

**ANALISIS SENTIMEN PENGGUNA APLIKASI BINAR ACADEMY  
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Joseph Nathanael Zebua**  
**00000042748**

**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS TEKNIK DAN INFORMATIKA**  
**UNIVERSITAS MULTIMEDIA NUSANTARA**  
**TANGERANG**  
**2024**

**ANALISIS SENTIMEN PENGGUNA APLIKASI BINAR ACADEMY  
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE**



**Joseph Nathanael Zebua**

**00000042748**

**UMMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Joseph Nathanael Zebua  
Nomor Induk Mahasiswa : 00000042748  
Program Studi : Informatika

Skripsi dengan judul:

**ANALISIS SENTIMEN PENGGUNA APLIKASI BINAR ACADEMY  
MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE**

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 22 Mei 2024



(Joseph Nathanael Zebua)

UMM  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

Skripsi dengan judul

### **ANALISIS SENTIMEN PENGGUNA APLIKASI BINAR ACADEMY MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE**

oleh

Nama : Joseph Nathanael Zebua  
NIM : 00000042748  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

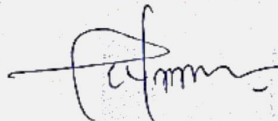
Telah diujikan pada hari Senin, 3 Juni 2024

Pukul 10.00 s/s 12.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

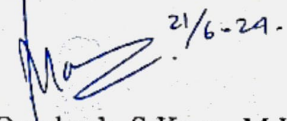
Ketua Sidang



(Sy Yuliani Yakub, S.Kom., M.T. PhD)

NIDN: 0411037904

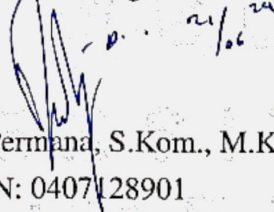
Penguji



(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 818038501

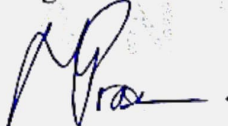
Pembimbing



(Angga Aditya Permana, S.Kom., M.Kom.)

NIDN: 0407128901

Ketua Program Studi Informatika,



(Dr. Eng. Niki Prastomo, S.T., M.Sc.)

NIDN: 0419128203

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK  
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Joseph Nathanael Zebua  
NIM : 00000042748  
Program Studi : Informatika  
Jenjang : S1  
Jenis Karya : Skripsi

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)\*\*.

Tangerang, 22 Mei 2024

Yang menyatakan



Joseph Nathanael Zebua

\*\* Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

**Halaman Persembahan / Motto**

”What you decide on will be done, and light will shine on your ways.”

Job 22:28 (NIV)



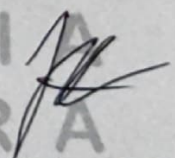
## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: ANALISIS SENTIMEN PENGGUNA APLIKASI BINAR ACADEMY MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Yaman Khaeruzzaman, M.Sc., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Angga Aditya Permana, S.Kom., M.Kom., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua, keluarga serta teman-teman saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 22 Mei 2024

  
Joseph Nathanael Zebua

# ANALISIS SENTIMEN PENGGUNA APLIKASI BINAR ACADEMY MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Joseph Nathanael Zebua

## ABSTRAK

*Binar Academy* merupakan salah satu perusahaan dibidang teknologi edukasi yang menyediakan *bootcamp* di Indonesia. Namun meskipun Indonesia merupakan salah satu negara dengan pengguna *smartphone* yang terbesar di Asia Tenggara, pada saat ini belum ada perusahaan dengan program *bootcamp* selain *Binar Academy* yang memiliki aplikasi berbasis *mobile*. Adanya pembelejaraan berbasis *mobile* mampu meningkatkan tingkat pemahaman dan minat peserta dalam mempelajari suatu materi. Oleh karena itu, Analisis sentimen pengguna pada aplikasi binar penting untuk mengetahui sentimen masyarakat terhadap aplikasi *bootcamp* berbasis *mobile*. Pada penelitian ini, analisis sentimen dilakukan pada ulasan aplikasi *Binar Academy* di *Google Play Store* menggunakan algoritma klasifikasi *Support Vector Machine* (SVM). Ketidakseimbangan pada populasi data dapat berdampak negatif pada hasil pengujian model karena dapat membuat hasil dari model menjadi bias terhadap kelas mayoritas. Metode *imbalance handling* yang digunakan pada penelitian ini adalah *SMOTE* dan *ADASYN*. Pembobotan dilakukan dengan TF-IDF untuk menentukan kepentingan data pada keperluan analisis sentimen. Algoritma SVM yang mengaplikasikan *imbalance handling* seperti *SMOTE* dan *ADA-SYN* dievaluasi menggunakan *confusion matrix* dan divalidasi menggunakan *K-Fold Cross Validation*. Performa model terbaik didapat pada model yang menggunakan *ADA-SYN* dengan *accuracy* sebesar 95%, *precision* sebesar 94%, *recall* sebesar 94% dan *f1-score* sebesar 94%.

