

**PEMODELAN ANALISIS SENTIMEN BERBASIS
NAÏVE BAYES DAN *SUPPORT VECTOR*
MACHINE UNTUK KASUS
PERSELINGKUHAN DI
INSTAGRAM**

SKRIPSI

Oleh:

Syafira Cessa Agustin

202010225109



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2024**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Proposal Tugas Akhir : *Pemodelan Analisis Sentimen Berbasis Naïve Bayes
Dan Support Vector Machine Untuk Kasus
Perselingkuhan Di Instagram*

Nama Mahasiswa : Syafira Cessa Agustin

Nomor Pokok Mahasiswa : 202010225109

Program Studi/Fakultas : Informatika/Ilmu Komputer

Jakarta, 05 Juli 2024

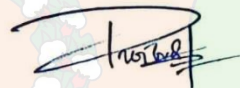
MENYETUJUI,

Pembimbing I

Pembimbing II

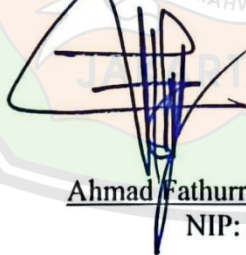


Dr. Herlawati, S.Si., M.M., M.Kom.
NIDN: 0311097302



Dwi Budi Srisulistiowati., S.Kom., M.M.
NIDN: 0323057701

Ketua Program Studi/Sesprodi



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

Program Studi Informatika

Fakultas Ilmu Komputer

Universitas Bhayangkara Jakarta Raya

2024

LEMBAR PENGESAHAN

Judul Tugas Akhir : *Pemodelan Analisis Sentimen Berbasis Naïve Bayes Dan Support Vector Machine Untuk Kasus Perselingkuhan Di Instagram*

Nama Mahasiswa : Syafira Cessa Agustin

Nomor Pokok Mahasiswa : 202010225109

Program Studi/Fakultas : Informatika/Illmu Komputer

Tanggal Lulus Ujian Tugas Akhir : 26 Juni 2024

Jakarta, 05 Juli 2024

MENGESAHKAN,

Ketua Tim Penguji : Dr. Tb. Ai Munandar, S.Kom., MT.

NIDN : 0413098403

Penguji I : Prima Dina Atika, S.Kom., M.Kom.

NIDN : 0311037107

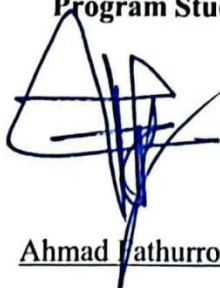
Penguji II : Dr. Herlawati, S.Si., M.M., M.Kom.

NIDN : 0311097302

MENGETAHUI,

Ketua

Program Studi Informatika



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

NIP. 2012486

Dekan

Fakultas Ilmu Komputer



Dr. Dra. Tyastuli Sri Lestari, M.M.

NIP. 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Syafira Cessa Agustin
NPM : 202010225109
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Pemodelan Analisis Sentimen Berbasis *Naïve Bayes*
Dan *Support Vector Machine* Untuk Kasus
Perselingkuhan Di Instagram

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 manapun.

Jakarta, 05 Juli 2024

Penulis



Syafira Cessa Agustin

ABSTRAK

Syafira Cessa Agustin. 202010225109. Pemodelan Analisis Sentimen Berbasis *Naïve Bayes* Dan *Support Vector Machine* Untuk Kasus Perselingkuhan Di Instagram

Pesatnya kemajuan teknologi, media sosial seperti Instagram menjadi platform utama bagi Masyarakat untuk berinteraksi dan berbagi pendapat mengenai berbagai isu, termasuk perselingkuhan. Dari pertengahan 2023 hingga awal 2024, kasus perselingkuhan yang melibatkan aktris, influencer, dan masyarakat biasa menjadi sorotan dan memancing berbagai reaksi pengguna Instagram. Penelitian ini bertujuan untuk menganalisis sentimen dari komentar terkait tiga kasus perselingkuhan pada akun Instagram populer yaitu @lambe_turah dan @awrekeh.id dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM). Hasil analisis menunjukkan bahwa dari total 1200 komentar, 918 bersentimen negatif dan 282 bersentimen positif. Pengujian model dengan pembagian dataset 80:20 menunjukkan bahwa SVM memiliki kinerja lebih baik dengan akurasi 91%, *precision* 87%, *recall* 96%, dan *f1-score* 92% dibandingkan dengan *Naïve Bayes* yang mencapai akurasi 86%, *precision* 82%, *recall* 94%, dan *f1-score* 88%. Temuan ini mengindikasikan bahwa model SVM lebih efektif dalam mengklasifikasikan sentimen komentar, yang dapat diimplementasikan untuk sistem analisis opini publik yang lebih efisien dan akurat.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *Support Vector Machine*, Instagram, Perselingkuhan

