

**TUGAS AKHIR**

**ANALISIS SENTIMEN TOPIK KORUPSI DI INDONESIA PADA  
SITUS X MENGGUNAKAN ALGORITMA BERNOULLI NAIVE  
BAYES**



**SEPTIAN DWI WICAKSONO**

**1910651053**

**PROGAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

**2024**

**TUGAS AKHIR**

**ANALISIS SENTIMEN TOPIK KORUPSI DI INDONESIA PADA  
SITUS X MENGGUNAKAN ALGORITMA BERNOULLI NAIVE  
BAYES**

Disusun Untuk Melengkapi Dan Memenuhi Syarat Kelulusan Guna Meraih Gelar  
Sarjana Komputer

Teknik Informatika Universitas Muhammadiyah Jember



**SEPTIAN DWI WICAKSONO**

**1910651053**

**PROGAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

**2024**

## SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Septian Dwi Wicaksono  
NIM : 1910651053  
Program Studi : Teknik Informatika  
Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul “*Analisis Sentimen Topik Korupsi di Indonesia Pada Situs X Menggunakan Algoritma Bernoulli Naïve Bayes*” bukan merupakan tugas akhir orang lain baik sebagian ataupun keseluruhan kecuali dalam bentuk kutipan yang disebutkan sumbernya pada tugas akhir ini.

Demikian surat pernyataan ini dibuat dengan sebesar – besarnya dan apabila pernyataan ini tidak sesuai, maka penulis bersedia mendapatkan sanksi akademik.

Jember, 22 Juli 2024



**Septian Dwi Wicaksono**  
NIM. 1910651053

**HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR**

**ANALISIS SENTIMEN TOPIK KORUPSI DI INDONESIA PADA  
SITUS X MENGGUNAKAN ALGORITMA BERNOULLI NAIVE  
BAYES**

**Oleh :**

**SEPTIAN DWI WICAKSONO**

**1910651053**

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

di

Universitas Muhammadiyah Jember

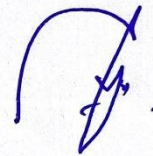
**Disetujui oleh,**

**Pembimbing I**

**Pembimbing II**



**Syarif Hidayatullah, S.Kom., M.Kom**  
NIDN. 0713049203



**Miftahur Rahman, S.Kom., M.Kom**  
NIDN. 0724039201

**HALAMAN PENGESAHAN**


**ANALISIS SENTIMEN TOPIK KORUPSI DI INDONESIA  
PADA SITUS X MENGGUNAKAN ALGORITMA BERNOULLI  
NAIVE BAYES**

Oleh :  
**SEPTIAN DWI WICAKSONO**

**1910651053**

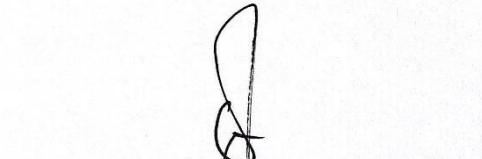
Telah mempertanggung jawabkan Proposal Tugas Akhirnya pada sidang tugas Akhir tanggal 22 Juli 2024 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom) di Universitas Muhammadiyah Jember  
**Disetujui oleh,**

**Penguji I**




**Guruh Wijaya, S.T., M.Kom**  
NIDN. 0729017601

**Pembimbing I**



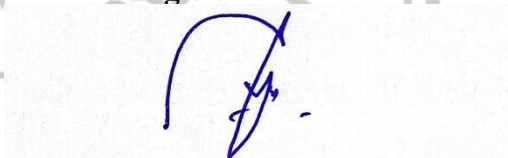
**Svarif Hidayatullah, S.Kom., M.Kom**  
NIDN. 0723049203