**Kata kunci:** *ADA-SYN*, Analisis Sentimen, *Binar Academy*, *bootcamp*, *Confusion Matrix*, *K-Fold Cross Validation*, *SMOTE*, *Support Vector Machine*, *TF-IDF*

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



# SENTIMENT ANALYSIS OF BINAR ACADEMY USER REVIEWS USING SUPPORT VECTOR MACHINE ALGORITHM

Joseph Nathanael Zebua

## ABSTRACT

Binar Academy is one of the companies in the field of educational technology that provides bootcamp in Indonesia. But even though Indonesia is one of the one of the countries with the largest smartphone users in Southeast Asia, currently there are no companies with bootcamp programs other than Binar Academy that have mobile-based applications. has a mobile-based application. The existence of mobile-based learning can increase the level of understanding and interest of participants in learning a material. Therefore, user sentiment analysis on the binar application is important to find out people's sentiment towards mobile-based bootcamp applications. In this study, sentiment analysis was carried out on the Binar Academy app reviews on the Google Play Store using the Support Vector Machine (SVM) classification algorithm. Imbalance in the data population can have a negative impact on model testing results because it can make the results of the model biased towards the majority class. The Weighting is done with TF-IDF to determine the importance of data for sentiment analysis purposes. SVM algorithms that apply imbalance handling such as SMOTE and ADA-SYN are evaluated using confusion matrix and validated using K-Fold Cross Validation. The best model performance is obtained in the model that uses ADA- SYN with an accuracy of 95%, precision of 94%, recall of 94% and f1-score of 94%.

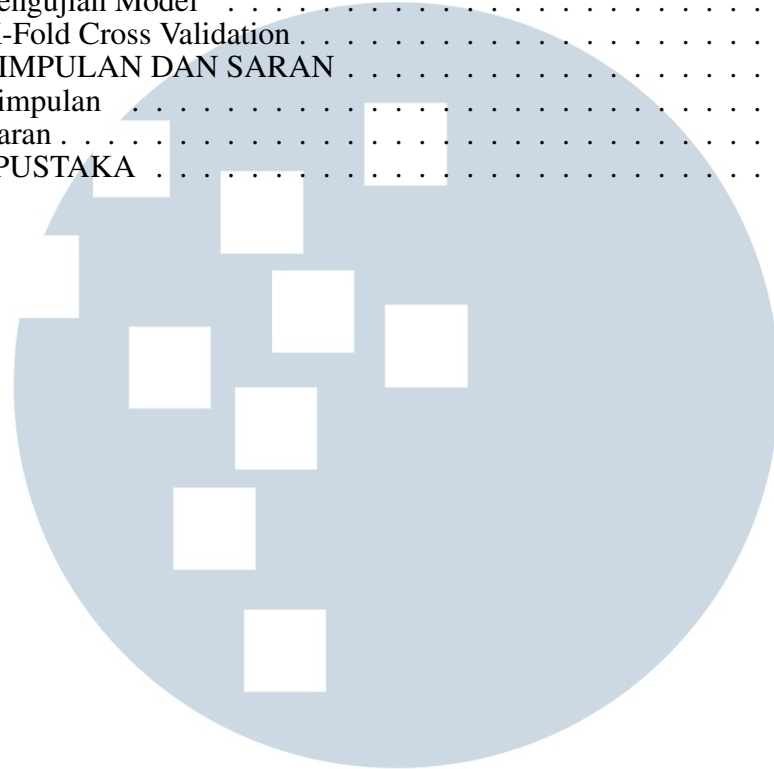
**Keywords:** *ADA-SYN, Binar Academy, bootcamp, Confusion Matrix, K-Fold Cross Validation, Sentiment Analysis, SMOTE, Support Vector Machine, TF-IDF*