ABSTRACT

Syafira Cessa Agustin. 202010225109. *Modeling Sentiment Analysis Based on Naïve Bayes and Support Vector Machine for Infidelity Cases on Instagram*

The rapid advancement of technology, social media such as Instagram has become the main platform for people to interact and share their opinions on various issues, including infidelity. From mid-2023 to early 2024, infidelity cases involving actresses, influencers, and ordinary people were in the spotlight and provoked various reactions from Instagram users. This study aims to analyze the sentiment of comments related to three infidelity cases on popular Instagram accounts, namely @lambe_turah and @awrekeh.id using the Naïve Bayes and Support Vector Machine (SVM) methods. The analysis results show that out of a total of 1200 comments, 918 have negative sentiments and 282 have positive sentiments. Model testing with an 80:20 dataset split showed that SVM performed better with 91% accuracy, 87% precision, 96% recall, and 92% f1-score compared to Naïve Bayes which achieved 86% accuracy, 82% precision, 94% recall, and 88% f1-score. These findings indicate that the SVM model is more effective in classifying comment sentiment, which can be implemented for a more efficient and accurate public opinion analysis system.

Keywords: *Sentiment Analysis, Naïve Bayes, Support Vector Machine, Instagram, Infidelity*

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

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

Nama : Syafira Cessa Agustin
NPM : 202010225109
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 :

Pemodelan Analisis Sentimen Berbasis *Naïve Bayes* Dan *Support Vector Machine* Untuk Kasus Perselingkuhan Di Instagram

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 ijin 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 : Jakarta
Pada tanggal : 05 Juli 2024
Yang Menyatakan



Syafira Cessa Agustin

KATA PENGANTAR

Puji syukur dipanjatkan atas berkah dan rahmat Allah SWT yang telah memberikan nikmat yang begitu besar, sehingga laporan skripsi yang berjudul **“Pemodelan Analisis Sentimen Berbasis Naïve Bayes Dan Support Vector Machine Untuk Kasus Perselingkuhan di Instagram”** dapat diselesaikan dengan lancar. Dalam menyusun penulisan skripsi ini, bantuan dari berbagai pihak sangat berperan penting. Oleh karena itu, ucapan terima kasih disampaikan kepada:

1. Bapak Irjen Pol. (Purn) Prof. Dr. Drs. Bambang Karsono, SH., M.M. Selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Tyastuti Sri Lestari, S.Si., M.M selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, SE., M.M.S.I selaku ketua Program studi Informatika.
4. Ibu Dr. Herlawati, S.Si., M.M., M.Kom selaku dosen pembimbing I skripsi yang telah memberikan masukan serta saran dan motivasi dalam mengerjakan tugas akhir ini.
5. Ibu Dwi Budi Srisulistiwati., S.Kom., M.M selaku dosen pembimbing II skripsi yang telah memberikan masukan serta saran dalam mengerjakan tugas akhir ini.
6. Kedua orang tua serta kakak yang selalu memberi semangat, dukungan, dan selalu mendoakan setiap harinya agar di berikan kelancaran, dan kemudahan dalam menyelesaikan tugas akhir ini.

7. Prilia Hashifah, Nida Rachmatin, Silviana Sevira yang selalu membantu dan mendukung serta berjuang melewati tantangan bersama sejak awal perkuliahan hingga akhir.

Penelitian ini masih jauh dari kesempurnaan dan masih banyak kekurangan. Oleh karenanya, diharapkan kepada pembaca untuk memberikan masukan-masukan berupa saran dan kritik yang bersifat membangun demi kesempurnaan penulisan skripsi ini.



Jakarta, 05 Juli 2024

Penyusun,

Syafira Cessa Agustin

202010225109

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Batasan Masalah	6

1.7	Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA		8
2.1	Penelitian Terdahulu.....	8
2.2	Landasan Teori	13
2.2.1	<i>Machine Learning</i>	13
2.2.2	<i>Data Mining</i>	14
2.2.3	CRISP-DM (<i>Cross-Industry Standard Process for Data Mining</i>)....	15
2.2.4	<i>Text Mining</i>	17
2.2.5	Analisis Sentimen	20
2.2.6	Instagram.....	21
2.2.7	Kasus	22
2.2.8	Perselingkuhan	22
2.2.9	<i>Python</i>	23
2.2.10	<i>Naïve Bayes</i>	24
2.2.11	<i>Support Vector Machine (SVM)</i>	25
2.2.12	<i>Confusion Matrix</i>	27
2.2.13	Lexicon Based.....	30
2.2.14	SMOTE	31
2.2.15	<i>Word Cloud</i>	31
BAB III METODOLOGI PENELITIAN		33

