

**PERBANDINGAN ALGORITMA *NAIVE BAYES* DAN
SVM PADA ANALISIS SENTIMEN KOMENTAR
PENGGUNA TWITTER (X) TERHADAP
KORUPSI PADA PT PERTAMINA**

SKRIPSI

**Oleh:
Farah Nabila
202110715200**



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2025**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Skripsi : Perbandingan Algoritma *Naïve Bayes* dan *SVM*
pada Analisis Sentimen Komentar Pengguna
Twitter Terhadap Korupsi pada PT Pertamina
Nama Mahasiswa : Farah Nabila
Nomor Pokok Mahasiswa : 202110715200
Program Studi/Fakultas : Informatika/Illmu Komputer
Tanggal Lulus Ujian Skripsi : 25 Juli 2025

Jakarta, 01 Agustus 2025

MENYETUJUI,

Pembimbing I

Pembimbing II



Prima Dina Atika, S.Kom., M.Kom.
NIDN: 0311037107



Khairunnisa Fadhillah Ramdhania, S.Si., M.Si.
NIDN: 0328039201

Ketua Program Studi



Ahmad Fathurrozi, S.E., M.M.S.I
NIP: 2012486

Program Studi Informatika
Fakultas Ilmu Komputer
Universitas Bhayangkara Jakarta Raya
2025

LEMBAR PENGESAHAN

Judul Skripsi : Perbandingan Algoritma *Naive Bayes* dan *SVM*
pada Analisis Sentimen Komentar Pengguna
Twitter (X) Terhadap Korupsi pada PT Pertamina

Nama Mahasiswa : Farah Nabila

Nomor Pokok Mahasiswa : 202110715200

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 25 Juli 2025

Jakarta, 01 Agustus 2025

MENGESAHKAN,

Ketua Tim Penguji : Adi Muhajirin, S.Kom., M.Kom., M.M.
NIDN : 0318038501

Penguji I : Allan D. Alexander, S.T., M.Kom.
NIDN : 0305127404

Penguji II : Prima Dina Atika, S.Kom., M.Kom.
NIDN : 0311037107

MENGETAHUI,

Ketua
Program Studi Informatika

Dekan
Fakultas Ilmu Komputer

Ahmad Fathurrozi, S.E., M.M.S.I.
NIP. 2012486

Dr. Robertus Suraji, S.S., M.A.
NIP. 1904406



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Farah Nabila
NPM : 202110715200
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Skripsi : Perbandingan Algoritma *Naïve Bayes* dan *SVM* pada Analisis Sentimen Komentar Pengguna Twitter (X) Terhadap Korupsi pada PT Pertamina

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak mana pun.

Bekasi, 01 Agustus 2025

Penulis



Farah Nabila

ABSTRAK

Farah Nabila. 202110715200. Perbandingan Algoritma *Naïve Bayes* dan *SVM* Pada Analisis Sentimen Komentar Pengguna Twitter (X) Terhadap Korupsi Pada PT Pertamina. Bekasi: Fakultas Ilmu Komputer. Universitas Bhayangkara Jakarta Raya. 2025.

Korupsi merupakan salah satu permasalahan serius yang merusak tatanan sosial, politik, dan ekonomi di Indonesia. PT Pertamina sebagai salah satu perusahaan BUMN strategis turut terseret dalam berbagai kasus korupsi, yang memicu respons luas dari masyarakat di media sosial, khususnya Twitter (X). Penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan sentimen komentar publik terkait kasus korupsi di PT Pertamina. Data yang digunakan dikumpulkan melalui metode *web scraping* dari Twitter (X) pada periode Februari–Maret 2025, menghasilkan 1.946 tweet, dengan 1.887 data bersih setelah proses *cleaning*. Sentimen diklasifikasikan ke dalam tiga kategori: positif, netral, dan negatif. Penilaian performa juga dilakukan dengan skenario pembagian data latih dan uji yaitu: 80:20. Proses analisis melibatkan tahapan *preprocessing*, pembobotan TF-IDF dan juga penyeimbangan data menggunakan *SMOTE*, serta evaluasi model dengan *Confusion Matrix*. Hasil pengujian menunjukkan bahwa algoritma *SVM* memperoleh akurasi sebesar 78%, lebih tinggi dibandingkan *Naïve Bayes* yang mencapai 64%. Penelitian ini menunjukkan bahwa *SVM* lebih efektif dalam mengklasifikasikan opini publik pada kasus korupsi PT Pertamina.

