

SKRIPSI

**ANALISIS SENTIMEN RANCANGAN UNDANG-UNDANG
TINDAK PIDANA KEKERASAN SEKSUAL/UNDANG-
UNDANG TINDAK PIDANA KEKERASAN SEKSUAL
PENGGUNA *TWITTER* MENGGUNAKAN ALGORITMA
*NAÏVE BAYES CLASSIFIER***



*Mencerdaskan dan
Memartabatkan Bangsa*

ADAM PANCA PUTRA PINARIA

1512618050

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

FAKULTAS TEKNIK


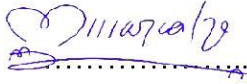
UNIVERSITAS NEGERI JAKARTA

2023

**ANALISIS SENTIMEN RANCANGAN UNDANG-UNDANG TINDAK
PIDANA KEKERASAN SEKSUAL/UNDANG-UNDANG TINDAK
PIDANA KEKERASAN SEKSUAL PENGGUNA *TWITTER*
MENGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER***

Adam Panca Putra Pinaria, 1512618050

HALAMAN PENGESAHAN

NAMA DOSEN	TANDA TANGAN	TANGGAL
Dr. Widodo, S.Kom, M.Kom (Dosen Pembimbing I)		06 Februari 2023
Murien Nugraheni, ST., M.Cs (Dosen Pembimbing II)		06 Februari 2023

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Hamdillah Ajie, S.Si, M.T (Ketua Penguji)		31 Januari 2023
Diat Nurhidayat, S.Pd., M.T.I (Dosen Penguji I)		31 Januari 2023
Ali Idrus, S.Kom, M.Kom (Dosen Penguji II)		06 Februari 2023

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa :

1. Skripsi ini merupakan karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di perguruan tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan, maka saya bersedia menerima sanksi akademik berupapencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 17 Januari 2023



Adam Panca Putra Pinaria

NIM.1512618050



**KEMENTERIAN PENDIDIKAN DAN
KEBUDAYAAN UNIVERSITAS NEGERI
JAKARTA
UPT PERPUSTAKAAN**

**Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id**

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Adam Panca Putra Pinaria
NIM : 1512618050
Fakultas/Prodi : Teknik/Pendidikan Teknik Informatika dan Komputer
Alamat email : pinaria16@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

**ANALISIS SENTIMEN RANCANGAN UNDANG-UNDANG TINDAK
PIDANA KEKERASAN SEKSUAL/UNDANG-UNDANG TINDAK PIDANA
KEKERASAN SEKSUAL PENGGUNA TWITTER MENGGUNAKAN
ALGORITMA *NAÏVE BAYES CLASSIFIER*.**

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 06 Februari 2023

Penulis

(Adam Panca Putra Pinaria)

KATA PENGANTAR

Segala puji dan syukur kehadirat Allah *subhanahu wa ta'ala* atas segala nikmat, rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi berjudul “Analisis Sentimen Rancangan Undang-Undang Tindak Pidana Kekerasan Seksual/Undang-Undang Tindak Pidana Kekerasan Seksual Pengguna Twitter Menggunakan Algoritma *Naïve Bayes Classifier*” yang merupakan persyaratan untuk meraih gelar Sarjana Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta. Shalawat serta salam selalu tercurah kepada Nabi Muhammad *shallallahu ‘alaihi wa sallam*, beserta keluarga dan para sahabatnya yang telah membimbing kita dari jalan kegelapan menuju jalan yang terang benderang.