3.1	Objek Penelitian	33
3.2	Kerangka Penelitian	34
3.2.1	<i>Business Understanding</i> (Pemahaman Bisnis).....	35
3.2.2	<i>Data Understanding</i> (Pemahaman Data)	35
3.2.3	<i>Data Preparation</i> (Persiapan Data)	37
3.2.4	<i>Modeling</i> (Pemodelan).....	38
3.2.5	<i>Evaluation</i> (Evaluasi).....	38
3.2.6	<i>Deployment</i> (Penerapan)	39
BAB IV HASIL DAN PEMBAHASAN.....		40
4.1	<i>Business Understanding</i>	40
4.2	<i>Data Understanding</i>	41
4.3	<i>Data Preparation</i>	42
4.3.1	<i>Case Folding</i>	42
4.3.2	<i>Tokenizing</i>	43
4.3.3	<i>Text Normalization</i>	45
4.3.4	<i>Stopword</i>	46
4.3.5	<i>Stemming</i>	47
4.3.6	<i>Labeling</i>	48
4.3.7	TF-IDF	51
4.3.8	SMOTE	56

4.4	<i>Modeling</i>	57
4.4.1	<i>Split Data</i>	57
4.4.2	<i>Naïve Bayes</i>	58
4.4.3	<i>Support Vector Machine (SVM)</i>	65
4.5	<i>Evaluation</i>	70
4.5.1	<i>Naïve Bayes</i>	70
4.5.2	<i>Support Vector Machine (SVM)</i>	73
4.6	<i>Word Cloud</i>	76
4.7	<i>Deployment</i>	80
BAB V	PENUTUP	84
5.1	Kesimpulan	84
5.2	Saran	85
DAFTAR PUSTAKA		87
LAMPIRAN		94

DAFTAR TABEL

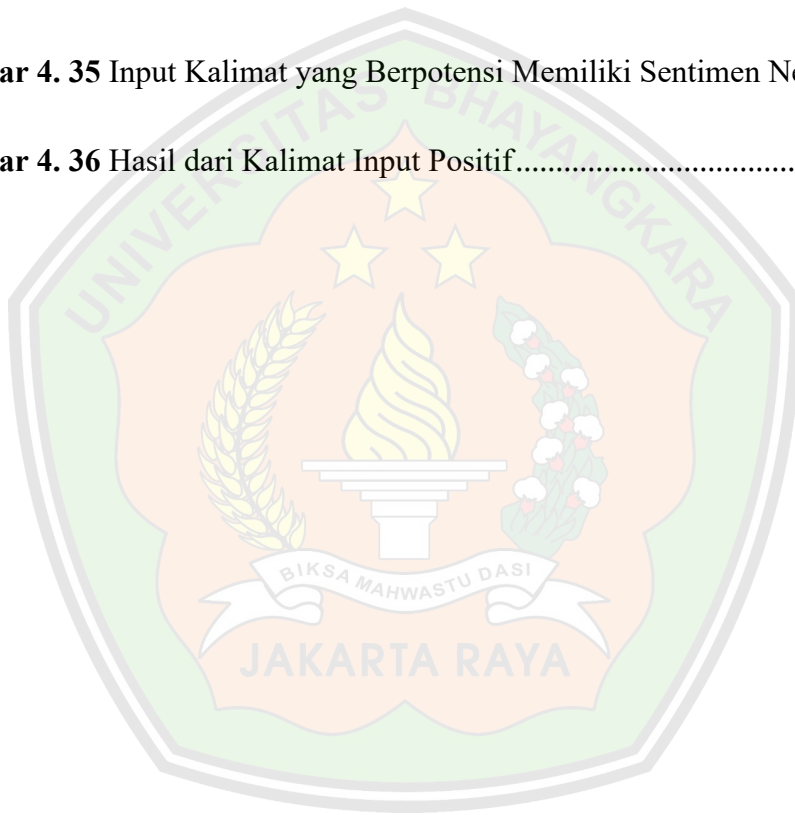
Tabel 2. 1 Penelitian Terkait	8
Tabel 4.1 Hasil <i>Case Folding</i>	43
Tabel 4. 2 Hasil <i>Tokenizing</i>	44
Tabel 4. 3 Hasil <i>Text Normalization</i>	46
Tabel 4. 4 Hasil <i>Stopword</i>	47
Tabel 4. 5 Hasil <i>Stemming</i>	48
Tabel 4. 6 Contoh <i>Dataset</i> Perhitungan.....	53
Tabel 4. 7 Perhitungan Manual TF Normalisasi	54
Tabel 4. 8 Perhitungan Manual DF dan IDF	54
Tabel 4. 9 Perhitungan Manual TF-IDF	55
Tabel 4. 10 <i>Data Training</i>	59
Tabel 4. 11 Jumlah Setiap Kata Dalam Setiap Kelas.....	60
Tabel 4. 12 Tabel Untuk Uji <i>Data Testing</i>	63
Tabel 4. 13 Hasil Probabilitas Pada <i>Data Testing</i>	64
Tabel 4. 14 Hasil TF-IDF	66