Kata Kunci: Analisis Sentimen, Twitter (X), *Naïve Bayes*, *Support Vector Machine*, Korupsi, PT Pertamina

ABSTRACT

Farah Nabila. 202110715200. *Comparison of Naïve Bayes and SVM Algorithms on Sentiment Analysis of Twitter (X) User Comments on Corruption at PT Pertamina. Bekasi: Faculty Of Computer Science. University Of Bhayangkara Jakarta Raya. 2025.*

Corruption is one of the serious problems that damage the social, political and economic order in Indonesia. PT Pertamina as one of the strategic state-owned companies has also been dragged into various corruption cases, which triggered a wide response from the public on social media, especially Twitter (X). This study aims to analyze and compare the performance of Naïve Bayes and Support Vector Machine (SVM) algorithms in classifying the sentiment of public comments related to corruption cases at PT Pertamina. The data used was collected through web scraping method from Twitter (X) in the period February-March 2025, resulting in 1,946 tweets, with 1,887 clean data after cleaning process. Sentiments were classified into three categories: positive, neutral, and negative. Performance assessment was also carried out with a training and test data split scenario of: 80:20. The analysis process involved preprocessing, TF-IDF weighting as well as data balancing using SMOTE, and model evaluation using Confusion Matrix. The test results showed that the SVM algorithm obtained an accuracy of 78%, higher than Naïve Bayes which reached 64%. This research shows that SVM is more effective in classifying public opinion in the PT Pertamina corruption case.

Keywords: *Sentiment Analysis, Twitter (X), Naïve Bayes, Support Vector Machine, Corruption, PT Pertamina*

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Farah Nabila
NPM : 202110715200
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*)**, atas karya ilmiah saya yang berjudul :

“Perbandingan Algoritma *Naïve Bayes* dan *SVM* Pada Analisis Sentimen Komentar Pengguna Twitter (X) Terhadap Korupsi Pada PT Pertamina”

Beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi
Pada tanggal : 01 Agustus 2025

Yang Menyatakan



Farah Nabila

KATA PENGANTAR

Dengan memanjatkan puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayahnya sehingga skripsi dengan judul **"Perbandingan Algoritma *Naïve Bayes* dan *SVM* Pada Analisis Sentimen Komentar Pengguna Twitter (X) Terhadap Korupsi Pada PT Pertamina"** ini dapat terselesaikan. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.

Skripsi ini tidak mungkin dapat terselesaikan tanpa adanya bantuan, dukungan, bimbingan, dan juga saran-saran dari berbagai pihak. Oleh karena itu, diucapkannya kata terima kasih yang sebesar-besarnya kepada kedua Orang Tua yang selalu memberikan semangat dan dukungan serta kepada:

1. Bapak Irjen Polisi (Purn) Prof. Dr. Drs. H. Bambang Karsono, S.H., M.M., Ph.D., D.Crim.(Hc) selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Bapak Dr. Robertus Suraji, S.S., M.A. selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. selaku Ketua Program Studi Informatika.
4. Ibu Prima Dina Atika, S.Kom., M.Kom dan juga Ibu Khairunnisa Fadhillah Ramdhania, S.Si., M.Si. Selaku dosen pembimbing I dan 2 dalam penyusunan laporan skripsi.

5. Bapak Agus Hidayat, S.Kom., M.Si. Selaku dosen pembimbing akademik, atas bimbingan, nasihat, dan arahan selama masa studi hingga penyusunan tugas akhir ini.
6. Kepada kedua orang tua, kakak dan adik yang selalu memberikan dukungan dan semangat serta mendoakan setiap harinya agar diberikan kemudahan dalam menyelesaikan laporan skripsi ini.
7. Terima kasih kepada teman-teman seperjuangan terutama “Princess A5” yang selalu memberikan dukungan dan selalu membantu dalam menyelesaikan laporan skripsi ini.

Disadari bahwa skripsi ini jauh dari kesempurnaan, untuk itu mohon maaf apabila terdapat kekurangan dan kesalahan baik dari pembahasan maupun penulisan. Oleh karena itu, diharapkan masukan berupa kritik dan saran yang membangun.

