



**ANALISIS SENTIMEN TERHADAP
PERMENDIKBUD NOMOR 30 TAHUN 2021 TENTANG
PENCEGAHAN DAN PENANGANAN KEKERASAN
SEKSUAL DI LINGKUNGAN PERGURUAN TINGGI
MENGUNAKAN METODE NAIVE BAYES**

SKRIPSI

DANIEL LAMHOT

1810511050

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2023



**ANALISIS SENTIMEN TERHADAP
PERMENDIKBUD NOMOR 30 TAHUN 2021 TENTANG
PENCEGAHAN DAN PENANGANAN KEKERASAN
SEKSUAL DI LINGKUNGAN PERGURUAN TINGGI
MENGUNAKAN METODE NAIVE BAYES**

SKRIPSI

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

DANIEL LAMHOT

1810511050

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2023

PERNYATAAN ORISINALITAS

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

Nama : Daniel Lamhot

NIM : 1810511050

Tanggal : 1 September 2023

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

Jakarta, 1 September 2023

Yang Menyatakan,



(Daniel Lamhot)

PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertandatangan di bawah ini :

Nama : Daniel Lamhot

NIM : 1810511050

Fakultas : Ilmu Komputer

Program Studi : S1 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:

ANALISIS SENTIMEN TERHADAP PERMENDIKBUD NOMOR 30 TAHUN 2021 TENTANG PENCEGAHAN DAN PENANGANAN KEKERASAN SEKSUAL DI LINGKUNGAN PERGURUAN TINGGI MENGGUNAKAN METODE NAIVE BAYES

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 mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat Di : Jakarta

Pada tanggal : 1 September 2023

Yang Menyatakan,



(Daniel Lamhot)

LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa Skripsi berikut:

Nama : Daniel Lamhot

NIM : 1810511050

Program Studi: S1-Informatika

Judul Skripsi : Analisis Sentimen Terhadap Permendikbud Nomor 30 Tahun 2021 Tentang Pencegahan Dan Penanganan Kekerasan Seksual Di Lingkungan Perguruan Tinggi Menggunakan Metode Naïve Bayes

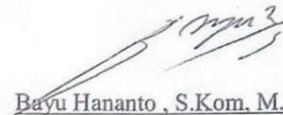
Telah berhasil dipertahankan dihadapan tim penguji dan diterima sebagai persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



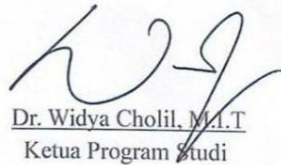
Yuni Widiastuti, S.Kom, M.Si.
Penguji Satu



Helena Nurramdhani Irmada, S.Pd, M.Kom
Penguji Dua

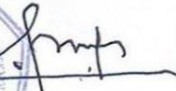


Bayu Hananto, S.Kom, M.Kom.
Pembimbing



Dr. Widya Cholil, M.I.T
Ketua Program Studi




Dr. Ermatita, M.Kom.
Dekan

Ditetapkan di : Jakarta

Tanggal Persetujuan : 17 Juli 2023



**ANALISIS SENTIMEN TERHADAP
PERMENDIKBUD NOMOR 30 TAHUN 2021 TENTANG PENCEGAHAN
DAN PENANGANAN KEKERASAN SEKSUAL DI LINGKUNGAN
PERGURUAN TINGGI MENGGUNAKAN METODE NAIVE BAYES**