**Penguji II**



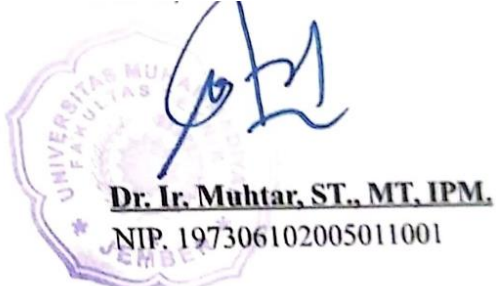

**Ilham Sajfudin, S.pd., M.Si**  
NIDN. 0731108903

**Pembimbing II**



**Miftahur Rahman, S.Kom., M.Kom**  
NIDN. 0724039201

**Mengesahkan,  
Dekan Fakultas Teknik**



**Dr. Ir. Muhtar, ST., MT, IPM.**  
NIP. 197306102005011001

**Mengetahui:**



**Ketua Program Studi  
Teknik Informatika**  
**Rosita Yanuarti, S.Kom., M.Cs**  
NIDN. 0629018601

## MOTTO

“Jika seseorang bepergian dengan tujuan menuntut ilmu, maka Allah akan menjadikan perjalanannya seperti perjalanan menuju surga”

*(Nabi Muhammad SAW)*

“Saat engkau bermalas-malasan sebaiknya ingatlah bahwa jutaan pesaingmu sedang berusaha mengalahkanmu”

*(Septian Dwi W)*



## PERSEMBAHAN

Alhamdulillah segala puji bagi Allah SWT yang telah melimpahkan rahmat serta hidayah yang luar biasa, sehingga penulis dapat menyelesaikan tugas akhir. Maka dari itu pada kesempatan kali ini penulis menyampaikan rasa ucapan terimakasih yang sedalam-dalamnya diberikan kepada:

1. Allah SWT berkat segala rida, rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan segala urusan dalam penyusunan Tugas Akhir dan diberikan kesempatan mendapatkan gelar Sarjana Komputer.
2. Ibunda Sri Endang Yulianti dan Ayahanda Banggawak Soedi Hartono tercinta serta Kakak saya Raditya Arif Wibisono yang selalu mendoakan dan memberi nasehat serta semangat yang sangat berarti bagi penulis, semoga Allah SWT membalasnya dan memberi barokah serta hidayah.
3. Dr. Ir. Muhtar, S.T, M.T, IPM. Selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
4. Ibu Rosita Yanuarti, S.Kom., M.Cs. Selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
5. Bapak Syarif Hidayatullah, S.Kom., M.Kom. Selaku dosen pembimbing I yang telah memberikan arahan dan membimbing penulis dalam menyelesaikan tugas akhir ini.
6. Bapak Miftahur Rahman, S.Kom., M.Kom. Selaku dosen pembimbing II yang telah memberikan arahan dan membimbing penulis dalam menyelesaikan tugas akhir ini.
7. Bapak Guruh Wijaya, S.T., M.Kom. Selaku dosen penguji I dan Bapak Ilham Saifudin, S.Pd., M.Si. Selaku dosen penguji II yang telah memberikan saran dan masukan dalam penyelesaian tugas akhir ini.
8. Kekasih saya Nadira Fadia Risanti yang telah memberikan dukungan, doa dan masukan selama penulisan Tugas Akhir ini.
9. Teman seperjuangan Teknik Informatika Angkatan 19 dan para kakak Tingkat yang senantiasa mendukung saya dalam Menyusun Tugas Akhir ini

## ABSTRAK

# ANALISIS SENTIMEN TOPIK KORUPSI DI INDONESIA PADA SITUS X MENGGUNAKAN ALGORITMA BERNOULLI NAIVE BAYES

Septian Dwi Wicaksono<sup>1</sup>, Syarif Hidayatullaah<sup>2</sup>, Miftahur Rahman<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : tianfalconers46@gmail.com