Penulis menyadari sepenuhnya bahwa skripsi ini selesai berkat bantuan, petunjuk, saran, bimbingan dan dorongan dari berbagai pihak. Untuk itu pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Allah SWT atas nikmat, rahmat dan karunia-Nya kepada penulis,
2. Bapak Dr. Widodo, S.Kom., M. Kom., selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik Universitas Negeri Jakarta dan juga selaku dosen pembimbing pertama yang telah memberikan banyak bantuan, bimbingan, serta arahan dalam penulisan skripsi ini,
3. Ibu Murien Nugraheni, S.T, M.Cs., selaku dosen pembimbing kedua yang juga telah memberikan banyak bantuan, bimbingan, koreksi, saran, serta arahan dalam penulisan skripsi ini,
4. Bapak Hamidillah Ajie, M.T., selaku pembimbing akademik penulis yang telah memberikan saran dan bantuan selama perkuliahan,
5. Ibu Dr. Ir. Dra. Erdawaty Kamaruddin, M. Pd., selaku pembimbing Praktik Keterampilan Mengajar penulis yang telah membimbing penulis dalam penyelesaian praktik,
6. Bapak Diat Nurhidayat, M.T.I, selaku pembimbing Praktik Kerja Lapangan penulis yang telah membimbing penulis dalam penyelesaian praktik tersebut,

7. Seluruh bapak ibu dosen Pendidikan Teknik Informatika dan Komputer dan Sistem Teknologi Informasi yang telah berbagi ilmu kepada penulis,
8. Orang tua, kakak dan adik penulis yang senantiasa memberikan doa, dukungan dan motivasi kepada penulis,
9. Pihak administrasi Program Studi Pendidikan Teknik Informatika dan Komputer yang membantu penulis dalam proses penyelesaian studi,
10. Seluruh mahasiswa/i Pendidikan Teknik Informatika dan Komputer Angkatan 2018,
11. Seluruh keluarga besar warga dan alumni Asrama Mahasiswa Islam Sunan Giri yang telah memberikan bantuan tempat tinggal terjangkau dan fasilitas pendukung untuk penulis,
12. Terakhir kepada seluruh pihak yang terlibat dan/atau memberikan bantuan kepada penulis baik dalam bentuk material atau non-material sehingga penulis dapat menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, maka dari itu penulis memohon maaf apabila terdapat kekurangan dan kesalahan, baik dari isi maupun tulisan. Akhir kata, penulis berharap agar penulisan dan penyusunan skripsi ini dapat bermanfaat bagi pada pembaca dan semua pihak yang terkait.

Jakarta, 17 Januari 2023

Adam Panca Putra Pinaria
NIM.1512618050

ABSTRAK

Adam Panca Putra Pinaria. Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta, Analisis Sentimen Rancangan Undang-Undang Tindak Pidana Kekerasan Seksual/Undang-Undang Tindak Pidana Kekerasan Seksual Pengguna Twitter Menggunakan Algoritma *Naïve Bayes Classifier*. 2023. Dosen Pembimbing: (1) Dr. Widodo, S.Kom., M. Kom. (2) Murien Nugraheni, S.T., M.Cs.

Penelitian ini bertujuan untuk melakukan analisis sentimen Rancangan Undang-Undang Tindak Pidana Kekerasan Seksual/Undang-Undang Tindak Pidana Kekerasan Seksual pengguna *Twitter* menggunakan metode *Naïve Bayes Classifier* dan mengevaluasi kinerja klasifikasinya. Analisis sentimen pada penelitian ini menggunakan metode *Naïve Bayes Classifier* dengan pembobotan *Term Frequency-Inverse Document Frequency*. Proses pengumpulan data dilakukan menggunakan pustaka *Snsrape Python*. Data pakai yang digunakan untuk dibersihkan hanya yang bersentimen positif dan negatif. Proses pembersihan data melalui tahapan preproses teks yaitu *Case Folding*, *Cleansing*, *Tokenizing*, *Stopword Removal*, *Stemming* dan *Normalization*. Analisis sentimen yang dihasilkan berupa 12.807 data, dengan sentimen negatif sebanyak 10.263 data (80,1%) dan sentimen positif sebanyak 2.544 data (19,9%). Sentimen pengguna *twitter* Indonesia mengarah pada sentimen negatif. Pengujian kinerja model klasifikasi metode analisis sentimen penelitian ini menggunakan teknik *Confusion Matrix* pada hasil klasifikasi data uji dengan nilai *Accuracy* klasifikasi algoritma sebesar 93%, nilai *Precision* klasifikasi sebesar 99%, nilai *Recall* klasifikasi sebesar 87% dan *F-Measure* sebesar 93%. Dari nilai pengujian yang dihasilkan dapat disimpulkan model klasifikasi dengan metode algoritma *Naïve Bayes Classifier* memiliki kinerja yang sangat baik.