DAFTAR GAMBAR

Gambar 1. 1 Jumlah Pengguna Instagram (Oktober 2023)	1
Gambar 2. 1 <i>Machine Learning in Computer Science</i>	13
Gambar 2. 2 Tahapan CRISP-DM	16
Gambar 2. 3 Tahap <i>Preprocessing Data</i>	18
Gambar 2. 4 Logo Instagram	21
Gambar 2. 5 Logo <i>Python</i>	23
Gambar 2. 6 Tabel <i>Confusion Matrix</i>	28
Gambar 3. 1 Objek Penelitian Kasus Perselingkuhan	34
Gambar 3. 2 Kerangka Penelitian CRISP-DM	34
Gambar 3. 3 Data Komentar Instagram	35
Gambar 3. 4 Data Miner	36
Gambar 4. 1 Contoh Process Scraping Data Miner	41
Gambar 4. 2 Data Komentar Instagram	42
Gambar 4. 3 Contoh Code Case Folding	43
Gambar 4. 4 Contoh Code Tokenizing	44
Gambar 4. 5 Contoh Code Text Normalization	45
Gambar 4. 6 Kamus Text Normalization	45
Gambar 4. 7 Contoh Code Stopword	46
Gambar 4. 8 Contoh Code Stemming	47

Gambar 4. 9 Contoh Code Labeling Lexicon Based	49
Gambar 4. 10 Hasil Pelabelan data.....	49
Gambar 4. 11 Grafik Kategori Komentar Instagram	50
Gambar 4. 12 Hasil Labeling Otomatis Pada Dataset	50
Gambar 4. 13 Pengubahan Kolom 'Sentimen' Menjadi Representasi Numerik ..	51
Gambar 4. 14 Contoh Code TF-IDF Perkalimat	52
Gambar 4. 15 Hasil TF-IDF Perkalimat	52
Gambar 4. 16 Contoh Hasil dan Code TF-IDF Seluruh Dokumen	53
Gambar 4. 17 Contoh Hasil dan Code SMOTE	56
Gambar 4. 18 Hasil dan Grafik SMOTE	57
Gambar 4. 19 Code Split Data.....	58
Gambar 4. 20 Contoh Code Modelling Naïve Bayes	59
Gambar 4. 21 Contoh Code Modelling Support Vector Machine	65
Gambar 4. 22 Grafik Kernel Linear SVM.....	66
Gambar 4. 23 Tahapan Evaluation Pada Algoritma Naïve Bayes	71
Gambar 4. 24 Confusion Matrix Dengan Naïve Bayes	71
Gambar 4. 25 Tahapan Evaluation Pada Algoritma Support Vector Machine.....	73
Gambar 4. 26 Confusion Matrix Dengan Support Vector Machine	74
Gambar 4. 27 Hasil Word Cloud Seluruh Data	76
Gambar 4. 28 Frekuensi Kata Yang Sering Muncul di Seluruh Dataset	77

Gambar 4. 29 Hasil Word Cloud Sentimen Positif.....	77
Gambar 4. 30 Frekuensi Kata Yang Sering Muncul Pada Sentimen Positif.....	78
Gambar 4. 31 Hasil Word Cloud Sentimen Negatif	78
Gambar 4. 32 Frekuensi Kata Yang Sering Muncul Pada Sentimen Negatif	79
Gambar 4. 33 Input Kalimat yang Berpotensi Memiliki Sentimen Positif	81
Gambar 4. 34 Hasil dari Kalimat Input Positif.....	81
Gambar 4. 35 Input Kalimat yang Berpotensi Memiliki Sentimen Negatif.....	82
Gambar 4. 36 Hasil dari Kalimat Input Positif.....	83



DAFTAR LAMPIRAN

- Lampiran 1 : Plagiarisme
- Lampiran 2 : Biodata Mahasiswa
- Lampiran 3 : Kartu Bimbingan Mahasiswa
- Lampiran 4 : Data Komentar Instagram