Indonesia merupakan negara dengan jumlah penduduk yang banyak sehingga penggunaan internet berkembang sangat pesat. Twitter merupakan salah satu media sosial yang populer di Indonesia maupun di berbagai negara lainnya. Hal ini membuat Twitter potensial sebagai sumber data yang efisien untuk analisis sentimen. Analisis sentimen adalah studi komputasi yang bertujuan untuk memahami, mengekstrak, dan memproses data teks untuk mendapatkan informasi tentang sentimen dalam suatu kalimat. Dari proses penelitian yang dilakukan tentang identifikasi kasus korupsi di Indonesia, maka hasil yang didapat dari klasifikasi komentar masyarakat dibagi dalam dua kelas yaitu positif dan negatif. Penerapan algoritma *Bernoulli Naive Bayes* dalam konteks ini relevan karena model ini mampu mengatasi kelemahan data teks yang sparse, dimana setiap dokumen dianggap sebagai sekumpulan kata yang diwakili oleh variabel biner yang menjelaskan proses pengumpulan data yang diambil dari Twitter untuk tujuan analisis sentimen terhadap topik korupsi. Hasil dari percobaan data dengan menggunakan percobaan pertama menggunakan perbandingan data latih 10% dan data uji 90% dengan nilai *precision* 83%, *recall* 91% dan *f1-score* 87% dari label negatif, sedangkan nilai *precision* 71%, *recall* 64% dan *f1-score* 71% dari label positif, menghasilkan akurasi sebesar 81,88%. Percobaan kedua menggunakan perbandingan data latih 85% dan data uji 15%, menghasilkan akurasi sebesar 77% dengan nilai *precision* 80%, *recall* 89% dan *f1-score* 84% dari label negatif, sedangkan nilai *precision* 72%, *recall* 55% dan *f1-score* 62% dari label positif. Percobaan ketiga data latih 30% dan 70% data uji menghasilkan akurasi 75% dengan nilai *precision* 79%, *recall* 87% dan *f1-score* 83% dari label negatif, sedangkan nilai *precision* 68%, *recall* 55% dan *f1-score* 61% dari label positif. Dari ketiga percobaan yang telah di uji dapat disimpulkan bahwa dengan data latih 10% dan data uji 90% menghasilkan akurasi yang lebih tinggi.

**Kata Kunci:** *Twitter, Analisis Sentimen, Korupsi, Bernoulli Naive Bayes*



## **ABSTRACT**

### **SENTIMENT ANALYSIS OF CORRUPTION TOPICS IN INDONESIA ON WEBSITE X USING BERNOULLI NAIVE BAYES ALGORITHM**

Septian Dwi Wicaksono<sup>1</sup>, Syarif Hidayatullaah<sup>2</sup>, Miftahur Rahman<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : tianfalconers46@gmail.com

*Indonesia is a country with a large population so internet usage is growing very rapidly. Twitter is one of the most popular social media in Indonesia and other countries. This makes Twitter a potential source of efficient data for sentiment analysis. Sentiment analysis is a computational study that aims to understand, extract, and process text data to obtain information about the sentiment in a sentence. From the research process conducted on the identification of corruption cases in Indonesia, the results obtained from the classification of public comments are divided into two classes, namely positive and negative. The application of the Bernoulli Naive Bayes algorithm in this context is relevant because this model is able to overcome the weaknesses of sparse text data. where each document is considered as a set of words represented by binary variables that explain the process of collecting data taken from Twitter for the purpose of sentiment analysis on the topic of corruption. Results from data experiments using the first experiment using a 10% training data ratio and 90% test data with a precision value of 83%, recall 91% and f1-score 87% of negative labels, while the precision value is 71%, recall 64% and f1-score 71% of positive labels, resulting in an accuracy of 81.88%. The second experiment used a ratio of 85% training data and 15% test data, resulting in an accuracy of 77% with a precision value of 80%, recall 89% and f1-score 84% from negative labels, while the precision value was 72%, recall 55% and f1-score 62% from positive labels. The third trial of 30% training data and 70% test data resulted in 75% accuracy with 79% precision, 87% recall and 83% f1-score from negative labels, while 68% precision, 55% recall and 61% f1-score from positive labels. From the three experiments that have been tested, it can be concluded that with 10% training data and 90% test data produces higher accuracy.*