Kata kunci:

Analisis Sentimen. Naïve Bayes Classifier. Confusion Matrix. UUTPKS. Twitter.

ABSTRACT

Adam Panca Putra Pinaria. Informatics and Computer Engineering Education, Faculty of Engineering, Jakarta State University, Sentiment Analysis of the Draft Law on Sexual Violence / Sexual Violence Crime Law on Twitter Users Using the Naïve Bayes Classifier Algorithm. 2023. Supervisor: (1) Dr. Widodo, S. Kom., M. Kom. (2) Murien Nugraheni, S.T., M.Cs.

This study aims to analyze the sentiment of the Draft Law on Sexual Violence Crimes / Law on Sexual Violence Crimes on Twitter users using the Naïve Bayes Classifier method and evaluate its classification performance. Sentiment analysis in this study uses the Naïve Bayes Classifier method with Term Frequency-Inverse Document Frequency weighted. The data collection process was carried out using the Snsrape Python library. The usage data that is used for cleaning is only positive and negative sentiment. The process of cleaning the data through the stages of text preprocessing, namely Case Folding, Cleansing, Tokenizing, Stopword Removal, Stemming and Normalization. The resulting sentiment analysis is 12,807 data, with negative sentiment as much as 10,263 data (80.1%) and positive sentiment as much as 2,544 data (19.9%). The sentiment of Indonesian Twitter users is towards negative sentiment. Testing the performance of the classification model for the sentiment analysis method of this study uses the Confusion Matrix technique on the results of the classification of test data with an algorithm classification Accuracy value of 93%, a classification Precision value of 99%, a classification Recall value of 87% and an F-Measure of 93%. From the resulting test values, it can be concluded that the classification model with the Naïve Bayes Classifier algorithm method has very good performance.

Keywords:

Sentiment Analysis. Naïve Bayes Classifier. Confusion Matrix. UUTPKS. Twitter.

DAFTAR ISI

HALAMAN PENGESAHAN	ii
LEMBAR PERNYATAAN	iii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	iv
KATA PENGANTAR	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah.....	4
1.3. Batasan Masalah	4
1.4. Rumusan Masalah.....	4
1.5. Tujuan Penelitian	4
1.6. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1. <i>Text Mining</i>	6
2.2. Analisis Sentimen	6
2.3. Rancangan Undang Undang Tindak Pidana Kekerasan Seksual / Undang Undang Tindak Pidana Kekerasan Seksual (RUUTPKS/UUTPKS).....	7
2.4. <i>Twitter</i>	7
2.5. <i>Google Colaboratory</i>	8
2.6. <i>Python</i>	8
2.6.1. <i>Snsrape</i>	8
2.7. <i>Naïve Bayes Classifier</i>	9
2.8. <i>Preprocessing</i>	10
2.8.1. <i>Case Folding</i>	10
2.8.2. <i>Cleansing</i>	10
2.8.3. <i>Tokenizing</i>	10
2.8.4. <i>Stopword Removal</i>	11

