

**ANALISIS SENTIMEN PROGRAM KAMPUS
MERDEKA IISMA BERBASIS KOMENTAR TIKTOK
DAN *TWEETS* TWITTER MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE DAN *FASTTEXT***



**OLEH
PUTU AGISNA FRANSISCO SARA
NIM 2015091005**

**PROGRAM STUDI SISTEM INFORMASI
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2024



**ANALISIS SENTIMEN PROGRAM KAMPUS
MERDEKA IISMA BERBASIS KOMENTAR TIKTOK
DAN *TWEETS* TWITTER MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE DAN *FASTTEXT***

SKRIPSI

**Diajukan kepada
Universitas Pendidikan Ganesha
Untuk Memenuhi Salah Satu Persyaratan dalam Menyelesaikan
Program Sarjana Sistem Informasi**



**Oleh
Putu Agisna Fransisco Sara
NIM 2015091005**

**PROGRAM STUDI SISTEM INFORMASI
JURUSAN PENDIDIKAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2024

SKRIPSI

DIAJUKAN UNTUK MELENGKAPI TUGAS DAN MEMENUHI SYARAT-SYARAT UNTUK MENCAPAI GELAR SARJANA SISTEM INFORMASI

Menyetujui

Dosen Pembimbing I,



I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T.
NIP. 198907112020122004

Dosen Pembimbing II,



I Made Dendi Maysanjaya, S.Pd., M.Eng.
NIP. 199005152019031008

Skripsi oleh Putu Agisna Fransisco Sara
Telah dipertahankan didepan dewan penguji
Pada tanggal 31 Januari 2024

Dewan Penguji,



Dr. I Made Gede Sunarya, S.Kom., M.Cs.
NIP. 198307252008011008

(Ketua)



Gede Surya Mahendra, S.Pd., M.Kom.
NIP. 199003132022031009

(Anggota)



I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T.
NIP. 198907112020122004

(Anggota)



I Made Dendi Maysanjaya, S.Pd., M.Eng.
NIP. 199005152019031008

(Anggota)

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
Guna memenuhi syarat-syarat untuk mencapai gelar Sarjana Komputer

Pada :

Hari : Rabu
Tanggal : 31 Januari 2024

Mengetahui,

Ketua Ujian,



Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.
NIP. 198211112008121001

Sekretaris Ujian,



I Made Dendi Maysanjaya, S.Pd., M.Eng.
NIP. 199005152019031008

Mengesahkan,

Dekan Fakultas Teknik dan Kejuruan



Dr. Kadek Rihendra Dantes, S.T., M.T.
NIP. 197912012006041001

PERNYATAAN

Dengan ini saya menyatakan bahwa karya tulis yang berjudul “ANALISIS SENTIMEN PROGRAM KAMPUS MERDEKA IISMA BERBASIS KOMENTAR TIKTOK DAN *TWEETS* TWITTER MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN *FASTTEXT*” beserta seluruh isinya adalah benar-benar karya sendiri dan saya tidak melakukan penjiplakan serta pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko dan sanksi yang dijatuhkan kepada saya apabila kemudian hari ditemukan adanya pelanggaran atas etika keilmuan dalam karya saya ini atau ada klaim terhadap keaslian karya.

Singaraja, 31 Januari 2024

Yang membuat pernyataan,



Putu Agisna Fransisco Sara

NIM. 2015091005

PRAKATA

Puji syukur penulis haturkan ke hadapan Tuhan Yang Maha Esa, *Ida Sang Hyang Widhi Wasa* karena berkat rahmat-Nya lah, penulis dapat menyelesaikan skripsi yang berjudul “ANALISIS SENTIMEN PROGRAM KAMPUS MERDEKA IISMA BERBASIS KOMENTAR TIKTOK DAN *TWEETS* TWITTER MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN *FASTTEXT*”. Skripsi ini disusun guna memenuhi persyaratan mencapai gelar sarjana komputer pada Universitas Pendidikan Ganesha.