Jakarta, 01 Agustus 2025

Hormat Saya



Farah Nabila
202110715200

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	6
1.3 Rumusan Masalah.....	6
1.4 Batasan Masalah.....	6
1.5 Tujuan Penelitian	7
1.6 Manfaat Penelitian	8
1.7 Sistematika Penulisan	9
BAB II TINJAUAN PUSTAKA.....	11
2.1 Penelitian Terkait	11
2.1.1 Tabel <i>Related Reseach / State Of The Art</i>	15
2.2 Teori Pendukung.....	18
2.2.1 Analisis Sentimen	18
2.2.2 Media Sosial Twitter (X)	19
2.2.3 Data Mining	20

2.2.4	<i>Text Mining</i>	22
2.2.5	<i>Data Collection</i>	23
2.2.6	<i>Preprocessing</i>	24
2.2.7	<i>Knowledge Discovery In Database (KDD)</i>	27
2.2.8	Pembobotan Kata	29
2.2.9	<i>Machine Learning</i>	31
2.2.10	Algoritma <i>Naïve Bayes</i>	32
2.2.11	Algoritma <i>Support Vector Machine (SVM)</i>	34
2.2.12	<i>Confusion Matrix</i>	36
2.2.13	<i>Lexicon Based</i>	39
2.2.14	<i>SMOTE</i>	40
2.2.15	<i>Python</i>	40
BAB III METODOLOGI PENELITIAN		42
3.1	Objek penelitian	42
3.2	Kerangka Penelitian	43
3.3	Metode Pengumpulan Data	47
3.4	<i>Data Cleaning</i>	52
3.5	<i>Preprocessing</i>	54
3.5.1	<i>Case Folding</i>	54
3.5.2	<i>Cleaning Data</i>	55
3.5.3	<i>Tokenizing</i>	57
3.5.4	<i>Normalization</i>	58
3.5.5	<i>Filtering/Stopword Removal</i>	59
3.5.6	<i>Stemming</i>	60
3.6	<i>Labeling</i>	62
3.7	<i>Transformation</i>	64
3.8	<i>Splitting Data</i>	69
3.9	Analisis Kebutuhan Sistem	71
BAB IV HASIL DAN PEMBAHASAN		72

4.1	<i>Data Mining</i>	72
4.1.1	<i>Naïve Bayes</i>	72
4.1.2	<i>Support Vector Machine</i>	79
4.2	<i>Evaluation</i>	82
4.2.1	Evaluasi <i>Naïve Bayes</i>	83
4.2.2	Evaluasi <i>SVM</i>	85
4.3	<i>WordCloud</i>	88
4.3.1	<i>WordCloud Naïve Bayes</i>	88
4.3.2	<i>WordCloud SVM</i>	91
BAB V PENUTUP		94
5.1	Kesimpulan	94
5.2	Saran.....	95
DAFTAR PUSTAKA		97
LAMPIRAN		102

DAFTAR TABEL

Tabel 2. 1 Tabel State Of The Art.....	15
Tabel 3. 1 Hasil Tahap <i>Case Folding</i>	55
Tabel 3. 2 Hasil Tahap <i>Cleaning</i>	56
Tabel 3. 3 Hasil Tahap <i>Tokenizing</i>	57
Tabel 3. 4 Hasil Tahap <i>Normalization</i>	58
Tabel 3. 5 Hasil Tahap <i>Filtering/Stopword Removal</i>	60
Tabel 3. 6 Hasil Tahap <i>Stemming</i>	61
Tabel 3. 7 Contoh Dataset Perhitungan Manual.....	65
Tabel 3. 8 Perhitungan Manual <i>TF-IDF</i>	65
Tabel 4. 1 Tabel Sampel <i>Data Train</i>	74
Tabel 4. 2 Tabel Sampel <i>Data Test</i>	74
Tabel 4. 3 Penentuan Prediksi dan Aktual.....	74
Tabel 4. 4 Data Untuk Perhitungan <i>SVM</i>	81
Tabel 4. 5 Perbandingan <i>Confusion Matrix SVM</i> dan <i>Naïve Bayes</i>	87