2.8.5.	<i>Stemming</i>	11
2.8.6.	<i>Normalization</i>	11
2.9.	<i>Term Frequency – Inverse Document Frequency (TF – IDF)</i>	11
2.10.	<i>Confusion Matrix</i>	12
2.11.	Penelitian Relevan	13
2.12.	Kerangka Berpikir.....	18
BAB III	METODOLOGI PENELITIAN	20
3.1.	Waktu dan Tempat Penelitian.....	20
3.2.	Alat dan Bahan Penelitian.....	20
3.2.1.	Alat Penelitian.....	20
3.2.2.	Bahan Penelitian	21
3.3.	Diagram Alir Penelitian	21
3.3.1.	Pembuatan <i>Dataset</i> Penelitian	22
3.3.2.	Pengolahan Data Penelitian	23
3.3.2.1	<i>Case Folding</i>	23
3.3.2.2	<i>Cleansing</i>	25
3.3.2.3	<i>Tokenizing</i>	26
3.3.2.4	<i>Stopword Removal</i>	27
3.3.2.5	<i>Stemming</i>	28
3.3.2.6	<i>Normalization</i>	28
3.3.3.	Pengklasifikasian dengan <i>Naïve Bayes Classifier</i>	31
3.3.3.1	Klasifikasi Data Latih	31
3.3.3.2	Klasifikasi Data Uji	32
3.3.4.	Analisis Kinerja Algoritma.....	33
3.4.	Teknik dan Prosedur Pengumpulan data.....	33
3.5.	Teknik Analisis Data.....	34
BAB IV	HASIL PENELITIAN	36
4.1.	Pengumpulan Data	36
4.2.	Preproses Teks Data.....	38
4.3.	Klasifikasi Dengan <i>Naïve Bayes Classifier (NBC)</i>	48
4.4.	Pembahasan.....	50
4.5.	Aplikasi Hasil Penelitian.....	52
BAB V	KESIMPULAN DAN SARAN	53

5.1. Kesimpulan	53
5.2. Saran	53
DAFTAR PUSTAKA	54
RIWAYAT PENULIS.....	57
LAMPIRAN.....	58



DAFTAR GAMBAR

Gambar 2.1 Kerangka Berpikir	19
Gambar 3.1 Diagram Alir Penelitian	22
Gambar 3.2 Tahapan Pembuatan <i>Dataset</i> Penelitian	22
Gambar 3.3 Tahapan <i>Preprocessing</i>	23
Gambar 3.4 Tahapan Klasifikasi <i>Naïve Bayes Classifier</i>	32
Gambar 3.5 Tahapan Analisis Kinerja Algoritma	33
Gambar 4.1 Kode <i>Scrape</i> Data Dengan Kata Kunci.....	36
Gambar 4.2 Kode <i>Case Folding</i>	39
Gambar 4.3 Kode <i>Cleansing</i>	41
Gambar 4.4 Kode <i>Cleansing</i>	43
Gambar 4.5 Kode <i>Stopword Removal</i>	44
Gambar 4.6 Kode <i>Stemming</i>	46
Gambar 4.7 Kode <i>Normalization</i>	47
Gambar 4.8 Kode Klasifikasi NBC.....	49

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	13
Tabel 2.2 Penelitian Relevan.....	16
Tabel 3.1 Daftar Perangkat Keras	20
Tabel 3.2 Daftar Perangkat Lunak	21
Tabel 3.3 Contoh Hasil Proses <i>Case Folding</i>	24
Tabel 3.4 Contoh Hasil Proses <i>Cleaning</i>	25
Tabel 3.5 Contoh Hasil Proses <i>Tokenizing</i>	26
Tabel 3.6 Contoh Hasil Proses <i>Stopword Removal</i>	27
Tabel 3.7 Contoh Hasil Proses <i>Stemming</i>	28
Tabel 3.8 Contoh Hasil Proses <i>Normalization</i>	29
Tabel 3.9 Contoh Hasil Proses <i>Preprocessing</i>	29
Tabel 3.10 Perhitungan TF dan IDF	30
Tabel 3.11 Perhitungan TF-IDF.....	31
Tabel 3.12 Detail <i>Confusion Matrix</i>	34
Tabel 4.1 Sampel Data Utama Mentah	37
Tabel 4.2 Sampel Hasil <i>Case Folding</i>	38
Tabel 4.3 Sampel Hasil <i>Cleansing</i>	40
Tabel 4.4 Sampel Hasil <i>Tokenizing</i>	41
Tabel 4.5 Sampel Hasil <i>Stopword Removal</i>	43
Tabel 4.6 Sampel Hasil <i>Stemming</i>	45
Tabel 4.7 Sampel Hasil <i>Normalization</i>	46
Tabel 4.8 Sampel Hasil Klasifikasi Dengan NBC	48
Tabel 4.9 <i>Confusion Matrix</i> Hasil Klasifikasi.....	50