U M N  
U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	vii
ABSTRACT . . . . .	viii
DAFTAR ISI . . . . .	ix
DAFTAR GAMBAR . . . . .	xi
DAFTAR TABEL . . . . .	xii
DAFTAR KODE . . . . .	xiii
DAFTAR LAMPIRAN . . . . .	xiv
BAB 1 PENDAHULUAN . . . . .	1
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Batasan Permasalahan . . . . .	3
1.4 Tujuan Penelitian . . . . .	3
1.5 Manfaat Penelitian . . . . .	4
1.6 Sistematika Penulisan . . . . .	4
BAB 2 LANDASAN TEORI . . . . .	5
2.1 Analisis Sentimen . . . . .	5
2.2 Bootcamp . . . . .	5
2.3 Binar Academy . . . . .	6
2.3.1 Support Vector Machine . . . . .	6
2.3.2 TF-IDF . . . . .	9
2.3.3 Synthetic Minority Oversampling Technique (SMOTE) . . . . .	9
2.3.4 Adaptive Synthetic Sampling (ADA-SYN) . . . . .	10
2.4 K-Fold Cross Validation . . . . .	10
BAB 3 METODOLOGI PENELITIAN . . . . .	13
3.1 Data Scraping . . . . .	14
3.2 Data Pre-processing . . . . .	14
3.2.1 Labeling . . . . .	14
3.2.2 Cleaning dan Case Folding . . . . .	15
3.2.3 Tokenizing . . . . .	15
3.2.4 Filtering . . . . .	16
3.2.5 Stemming . . . . .	16
3.3 Pembuatan model . . . . .	16
3.3.1 Train Test Split dan TF-IDF . . . . .	17
3.3.2 Imbalance Handling . . . . .	18
3.3.3 Support Vector Machine . . . . .	18
3.4 Evaluasi Model . . . . .	18
BAB 4 HASIL DAN DISKUSI . . . . .	20
4.1 Spesifikasi Sistem . . . . .	20
4.2 Hasil Implementasi . . . . .	20
4.2.1 Data Scraping . . . . .	20
4.2.2 Data Pre-processing . . . . .	22
4.3 TF-IDF . . . . .	29

4.4	Imbalance Handling . . . . .	30
4.5	Pembuatan model . . . . .	30
4.6	Pengujian Model . . . . .	31
4.7	K-Fold Cross Validation . . . . .	42
BAB 5	SIMPULAN DAN SARAN . . . . .	44
5.1	Simpulan . . . . .	44
5.2	Saran . . . . .	44
	DAFTAR PUSTAKA . . . . .	46



UMMN  
 UNIVERSITAS  
 MULTIMEDIA  
 NUSANTARA

## DAFTAR GAMBAR

Gambar 2.1	Algoritma <i>Support Vector Machine</i> . . . . .	7
Gambar 2.2	Linear vs Non-Linear . . . . .	7
Gambar 2.3	Confusion Matrix . . . . .	11
Gambar 3.1	Flowchart pengerjaan penelitian . . . . .	13
Gambar 3.2	Flowchart data pre-processing . . . . .	14
Gambar 3.3	Pembuatan Model . . . . .	17
Gambar 3.4	Contoh <i>Confusion Matrix</i> . . . . .	19
Gambar 4.1	Data Hasil Scraping . . . . .	22
Gambar 4.2	Labeling data . . . . .	24
Gambar 4.3	Cleaning Text dan Case Folding data . . . . .	25
Gambar 4.4	Tokenizing data . . . . .	26
Gambar 4.5	Filtering data . . . . .	27
Gambar 4.6	Stemming data . . . . .	29
Gambar 4.7	<i>Confusion Matrix</i> dataset tanpa <i>imbalance handling</i> dengan perbandingan dataset 60:40 . . . . .	33
Gambar 4.8	<i>Confusion Matrix</i> dataset <i>SMOTE</i> dengan perbandingan dataset 60:40 . . . . .	34
Gambar 4.9	<i>Confusion Matrix</i> dataset <i>ADA-SYN</i> dengan perbandingan dataset 60:40 . . . . .	34
Gambar 4.10	<i>Confusion Matrix</i> dataset tanpa <i>imbalance handling</i> dengan perbandingan dataset 70:30 . . . . .	37
Gambar 4.11	<i>Confusion Matrix</i> dataset <i>SMOTE</i> dengan perbandingan dataset 70:30 . . . . .	37
Gambar 4.12	<i>Confusion Matrix</i> dataset <i>ADA-SYN</i> dengan perbandingan dataset 70:30 . . . . .	38
Gambar 4.13	<i>Confusion Matrix</i> dataset tanpa <i>imbalance handling</i> dengan perbandingan dataset 80:20 . . . . .	40
Gambar 4.14	<i>Confusion Matrix</i> dataset <i>SMOTE</i> dengan perbandingan dataset 80:20 . . . . .	41
Gambar 4.15	<i>Confusion Matrix</i> dataset <i>ADA-SYN</i> dengan perbandingan dataset 80:20 . . . . .	41