**Keywords:** *Twitter, Sentiment Analysis, Corruption, Bernoulli Naive Bayes*

## **KATA PENGANTAR**

Puji Syukur kehadiran Allah SWT. yang telah melimpahkan rahmat serta hidayah dan karunia-Nya pada penulis, sehingga penulis dapat menyelesaikan penulisan Tugas Akhir yang berjudul “*ANALISIS SENTIMEN TOPIK KORUPSI DI INDONESIA PADA SITUS X MENGGUNAKAN ALGORITMA BERNOULLI NAÏVE BAYES*”. Tugas Akhir ini ditulis dalam rangka memenuhi syarat untuk mencapai gelar sarjana Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Penulis sangat berterima kasih kepada Bapak Syarif Hidayatullah, S.Kom., M.Kom. selaku pembimbing pertama dan Bapak Miftahur Rahman, S.Kom., M.Kom. selaku pembimbing kedua, atas segala perhatian dan bimbingannya serta arahan-arahan yang diberikan kepada penulis dalam upaya menyelesaikan tugas akhir ini.

Ucapan terima kasih yang sebesar-besarnya penulis sampaikan kepada Dr. Ir. MUHTAR, S.T, M.T, IPM. selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember, atas kesediannya penulis belajar Fakultas Teknik, Program Studi Teknik Informatika, Universitas Muhammadiyah Jember.

Penulis menyadari bahwa Tugas Akhir ini jauh dari kata sempurna, baik menyangkut aspek penulisan maupun materi. Untuk itu penulis memohon maaf yang sebesar-besarnya. Tugas Akhir ini diajukan guna memenuhi ujian akhir Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Demikian kata pengantar dari penulis, Semoga bermanfaat bagi pembaca

# DAFTAR ISI

<b>SURAT PERNYATAAN</b> .....	ii
<b>HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR</b> .....	iii
<b>HALAMAN PENGESAHAN</b> .....	iv
<b>MOTTO</b> .....	v
<b>PERSEMBAHAN</b> .....	vi
<b>ABSTRAK</b> .....	vii
<b>ABSTRACT</b> .....	viii
<b>KATA PENGANTAR</b> .....	ix
<b>DAFTAR ISI</b> .....	x
<b>BAB 1. PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan penelitian .....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
<b>BAB 2. LANDASAN TEORI</b> .....	4
2.1 Korupsi Di Indonesia.....	4
2.2 Twitter X .....	4
2.3 Analisis Sentimen.....	4
2.4 Text Mining .....	5
2.5 Preprocessing .....	5
2.5.1 <i>Case Folding</i> .....	5
2.5.2 <i>Stopword</i> .....	6
2.5.3 <i>Stemming</i> .....	6
2.5.4 <i>Tokenizing</i> .....	6
2.6 TF-IDF.....	6
2.7 Algoritma Bernoulli Naïve Bayes .....	7
2.8 Penerapan Bernoulli Naïve Bayes pada Analisis Sentimen Korupsi .....	8
2.9 Confusion Matrik .....	9
2.10 Phyton.....	10

2.11	Penelitian Terdahulu .....	10
<b>BAB 3. METODE PENELITIAN .....</b>		<b>13</b>
3.1	Tahapan Penelitian .....	13
3.2.	Pelabelan data.....	14
3.3.	Pengumpulan Data .....	14
3.4.	Preprocessing Data .....	15
3.4.1	<i>Cleaning</i> .....	15
3.4.2	<i>Case Folding</i> .....	16
3.4.3	<i>Stopword</i> .....	17
3.4.4	<i>Stemming</i> .....	17
3.4.5	<i>Tokenizing</i> .....	18
3.5.	TF-IDF.....	19
3.6.	Algoritma Bernoulli Naive Bayes .....	29
3.7.	Evaluasi .....	36
3.7.1	<i>Confusion Matrix</i> .....	36
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>39</b>
4.1.	Gambaran Data.....	39
4.2.	Pelabelan Data.....	40
4.3.	Pengujian dan Pembagian Normal Dataset .....	47
<b>BAB V. KESIMPULAN DAN SARAN .....</b>		<b>51</b>
5.1	Kesimpulan.....	51
5.2.	Saran.....	51
<b>LAMPIRAN VALIDASI DATA PENELITIAN.....</b>		<b>52</b>
<b>DAFTAR PUSTAKA .....</b>		<b>53</b>