DANIEL LAMHOT

ABSTRAK

Salah satu penanganan untuk kekerasan seksual di lingkungan perguruan tinggi yaitu dengan PERKEMDIKBUD Nomor 30 Tahun 2021. Hal ini dibahas disalah satu konten youtube dari channel Narasi Newsroom dengan judul “Ringkus Predator Seksual Kampus (FULL VERSION) | Mata Najwa” dan menimbulkan pro kontra antar masyarakat terhadap konten yang membahas PERKEMDIKBUD tersebut. Komentar pada konten youtube tersebut menjadi data dari penelitian ini untuk menguji klasifikasi dengan menggunakan metode Naïve Bayes. Namun karena data komentar tidak seimbang sehingga harus menggunakan metode SMOTE untuk menyeimbangkan data. Hasil dari penelitian ini terdiri dari data yang tidak seimbang, data yang seimbang dengan cara undersampling, dan data yang seimbang dengan cara SMOTE. Hasil data yang tidak seimbang memiliki akurasi 58.79%, presisi 58.10% dan recall 100%. Kemudian pada hasil data yang sudah di undersampling memiliki akurasi 61.76%, presisi 58.16% dan recall 83.82%, pada data yang sudah di undersampling akurasi mengalami peningkatan. Selanjutnya, pada data yang sudah seimbang dengan cara SMOTE memiliki akurasi 73.68%, presisi 70.14% dan recall 82.45%. Pada data yang seimbang dengan SMOTE mengalami peningkatan akurasi lagi, sehingga penelitian ketiga mendapatkan hasil akurasi terbaik yaitu 73.68%.

Kata kunci : Sentimen, *Naïve Bayes*, SMOTE

**SENTIMENT ANALYSIS OF PERMENDIKBUD NUMBER 30 OF 2021
CONCERNING PREVENTION AND MANAGEMENT OF SEXUAL
VIOLENCE IN HIGHER EDUCATION USING THE NAIVE BAYES
METHOD**

DANIEL LAMHOT

ABSTRACT

One way to handle sexual violence in a higher education environment is PERKEMDIKBUD Number 30 of 2021. This is discussed in one of the YouTube content from the Narasi Newsroom channel with the title "Ringkus Campus Sexual Predators (FULL VERSION) | Mata Najwa" and raises pros and cons among the community towards the content that discusses PERKEMDIKBUD. Comments on the YouTube content become data from this study to test classification using the Naïve Bayes method. However, because the comment data is unbalanced, the SMOTE method must be used to balance the data. The results of this study consist of unbalanced data, balanced data by undersampling, and balanced data by SMOTE. Unbalanced data results have 58.79% accuracy, 58.10% precision and 100% recall. Then the results of the undersampled data have an accuracy of 61.76%, 58.16% precision and 83.82% recall, the accuracy of the undersampled data has increased. Furthermore, the data that has been balanced by means of SMOTE has an accuracy of 73.68%, 70.14% precision and 82.45% recall. The data that is balanced with SMOTE has increased accuracy again, so that the third study obtained the best accuracy results, namely 73.68%.

Keywords : *Sentiment, Naïve Bayes, SMOTE*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas segala anugerah-Nya sehingga proposal ini dapat terselesaikan. Judul dari penelitian ini yang dilaksanakan sejak bulan November 2021 ini adalah “ANALISIS SENTIMEN TERHADAP PERMENDIKBUD NOMOR 30 TAHUN 2021 TENTANG PENCEGAHAN DAN PENANGANAN KEKERASAN SEKSUAL DI LINGKUNGAN PERGURUAN TINGGI MENGGUNAKAN METODE NAIVE BAYES” berhasil diselesaikan. Tak lupa penulis ingin mengucapkan banyak terima kasih kepada:

1. Orang tua, keluarga yang selalu memberikan dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Bayu Hananto, S.Kom, M.Kom selaku dosen pembimbing yang telah memberikan saran yang bermanfaat selama proses pembuatan Proposal hingga menyelesaikan skripsi.
3. Bapak Henki Bayu Seta, S.Kom, MTI. selaku dosen pembimbing akademik.
4. Ibu Dr. Widya Cholil, S.Kom., M.I.T. selaku Kaprodi Informatika yang telah memberikan informasi mengenai tugas akhir.
5. Ibu Dr. Ermatita, M.Kom. selaku dekan Fakultas Ilmu Komputer.
6. Teman-teman Informatika 2018 yang telah berjuang bersama dalam setiap proses perkuliahan serta saling memberikan semangat untuk dapat menyelesaikan Skripsi.

Dan semua pihak yang telah membantu penulisan dalam menyelesaikan skripsi ini.