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## DAFTAR TABEL

Tabel 2.1	Confusion Matrix . . . . .	11
Tabel 3.1	Contoh <i>labeling</i> sentimen . . . . .	15
Tabel 3.2	Contoh <i>cleaning</i> dan <i>case folding</i> . . . . .	15
Tabel 3.3	Contoh <i>tokenizing</i> . . . . .	15
Tabel 3.4	Contoh <i>filtering</i> . . . . .	16
Tabel 3.5	Contoh <i>stemming</i> . . . . .	16
Tabel 4.1	Jumlah dari setiap sentimen setelah proses labeling . . . . .	24
Tabel 4.2	Jumlah data sebelum dan sesudah imbalance handling pada perbandingan 60:40 . . . . .	33
Tabel 4.3	Hasil pemodelan dengan perbandingan dataset 60:40) . . . . .	33
Tabel 4.4	Jumlah data sebelum dan sesudah imbalance handling pada perbandingan 70:30 . . . . .	36
Tabel 4.5	Hasil pemodelan dengan perbandingan dataset 70:30) . . . . .	36
Tabel 4.6	Jumlah data sebelum dan sesudah imbalance handling pada perbandingan 80:20 . . . . .	39
Tabel 4.7	Hasil pemodelan dengan perbandingan dataset 80:20 . . . . .	40
Tabel 4.8	Hasil <i>K-Fold Cross Validation</i> dengan dataset Tanpa <i>Imbalance Handling</i> . . . . .	42
Tabel 4.9	Hasil <i>K-Fold Cross Validation</i> dengan dataset <i>SMOTE</i> . . . . .	43
Tabel 4.10	Hasil <i>K-Fold Cross Validation</i> dengan dataset <i>ADA-SYN</i> . . . . .	43



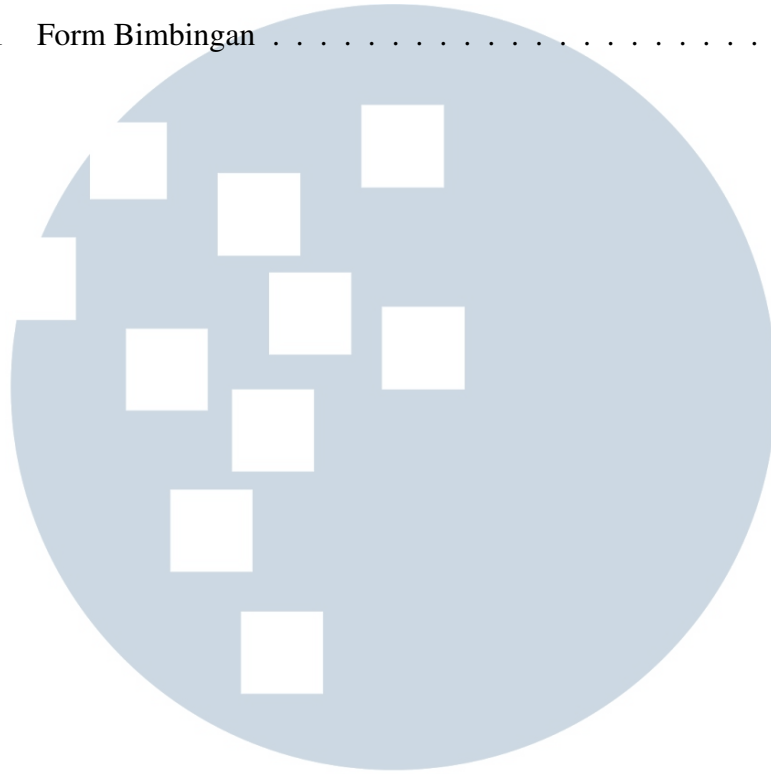
## DAFTAR KODE

4.1	Script untuk melakukan scraping data . . . . .	21
4.2	Pembuatan dataframe dan konversi kedalam bentuk csv . . . . .	21
4.3	Script untuk melakukan labeling data . . . . .	23
4.4	Script Cleaning dan Case Folding . . . . .	24
4.5	Script Tokenizing . . . . .	26
4.6	Script Filtering . . . . .	26
4.7	Script Stemming . . . . .	28
4.8	Penggabungan text string . . . . .	29
4.9	TF-IDF . . . . .	29
4.10	Imbalance handling menggunakan SMOTE . . . . .	30
4.11	Imbalance handling menggunakan ADA-SYN . . . . .	30
4.12	Implementasi algoritma SVM . . . . .	30
4.13	Imbalance handling menggunakan SMOTE pada dataset 60:40 . . . . .	32
4.14	Imbalance handling menggunakan ADA-SYN pada dataset 60:40 . . . . .	32
4.15	Imbalance handling menggunakan SMOTE pada dataset 70:30 . . . . .	35
4.16	Imbalance handling menggunakan ADA-SYN pada dataset 70:30 . . . . .	35
4.17	Imbalance handling menggunakan SMOTE pada dataset 80:20 . . . . .	39
4.18	Imbalance handling menggunakan ADA-SYN pada dataset 80:20 . . . . .	39



**DAFTAR LAMPIRAN**

Lampiran 1 Form Bimbingan . . . . . 50



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA