya bayangan yang diabaikan. Terima kasih sudah membantu penulis untuk tetap waras.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas segala karunia dan kesempatan yang telah diberikan, sehingga penulis dapat menyelesaikan Skripsi yang berjudul **“Perbandingan Pre-Processing Opini Netizen Terhadap RUU PKS Menggunakan Algoritma Naive Bayes Classifier”** dengan baik dan lancar.

Laporan Skripsi ini penulis ajukan sebagai salah satu syarat kelulusan program studi Strata satu Informatika Universitas Amikom Yogyakarta.

Pada kesempatan kali ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada Bapak M. Suyanto, Prof. M.M. sebagai Ketua Universitas AMIKOM Yogyakarta dan Bapak Hanif Al Fatta, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta dan Ibu Windha Mega Pradnya D., M.Kom selaku Ketua Jurusan Informatika S1 reguler Universitas AMIKOM Yogyakarta serta Ibu Dina Maulina, M.Kom selaku dosen pembimbing.

Penulis menyadari bahwa implementasi dan laporan Skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik serta saran yang membangun sangat penulis harapkan demi kemajuan yang lebih baik dimasa mendatang. Semoga penelitian ini dapat memberikan tambahan ilmu pengetahuan dan manfaat pada semua pihak.

Yogyakarta, 23 September 2021



Mia Andhara Corry

DAFTAR ISI

JUDUL.....	i
PERSETUJUAN.....	ii
PENGESAHAN.....	iii
PERNYATAAN.....	IV
MOTTO.....	V
PERSEMBAHAN.....	VI
KATA PENGANTAR.....	VII
DAFTAR ISI.....	VIII
DAFTAR TABEL.....	XIV
DAFTAR GAMBAR.....	XV
INTISARI.....	XVII
ABSTRACT.....	XVIII
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH.....	3
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	4
1.4.1 Maksud Penelitian.....	4
1.4.2 Tujuan Penelitian.....	4
1.5 MANFAAT PENELITIAN.....	5

1.6	METODE PENELITIAN	5
1.6.1	Metode Pengumpulan Data	6
1.6.1.1	Metode Kepustakaan	6
1.6.1.2	Metode Studi Literature.....	6
1.6.1.3	Metode Crawling Data	6
1.6.2	Metode Analisis.....	6
1.6.3	Metode Perancangan	6
1.7	SISTEMATIKA PENULISAN.....	7
BAB II LANDASAN TEORI.....		9
2.1	PENELITIAN YANG RELEVAN	9
2.2	RANCANGAN UNDANG-UNDANG.....	17
2.2.1	Rancangan Undang Undang - Penghapusan Kekerasan Seksual	17
2.3	ANALISIS	18
2.3.1	Analisis Sistem.....	19
2.3.2	Analisis Kebutuhan Sistem.....	19
2.3.2.1	Kebutuhan Fungsional.....	20
2.3.2.2	Kebutuhan Nonfungsional.....	20
2.3.3	Analisis Sentimen.....	21
2.4	MEDIA SOSIAL.....	22
2.4.1	Twitter	23
2.5	DATA MINING	23
2.5.1	Pengantar Data Mining.....	24
2.5.2	Kegunaan Data Mining.....	25
2.5.3	Jenis-Jenis Data Mining	26

2.5.3.1	Text Mining	27
2.6	PERANCANGAN SISTEM	28
2.6.1	Flowchart	29
2.7	PERANCANGAN SISTEM ANALISIS SENTIMEN	32
2.7.1	Crawling Data	32
2.7.2	Pre-processing	32
2.7.2.1	Cleansing Text	33
2.7.2.2	Tokenisasi	33
2.7.2.3	Case Folding	33
2.7.2.4	Remove Stopword	33
2.7.2.5	Stemming	34
2.7.2.6	Tweet Normalized	34
2.7.3	Labelisasi	34
2.7.4	Data Split	35
2.7.4.1	Data Latih	35
2.7.4.2	Data Uji	35
2.8	PYTHON	36
2.8.1	Sejarah Python	36
2.8.2	Fitur Python	37
2.9	SOFTWARE YANG DIGUNAKAN	38
2.9.1	JupyterLab	38
2.9.1.1	Sejarah JupyterLab	39
2.9.1.2	Fitur JupyterLab	39
2.10	PENGUJIAN SISTEM	40