DAFTAR GAMBAR

Gambar 1. 1 Posisi Indonesia Dalam Negara Terkorup di Asia Tenggara	2
Gambar 1. 2 Kasus Korupsi di Lingkungan Pertamina Indonesia	3
Gambar 2. 2 Gambar Tahapan <i>KDD</i>	28
Gambar 2. 3 Penggambaran <i>Support Vector Machine</i>	35
Gambar 2. 4 <i>Confusion Matrix</i>	37
Gambar 3. 1 Logo Pertamina	42
Gambar 3. 2 Kerangka Penelitian	46
Gambar 3. 3 <i>Tweet Harvest</i> Menggunakan Google Collab.....	47
Gambar 3. 4 <i>Script Crawl Data 1</i>	48
Gambar 3. 5 Hasil <i>Script Crawl Data 1</i>	48
Gambar 3. 6 <i>Script Crawl Data 2</i>	48
Gambar 3. 7 Hasil <i>Script Crawl Data 2</i>	49
Gambar 3. 8 <i>Script Crawl Data 3</i>	49
Gambar 3. 9 Hasil <i>Script Crawl Data 3</i>	49
Gambar 3. 10 <i>Script Crawl Data 4</i>	50
Gambar 3. 11 Hasil <i>Script Crawling Data 4</i>	50
Gambar 3. 12 <i>Script Crawl Data 5</i>	50
Gambar 3. 13 Hasil <i>Script Crawling Data 5</i>	51
Gambar 3. 14 Hasil <i>Merged Data</i> Pertamina.....	51
Gambar 3. 15 <i>Syntax/Script</i> untuk memeriksa <i>Missing Value</i> dan <i>Duplicate</i>	52
Gambar 3. 16 Hasil dari <i>Missing Value</i> dan <i>Duplicate</i>	53
Gambar 3. 17 Proses Pembersihan <i>Missing Value</i> dan <i>Duplicate</i>	54
Gambar 3. 18 <i>Script/Syntax Case Folding</i>	54
Gambar 3. 19 <i>Script/Syntax Cleaning</i>	56
Gambar 3. 20 <i>Script/Syntax Tokenizing</i>	57

Gambar 3. 21 <i>Script/Syntax Normalization</i>	58
Gambar 3. 22 <i>Script/Syntax Filtering/Stopword Removal</i>	59
Gambar 3. 23 <i>Script/Syntax Stemming</i>	61
Gambar 3. 24 Visualisasi <i>WordCloud</i> setelah <i>Stemming</i>	62
Gambar 3. 25 <i>Script/Syntax Labeling</i>	63
Gambar 3. 26 Hasil dari Labeling Otomatis	64
Gambar 3. 27 Visualiasi Menggunakan <i>Bar Chart</i> dan <i>Pie Chart</i>	64
Gambar 3. 28 <i>script/syntax TF-IDF</i>	68
Gambar 3. 29 Hasil Pembobotan Kata <i>TF-IDF</i>	68
Gambar 3. 30 Proses <i>Split Data</i>	69
Gambar 4. 1 <i>Import Library Sklearn Naïve Bayes</i>	73
Gambar 4. 2 Penerapan <i>SMOTE</i>	73
Gambar 4. 3 <i>Import Library Sklearn Support Vector Machine</i>	80
Gambar 4. 4 Penerapan <i>SMOTE</i>	80
Gambar 4. 5 Hasil Evaluasi <i>Naïve Bayes</i>	83
Gambar 4. 6 <i>Confusion Matrix Naïve Bayes</i>	84
Gambar 4. 7 Hasil Evaluasi <i>SVM</i>	85
Gambar 4. 8 <i>Confusion Matrix SVM</i>	86
Gambar 4. 9 <i>WordCloud Negative Naïve Bayes</i>	88
Gambar 4. 10 <i>WordCloud Neutral Naïve Bayes</i>	89
Gambar 4. 11 <i>WordCloud Positive Naïve Bayes</i>	90
Gambar 4. 12 <i>WordCloud Negative SVM</i>	91
Gambar 4. 13 <i>WordCloud Neutral SVM</i>	92
Gambar 4. 14 <i>WordCloud Positive SVM</i>	93

DAFTAR LAMPIRAN

Uji Plagiarisme

Biodata Mahasiswa

Kartu Konsultasi Bimbingan

Kartu Bimbingan Skripsi Dosen Pembimbing I

Kartu Bimbingan Skripsi Dosen Pembimbing II

Surat Rekomendasi Dari Pembimbing I

Surat Rekomendasi Dari Pembimbing II

Infografis

Dataset