Jakarta, 31 Desember 2022

Daniel Lamhot

DAFTAR ISI

PERNYATAAN ORISINALITAS	ii
PERNYATAAN PERSETUJUAN PUBLIKASI	iii
LEMBAR PENGESAHAN	iv
ABSTRAK	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian.....	3
1.4. Batasan Masalah.....	4
1.5. Manfaat Penelitian.....	4
1.6. Luaran Yang Diharapkan.....	4
1.7. Sistematis Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1. Youtube.....	6
2.2. PERMENDIKBUD Nomor 30 Tahun 2021	6
2.3. Analisis Sentimen.....	6
2.4. Naive Bayes	7
2.5. Text Mining.....	8
2.6. Term Frequency-Inverse Document Frequency (TF-IDF)	8

2.7.	Evaluasi.....	8
2.1.	Penelitian Terdahulu	10
BAB III METODE PENELITIAN		13
3.1.	Kerangka Pikir.....	13
3.1.1.	Identifikasi Masalah.....	13
3.1.2.	Studi Literatur	14
3.1.3.	Pengumpulan Data	14
3.1.4.	Praproses Data	14
3.1.5.	Term Frequency-Inverse Document Frequency (TF-IDF)	15
3.1.6.	Klasifikasi Naive Bayes	16
3.1.7.	Evaluasi	16
3.1.8.	Analisis Hasil.....	16
3.2.	Perangkat Penelitian	16
3.3.	Jadwal Penelitian	17
BAB IV		19
4.1.	Data.....	19
4.2.	Pelabelan Data.....	19
4.3.	Praproses Data.....	21
4.1.1.	Data cleaning	22
4.1.2.	Case folding.....	22
4.1.3.	Spelling normalization	23
4.1.4.	Tokenizing.....	24
4.1.5.	Filtering	25
4.1.6.	Stemming.....	25
4.4.	Term Frequency - Inverse Document Frequency	26
4.5.	Word Cloud data fitur	28
4.5.1.	Word cloud data positif	28
4.5.2.	Word cloud data negatif	29
4.6.	SMOTE (Synthetic Minority Oversampling Technique)	30
4.7.	Klasifikasi Naive Bayes.....	30
4.7.1.	Proses Latih.....	31

4.7.2. Proses Uji	33
4.8. Evaluasi.....	35
4.9. Analisis Perbandingan Hasil	35
BAB V	36
Kesimpulan.....	36
Saran.....	37
DAFTAR PUSTAKA	39
RIWAYAT HIDUP	40
LAMPIRAN	41

DAFTAR GAMBAR

Gambar 3. 1 Kerangka Pikir	12
Gambar 4. 1 Label Sentimen	20
Gambar 4. 2 Bagan Alur Data Cleaning	21
Gambar 4. 3 Bagan Alur Case Folding	22
Gambar 4. 4 Bagan Alur Spelling Normalization	23
Gambar 4. 5 Bagan Alur Tokenizing	23
Gambar 4. 6 Bagan Alur Filtering	24
Gambar 4. 7 Bagan Alur Stemming	25
Gambar 4. 8 Word Cloud Label Positif	27
Gambar 4. 9 Diagram Batang Word Cloud Label Positif	28
Gambar 4. 10 Word Cloud Label Negatif	29
Gambar 4. 11 Diagram Batang Word Cloud Label Negatif	29

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	9
Tabel 2. 2 Penelitian Sebelumnya	10
Tabel 3. 1 Jadwal Penelitian	16
Tabel 4. 1 Sampel Data Ulasan	18
Tabel 4. 2 Hasil Pelabelan Data	19
Tabel 4. 3 Sampel Data Cleaning	21
Tabel 4. 4 Sampel Case Folding	22
Tabel 4. 5 Sampel Spelling Normalization	23
Tabel 4. 6 Sampel Tokenizing	24
Tabel 4. 7 Sampel Filtering	24
Tabel 4. 8 Sampel Stemming	25
Tabel 4. 9 Data Komentar	25
Tabel 4. 10 Data TFIDF	26
Tabel 4. 11 Bobot Kata	30
Tabel 4. 12 Probabilitas Setiap Kelas	32
Tabel 4. 13 Data Sampel Uji	32
Tabel 4. 14 Data Bobot Data Uji	33
Tabel 4. 15 Evaluasi Confusion Matrix Data SMOTE	34
Tabel 4. 16 Perbandingan Hasil Penelitian	35

DAFTAR LAMPIRAN

Lampiran 1 Daftar Stopword	41
Lampiran 2 Data Kamus Normalisasi	45