2.10.1	Term Frequency-Inverse Document Frequency (TF-IDF).....	40
2.10.1.1	Term Frequency (TF)	41
2.10.1.2	Inverse Document Frequency (IDF).....	41
2.10.2	Algoritma Naive Bayes	41
2.10.2.1	Multinomial Naïve Bayes.....	42
2.11	EVALUASI SISTEM	44
2.11.1	Confusion Matrix.....	44
BAB III METODE PENELITIAN.....		47
3.1	ANALISIS KEBUTUHAN	47
3.1.1	Analisis Kebutuhan Fungsional.....	47
3.1.2	Analisis Kebutuhan Non-Fungsional	48
3.2	GENERAL ARSITERKTUR SISTEM	48
3.3	PERANCANGAN SISTEM ANALISIS SENTIMEN	50
3.3.1	Crawling Data.....	50
3.3.2	Pre-Processing	51
3.3.2.1	Cleansing Text.....	52
3.3.2.2	Tokenisasi.....	53
3.3.2.3	Case Folding.....	55
3.3.2.4	Remove Stopword	56
3.3.2.5	Stemming.....	60
3.3.2.6	Tweet Normalized	63
3.3.3	Data Split	65
3.4	ANALISIS SENTIMEN.....	66
3.4.1	Term Frequency (TF)	66

3.4.2	Klasifikasi menggunakan Algoritma Naïve Bayes.....	68
3.5	METODE PENGUJIAN.....	71
3.5.1	EVALUASI SISTEM	71
BAB IV	HASIL DAN PEMBAHASAN.....	73
4.1	IMPLEMENTASI	73
4.2	CRAWLING DATA	73
4.2.1	API Twitter.....	73
4.2.2	Membuat File Penyimpanan Data	74
4.2.3	Pengambilan Data.....	74
4.3	PRE-PROCESSING.....	75
4.3.1	Membaca File Data.....	75
4.3.2	Cleansing Text.....	76
4.3.3	Tokenisasi dan Case Folding	76
4.3.4	Remove Stopword	77
4.3.5	Stemming.....	79
4.3.6	Tweet Normalized	80
4.3.7	Menyimpan Hasil Pre-Processing	81
4.4	SENTIMENT ANALYSIS.....	81
4.4.1	Mengkonvrensi Label.....	82
4.4.2	Count Vectorize dan TF-ID.....	83
4.4.3	Spliting Data.....	84
4.4.4	Klasifikasi Data Dengan Multinomial Naïve Bayes	85
4.4.5	Word Cloud	86
4.4.6	Pie Chart	89

4.5	PENGUJIAN SISTEM.....	91
4.5.1	White Box Testing.....	91
4.6	PENGUJIAN EMPLEMENTASI PADA DATA	92
4.6.1	Menghitung Term Frequency	93
4.6.2	Klasifikasi Dengan Multinomial Naive Bayes	94
4.7	PENGHITUNGAN DENGAN MULTINOMIAL NAIVE BAYES.....	94
4.7.1	Prio Probabilitas	94
4.7.2	Conditional Probabilitas	95
4.7.4	Laplace Smoothing	96
4.7.5	Perhitungan Klasifikasi Data.....	98
4.8	HASIL EVALUASI SISTEM.....	100
	BAB V PENUTUP	106
5.1	KESIMPULANN.....	106
5.2	SARAN.....	107
	DAFTAR PUSTAKA	109
	LAMPIRAN	113