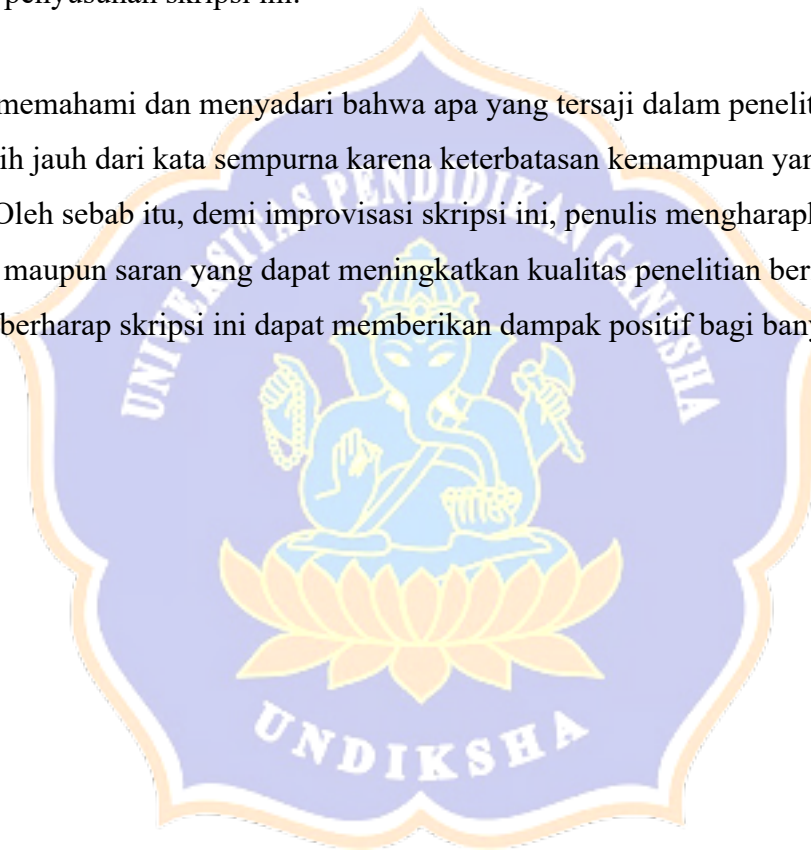
Pada perjalanan menyelesaikan skripsi ini, penulis banyak mendapat bantuan baik berupa moral maupun material dari berbagai pihak. Oleh sebab itu, dalam kesempatan ini, penulis mengucapkan rasa terima kasih kepada :

1. Bapak Dr. Kadek Rihendra Dantes, S.T., M.T. , selaku Dekan Fakultas Teknik dan Kejuruan atas kinerja, motivasi dan fasilitas yang diberikan sehingga penulis bisa menyelesaikan studi sesuai dengan rencana.
2. Bapak Putu Hendra Suputra, S.Kom., M.Cs. , selaku Ketua Jurusan Teknik Informatika atas kinerja, motivasi dan fasilitas yang diberikan dalam penyelesaian skripsi ini.
3. Ibu I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. , selaku Sekretaris Jurusan Teknik Informatika sekaligus sebagai Dosen Pembimbing I yang telah memberikan bimbingan, arahan, petunjuk, dan motivasi kepada penulis dalam penyelesaian skripsi ini.
4. Bapak I Made Dendi Maysanjaya, S.Pd., M.Eng., Selaku Koordinator Program Studi Sistem Informasi sekaligus Pembimbing II yang telah memberikan bimbingan, arahan, petunjuk, dan motivasi kepada penulis dalam penyelesaian skripsi ini.
5. Ibu Dra. I Wayan Adi Wahyuni dan Ibu Dra. Kadek Widhiasih selaku Guru Bahasa Indonesia SMA Negeri 4 Singaraja, Buleleng yang bersedia meluangkan waktunya untuk membantu penulis dalam melabeli data penelitian dalam penyelesaian skripsi ini.
6. Seluruh Dosen jurusan Teknik Informatika, Program Studi Sistem Informasi yang telah berbagi ilmu dan pengalamannya selama kuliah

semoga semua yang peneliti pelajari di kampus dapat bermanfaat bagi diri sendiri, orang terdekat dan masyarakat luas.

7. Secara khusus keluarga yaitu orang tua, saudara yang telah mendoakan dan memberikan dukungan yang sangat berarti kepada peneliti dalam menyelesaikan skripsi ini.
8. Rekan mahasiswa Program Studi Sistem Informasi yang telah memberikan motivasi kepada penulis dalam penyelesaian skripsi ini.
9. Semua pihak yang tidak bisa disebutkan satu per satu yang telah membantu penyusunan skripsi ini.

Penulis memahami dan menyadari bahwa apa yang tersaji dalam penelitian skripsi ini masih jauh dari kata sempurna karena keterbatasan kemampuan yang penulis miliki. Oleh sebab itu, demi improvisasi skripsi ini, penulis mengharapkan segala kritik maupun saran yang dapat meningkatkan kualitas penelitian berikutnya. Penulis berharap skripsi ini dapat memberikan dampak positif bagi banyak orang.



DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN LOGO	ii
HALAMAN JUDUL.....	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
LEMBAR PERSETUJUAN DOSEN PENGUJI.....	v
LEMBAR PERSETUJUAN DAN PENGESAHAN PANITIA UJIAN.....	vi
PERNYATAAN.....	vii
KATA PERSEMBAHAN	viii
MOTTO.....	ix
PRAKATA	x
ABSTRAK	xii
<i>ABSTRACT</i>	xiii
DAFTAR ISI.....	xiv
DAFTAR GAMBAR	xvi
DAFTAR TABEL.....	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian (Lihat Rumusan Masalah).....	6
1.4 Ruang Lingkup Penelitian.....	6
1.5 Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA.....	8
2.1 Studi Sebelumnya.....	8
2.2 Dasar Teori.....	11
2.2.1 Merdeka Belajar Kampus Merdeka (MBKM).....	11
2.2.2 Indonesian International Student Mobility Awards (IISMA).....	12
2.2.3 Twitter	13
2.2.4 TikTok.....	14
2.2.5 Data Mining	17
2.2.6 Cross-Industry Standard Process for Data Mining (CRISP-DM ..	19
2.2.7 Text Mining.....	21
2.2.8 Scraping Data (TikTok.....	22
2.2.9 Crawling Data (Twitter).....	23

2.2.10	Data Imbalance.....	23
2.2.11	Analisis Sentimen (Opinion Mining).....	23
2.2.12	Klasifikasi	24
2.2.13	FastText.....	24
2.2.14	Support Vector Machine (SVM).....	25
2.2.15	Confusion Matrix	29
2.2.16	K-Fold Cross Validation	31
BAB III METODOLOGI PENELITIAN.....		33
3.1	Jenis Penelitian.....	33
3.2	Tahapan Penelitian	33
3.2.1	Business Understanding.....	35
3.2.2	Data Understanding.....	35
3.2.3	Data Preparation.....	35
3.2.4	Modeling	46
3.2.5	Evaluation	49
3.2.6	Deployment.....	50
BAB IV HASIL DAN PEMBAHASAN		51
4.1	Hasil	51
4.1.1	Business Understanding.....	51
4.1.2	Data Understanding.....	51
4.1.3	Data Preparation.....	51
4.1.4	Modeling	65
4.1.5	Evaluation	70
4.1.6	Deployment.....	76
4.2	Pembahasan.....	82
4.2.1	Data Preparation.....	82
4.2.2	Modeling	85
4.2.3	Evaluation	87
4.2.4	Deployment.....	90
BAB V PENUTUP.....		94
5.1	Kesimpulan	94
5.2	Saran.....	96
DAFTAR PUSTAKA		98
LAMPIRAN.....		103

DAFTAR GAMBAR

Gambar 2.1 Tahapan-tahapan <i>data mining</i> (Aryuni, 2016).....	18
Gambar 2.2 Tahapan CRISP-DM (Puspitasari, 2022).....	20
Gambar 2.3 Ilustrasi mengubah kata menjadi vektor(Adam, 2019).....	24
Gambar 2.4 <i>Pretrained Model</i> Bahasa Indonesia.....	25
Gambar 2.5 SVM Model Linier(Dahman, 2021).....	26
Gambar 2.6 Transformasi data ruang input ke ruang fitur(Wilimitis, 2018).....	27
Gambar 2.7 Ilustrasi <i>10-fold cross-validation</i> (Andrade et al., 2020).....	32
Gambar 3.1 Alur Tahapan Penelitian.....	34
Gambar 3.2 Tahapan <i>Text Pre-processing</i>	42
Gambar 3.3 Tahapan Pemodelan SVM OVO.....	47
Gambar 4.1 Grafik <i>Pie</i> Distribusi Sentimen <i>Dataset</i> TikTok.....	57
Gambar 4.2 Grafik <i>Pie</i> Distribusi Sentimen <i>Dataset</i> Twitter.....	58
Gambar 4.3. Hasil Distribusi Sentimen setelah <i>Random Undersampling</i>	68
Gambar 4.4 Hasil <i>Learning Curve</i> dan <i>Validation Error</i>	70
Gambar 4.5. Hasil Evaluasi <i>Confussion Matrix Dataset</i> Gabungan.....	74
Gambar 4.6. Hasil Evaluasi <i>Confussion Matrix Dataset</i> Twitter.....	72
Gambar 4.7 Hasil Evaluasi <i>Confussion Matrix Dataset</i> TikTok.....	73
Gambar 4.8 Wordcloud <i>Dataset</i> TikTok untuk Sentimen Netral.....	77
Gambar 4.9 Wordcloud <i>Dataset</i> TikTok untuk Sentimen Negatif.....	78
Gambar 4.10 Wordcloud <i>Dataset</i> TikTok untuk Sentimen Positif.....	78
Gambar 4.11 Wordcloud <i>Dataset</i> Twitter untuk Sentimen Positif.....	79
Gambar 4.12 Wordcloud <i>Dataset</i> Twitter untuk Sentimen Netral.....	80
Gambar 4.13 Wordcloud <i>Dataset</i> Twitter untuk Sentimen Negatif.....	80
Gambar 4.14 Wordcloud <i>Dataset</i> Gabungan untuk Sentimen Netral.....	81
Gambar 4.15 Wordcloud <i>Dataset</i> Gabungan untuk Sentimen Negatif.....	81
Gambar 4.16 Wordcloud <i>Dataset</i> Gabungan untuk Sentimen Positif.....	82
Gambar 5.1 Distribusi Aktual dan Prediksi <i>Dataset</i> Gabungan.....	88
Gambar 5.2 Distribusi Aktual dan Prediksi <i>Dataset</i> Twitter.....	89
Gambar 5.3 Distribusi Aktual dan Prediksi <i>Dataset</i> TikTok.....	90

DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix Biner</i>	29
Tabel 2.2 <i>Confusion Matrix Multiclass</i> (Trivusi, 2022)	30
Tabel 3.1 Kata kunci dan <i>hashtag</i>	35
Tabel 3.2 Contoh <i>Dataset</i> Komentar TikTok	36
Tabel 3.3 Contoh <i>Dataset</i> Twitter.....	37
Tabel 3.4 Contoh Pelabelan <i>Dataset</i> Twitter	39
Tabel 3.5 Contoh Pelabelan <i>Dataset</i> TikTok.....	41
Tabel 3.6 Contoh Proses <i>Cleansing</i>	42
Tabel 3.7 Contoh Proses <i>Case Folding</i>	43
Tabel 3.8 Contoh Proses <i>Tokenizing</i>	44
Tabel 3.9 Contoh Proses <i>Normalization</i>	44
Tabel 3.10 Contoh Proses <i>Stopword Removing</i>	45
Tabel 3.11 Contoh Proses <i>Stemming</i>	46
Tabel 3.12 Contoh Tabel <i>Confusion Matrix Multiclass</i>	49
Tabel 4.1 Kode Program <i>Crawling</i> Data.....	52
Tabel 4.2 Contoh Daftar Video TikTok Dalam Bentuk JSON	53
Tabel 4.3 Hasil Pengumpulan Data Komentar <i>TikTok</i>	53
Tabel 4.4. Hasil Pengumpulan Data <i>Tweets</i> Twitter.....	54
Tabel 4.5. Rumus <i>Microsoft Excel</i> untuk Filter Data.....	54
Tabel 4.6 Hasil Filter Data Komentar TikTok Melalui <i>Microsoft Excel</i>	55
Tabel 4.7 Hasil Filter Data <i>Tweets</i> Twitter Melalui <i>Microsoft Excel</i>	55
Tabel 4.8 Hasil Pelabelan <i>Dataset</i> Komentar TikTok	56
Tabel 4.9 Hasil Pelabelan <i>Dataset Tweets</i> Twitter.....	56
Tabel 4.10 Kode Program Proses <i>Cleansing</i>	58
Tabel 4.11 Hasil Proses <i>Cleansing</i>	59
Tabel 4.12 Kode Program Proses <i>Case Folding</i>	59
Tabel 4.13 Hasil Proses <i>Case Folding</i>	60
Tabel 4.14 Kode Program Proses <i>Tokenizing</i>	60
Tabel 4.15 Hasil Proses <i>Tokenizing</i>	61
Tabel 4.16. Kode Program Proses <i>Normalization</i>	61
Tabel 4.17 Hasil Proses <i>Normalization</i>	62

Tabel 4.18 Kode Program Proses <i>Stopword Removing</i>	62
Tabel 4.19 Hasil Proses <i>Stopword Removing</i>	63
Tabel 4.20 Kode Program Proses <i>Stemming</i>	63
Tabel 4.21 Hasil Proses <i>Stemming</i>	64
Tabel 4.22 Kode Program <i>Export Dataset</i>	65
Tabel 4.23 Penjelasan Hasil <i>Dataset</i>	65
Tabel 4.24 Hasil Eksperimen Model Ekstraksi Fitur <i>FastText</i>	66
Tabel 4.25 Kode Program Ekstraksi Fitur <i>FastText</i>	66
Tabel 4.26 Hasil Proses Ekstraksi Fitur <i>FastText</i>	67
Tabel 4.27 Perbandingan Model Klasifikasi	69
Tabel 4.28 Kode Program Model Klasifikasi SVM	69
Tabel 4.29 Kode Program Evaluasi <i>Cross-Validation</i>	71
Tabel 4.30 Hasil Evaluasi <i>K-Fold Cross Validation Dataset</i> Twitter	72
Tabel 4.31 Hasil Evaluasi <i>K-Fold Cross Validation Dataset</i> TikTok	74
Tabel 4.32 Hasil Evaluasi <i>K-Fold Cross Validation Dataset</i> Gabungan	75
Tabel 4.33. Kode Program Evaluasi <i>Confussion Matrix</i>	75
Tabel 4.34 Kode Program Visualisasi	76
Tabel 4.37 Pencocokan Hasil <i>Wordcloud</i> Terhadap <i>Dataset</i>	91