DAFTAR TABEL

Tabel 2. 1 Penelitian Yang Relevan	12
Tabel 2. 2 Simbol Flowchart Sistem	29
Tabel 2. 3 Confusion Matrix	45
Tabel 3. 1 Hasil Proses Cleansing Text.....	53
Tabel 3. 2 Hasil Tokenisasi	54
Tabel 3. 3 Hasil Case Folding	56
Tabel 3. 4 Hasil Remove Stopwords	58
Tabel 3. 5 Hasil Remove Stemming.....	63
Tabel 3. 6 Tweet Normalized	65
Tabel 3. 7 Data Kalimat.....	66
Tabel 3. 8 Kumpulan Kata Pada Kalimat.....	67
Tabel 3. 9 Penyusunan Daftar Kata dan Perhitungan.....	67
Tabel 3. 10 Contoh Kalimat Klasifikasi	68
Tabel 3. 11 Contoh Nilai Variabel Evaluasi Sistem.....	71
Tabel 4. 1 Contoh Data.....	92
Tabel 4. 2 Daftar Term dan Frekuensi.....	93
Tabel 4. 3 Hasil Perhitungan Conditional P dan Lapalce S Pada Term	97
Tabel 4. 4 Jumlah Term Pada D4	98
Tabel 4. 5 Nilai Match Making Pada D4.....	99
Tabel 4. 6 Nilai Perhitungan Conditional P dan Lapalce S Pada D4	99
Tabel 4. 7 Hasil Pengujian ke-4 Kondisi.....	101
Tabel 4. 8 Nilai Rata-Rata F-Score	103
Tabel 4. 9 Nilai Rata-Rata F1-Score	104

DAFTAR GAMBAR

Gambar 2. 1 Akar Ilmu Data Mining	26
Gambar 2. 2 Metode Pembagian Data Menjadi Dua Kelompok Data	35
Gambar 3. 1 Arsitektur Umum.....	49
Gambar 3. 2 Flowchart Crawling Data.....	51
Gambar 3. 3 Flowchart Diagram Cleansing Text.....	52
Gambar 3. 4 Flowchart Diagram Tokenisasi.....	54
Gambar 3. 5 Flowchart Case Folding.....	55
Gambar 3. 6 Flowchart Diagram Remove Stopword	57
Gambar 3. 7 Flowchart Diagram Remove Stopword Manual.....	59
Gambar 3. 8 Flowchart Proses Stemming	60
Gambar 3. 9 Flowchart Gambaran Proses Stemming Sastrawi.....	61
Gambar 3. 10 Flowchart Diagram Tweet Normalized	64
Gambar 3. 11 Flowchart Data Split.....	65
Gambar 4. 1 API Twitter	74
Gambar 4. 2 Membuat File Penyimpanan	74
Gambar 4. 3 Pengambilan dan Penyimpanan Data	75
Gambar 4. 4 Membaca File Data.....	75
Gambar 4. 5 Source Code Cleansing Text	76
Gambar 4. 6 Hasil Cleansing Text	76
Gambar 4. 7 Source Code Tokenisasi	77
Gambar 4. 8 Hasil Tokenisasi	77
Gambar 4. 9 Source Code Remove Stopword.....	78
Gambar 4. 10 Hasil Remove Stopword.....	78
Gambar 4. 11 Source Code Stemming	79
Gambar 4. 12 Hasil Stemming	79
Gambar 4. 13 Source Code Tweet Normalized.....	80
Gambar 4. 14 Hasil Tweet Normalized.....	80
Gambar 4. 15 Menyimpan Hasil Pre-Processing	81
Gambar 4. 16 Hasil Pre-Processing.....	81

Gambar 4. 17 Hasil Pelabelan Sentimen	82
Gambar 4. 18 Source Code Konvensi Polaritas dan Variabel.....	82
Gambar 4. 19 Source Code CountVectorize dan TFID.....	83
Gambar 4. 20 Bentuk Matrik ‘Hasil Akhri’	84
Gambar 4. 21 Spliting Data	84
Gambar 4. 22 Source Code dan Nilai Klasifikasi.....	85
Gambar 4. 23 Hasil Analisis Data Test	86
Gambar 4. 24 Source Code Word Cloud Semua Cuitan	87
Gambar 4. 25 Hasil Word Cloud Semua Cuitan	87
Gambar 4. 26 Source Code Word Cloud Sentimen Positif	87
Gambar 4. 27 Hasil Word Cloud Sentimen Positif	88
Gambar 4. 28 Source Code Word Cloud Sentimen Netral.....	88
Gambar 4. 29 Hasil Word Cloud Sentimen Netral.....	88
Gambar 4. 30 Source Code Word Cloud Sentimen Negatif.....	89
Gambar 4. 31 Hasil Word Cloud Sentimen Negatif.....	89
Gambar 4. 32 Source Code Pie Chart.....	90
Gambar 4. 33 Hasil Pie Chart.....	90
Gambar 4. 34 Error Pemanggilan Data	91
Gambar 4. 35 Hasil Perbandingan F1-Score	104

INTISARI

Kejahatan seksual merupakan kejahatan yang paling ditakuti, tidak hanya bagi wanita dan anak kecil pria juga dapat mengalami kejahatan ini, kejahatan ini sangat beresiko dikarenakan dapat melukai korban baik secara fisik ataupun psikis. Untuk meredam kejahatan ini pemerintah mulai membuat Rancangan Undang-Undang Penghapusan Kekerasan Seksual (RUU-PKS) namun hingga saat ini proses RUU ini masih tetap dalam tahap pengajuan rancangan.

Banyak dari masyarakat ataupun kalangan aktivis yang memberikan respon terhadap situasi ini melalui sosial media mereka, Twitter salah satunya. Sebagian dari pengguna sosial media atau Netizen menyampaikan opini mereka terhadap RUU-PKS dengan berbagai sentimen. Banyaknya cuitan menyebabkan sulitnya Netizen untuk mengetahui opini negatif, positif atau netral. Dengan analisis sentimen opini acak tersebut dapat digabungkan dan diolah menjadi data sehingga menghasilkan sebuah informasi yang lebih jelas, pada proses *sentiment analysis* terdapat sebuah tahapan *pre-processing*, pada tahapan ini setiap data teks akan dilakukan pembersihan untuk mendapatkan data yang bersih dan jelas, terdapat beberapa tahapan dalam proses ini, diantaranya *cleansing text*, *case folding*, tokenisasi, *remove stopword*, *stemming*, normalisasi, dan tahapan-tahapan lainnya. dengan teknik *machine learning* bernama *Naive Bayes* sebuah metode yang dibuat untuk mengklasifikasikan data berbentuk text dapat digunakan untuk memprediksi suatu nilai dari sekumpulan data, maka dari itu peneliti akan melakukan penelitian perbedaan *pre-processing* opini Netizen terhadap RUU-PKS dengan dengan memanfaatkan algoritma tersebut.

Hasil dari penelitian ini dengan menggunakan empat kondisi *pre-processing* dengan 2.021 data opini dari Twitter mendapatkan nilai *f1-score* tertinggi pada kondisi *pre-processing* C yang tidak melakukan tahapan *stopword* namun melakukan tahapan normalisasi sebesar 72%.

Kata Kunci: RUU-PKS, *Pre-Processing*, Sentimen Analisis, Twitter, Naive Bayes Classifier

ABSTRACT

Sexual harassment is one of the most feared crimes, not only for women and children, even men can also get sexual harassment, this crime is very risky because it can injure the victim both physically and psychologically. To reduce it, the government has started to make a Draft Law on the Elimination of Sexual Violence (RUU-PKS) but until at this time the process of this bill is still in the stage of just submitting the draft.

There is a lot of the public or activists giving some response to this situation using their social media, especially Twitter. Some of the social media users or Netizens expressing their opinions about RUU-PKS with various sentiments. But there is a lot of tweets and it makes Netizens hard to differentiate negative, positive or neutral opinions. With sentiment analysis, those random opinions can be processed into data to get more clearer information, in sentiment analysis have a step named pre-processing, in pre-processing every single data having cleansing data to get a clear data for the modeling step. Pre-processing have some step in it, like cleansing text, case folding, tokenization, stopword, stemming, normalize. using machine learning technique Naive Bayes, this method used to classify data form text can be used to predict a value from a set of data, therefore researchers trying to do some research on the differences in the pre-processing of netizens' opinions on the RUU-PKS by utilizing those algorithm.

The results of this study using four pre-processing conditions with 2,021 opinion data from Twitter getting the highest f1-score value in the pre-processing C condition which did not perform the stopword stage but performed the normalization stage of 72%.

Keyword: RUU-PKS, Pre-Processing, Sentiment Analysis, Twitter, Naive Bayes Classifier