

**PENERAPAN *SUPPORT VECTOR MACHINE*
UNTUK MENGANALISIS SENTIMEN
MASYARAKAT TERHADAP PROGRAM
MAGANG DAN STUDI INDEPENDEN
BERSERTIFIKAT**

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Matematika**



**ERSTYNA KATHLYA PRABOWO
1305619017**

**PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA
2023**

ABSTRAK

ERSTYNA KATHLYA PRABOWO. Penerapan *Support Vector Machine* untuk Menganalisis Sentimen Masyarakat terhadap Program Magang dan Studi Independen Bersertifikat. Skripsi, Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta, Juli 2023.

Program Magang dan Studi Independen Bersertifikat (MSIB) merupakan salah satu program yang dicetuskan oleh Kemendikbud untuk mendukung salah satu kebijakan kurikulum Kampus Merdeka yang memberikan hak kepada mahasiswa untuk belajar di luar kampus selama tiga semester. Selama berjalan hampir dua tahun, program MSIB menuai opini dan pendapat dari masyarakat terhadap program ini, baik yang mengandung sentimen negatif, positif, atau netral. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program MSIB menggunakan *Support Vector Machine* (SVM). *Support Vector Machine* merupakan model *machine learning* yang dapat bekerja dengan baik pada data berdimensi tinggi seperti data teks. Hasil penelitian menunjukkan bahwa rata-rata *F1 score* yang dihasilkan oleh *Support Vector Machine* pada saat tahap *5-Fold Cross Validation* adalah sebesar 68,2%. Kinerja *Support Vector Machine* dalam mengklasifikasikan data baru termasuk pada kriteria cukup baik, yaitu menghasilkan nilai *F1 score* sebesar 74,12% ketika digunakan untuk mengklasifikasikan sentimen pada data uji. Hasil keseluruhan klasifikasi sentimen menggunakan *Support Vector Machine* menunjukkan bahwa sentimen netral menjadi kelas sentimen paling dominan dalam pendapat dan opini masyarakat terhadap program MSIB (74,39%), sedangkan sentimen negatif menjadi kelas sentimen kedua yang frekuensinya paling banyak muncul (13,60%). Sentimen positif menjadi kelas sentimen yang paling sedikit muncul pada pendapat dan opini masyarakat terhadap program MSIB (12%). Sentimen netral mendominasi pendapat dan opini masyarakat terhadap program MSIB terjadi karena banyaknya data opini yang cenderung merupakan kalimat pertanyaan atau kalimat berita tentang program MSIB itu sendiri.

Kata kunci. Analisis sentimen, *Support Vector Machine*, Program MSIB

ABSTRACT

ERSTYNA KATHLYA PRABOWO. Implementation of Support Vector Machine for Analyzing Public Sentiment of Certified Internship and Independent Study Program. Undergraduate Thesis, Mathematics Study Program, Faculty of Mathematics and Natural Sciences, State University of Jakarta, July 2023.

Certified Internship and Independent Study Program (MSIB) is one of the programs which initiated by the Ministry of Education, Culture, Research and Technology to support one of the Kampus Merdeka curriculum policies that give opportunities for students to learn and develop themselves through activities outside class for three semesters. After almost two years, the MSIB program has received opinions from the public, whether it contains negative, positive, or neutral sentiments. This research aims to analyze public sentiment towards the MSIB program using Support Vector Machine (SVM). Support Vector Machine is a machine learning model that can work well on high-dimensional data, such as text data. The result shows that the average F1 score given by the Support Vector Machine when passed the 5-Fold Cross Validation process is 68,2%. The performance of the Support Vector Machine in classifying new data is categorized as tolerable, which produced an F1 Score of 74,12%, when using SVM to classify sentiments on data test. The overall results of sentiment classification using the Support Vector Machine show that neutral sentiment dominates public opinion about the MSIB program (74,39%), while negative sentiment becomes the second class of sentiment with the highest frequency (13,60%). Positive sentiment is the class sentiment that appears the least in public opinions about the MSIB program (12%). Neutral sentiment dominates public opinion toward the MSIB program because a large amount of opinion data contains interrogative sentences or news about the MSIB program itself




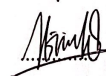
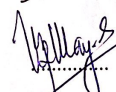

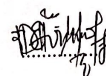
Keywords. *Sentiment analysis, Support Vector Machine, MSIB Program*

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

PENERAPAN *SUPPORT VECTOR MACHINE* UNTUK MENGANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM MAGANG DAN STUDI INDEPENDEN BERSERTIFIKAT

Nama : Erstyna Kathlya Prabowo

No. Registrasi : 1305619017

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			29 Agustus 2023
Dekan	: <u>Prof. Dr. Mukhtiningsih N, M.Si.</u> NIP. 196405111989032001	
Wakil Penanggung Jawab			29 Agustus 2023
Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT.</u> NIP. 197207281999031002	
Ketua	: <u>Dr. Eti Dwi Wiraningsih, S.Pd., M.Si.</u> NIP. 198102032006042001		21 Agustus 2023
Sekretaris	: <u>Ibnu Hadi, M.Si.</u> NIP. 198107182008011017		21 Agustus 2023
Penguji	: <u>Vera Maya Santi, M.Si.</u> NIP. 197905312005012006		16 Agustus 2023
Pembimbing I	: <u>Dr. Dian Handayani, M.Si.</u> NIP. 197404151998032001		21 Agustus 2023
Pembimbing II	: <u>Devi Eka Wardani M., S.Pd., M.Si.</u> NIP. 199005162019032014		21 Agustus 2023

Dinyatakan lulus ujian skripsi tanggal: 10 Agustus 2023

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul **"Penerapan *Support Vector Machine* untuk Menganalisis Sentimen Masyarakat Terhadap Program Magang dan Studi Independen Bersertifikat"** yang disusun sebagai syarat untuk memperoleh gelar Sarjana Matematika dari Program Studi Matematika Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 30 Juli 2023



Erstyna Kathlya Prabowo



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Erstyna Kathlya Prabowo
NIM : 1305619017
Fakultas/Prodi : Fakultas Matematika & Ilmu Pengetahuan Alam (FMIPA) / Matematika
Alamat email : erstynakth@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

PENERAPAN SUPPORT VECTOR MACHINE UNTUK MENGANALISIS
SENTIMEN MASYARAKAT TERHADAP PROGRAM MAGANG DAN
STUDI INDEPENDEN BERSEKERTIPAT

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 29 Agustus 2023

Penulis

(Erstyna Kathlya P)
nama dan tanda tangan

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT atas limpahan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi yang berjudul “*Penerapan Support Vector Machine* untuk Menganalisis Sentimen Masyarakat terhadap Program Magang dan Studi Independen Bersertifikat” disusun untuk melengkapi syarat mendapatkan gelar sarjana pada Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta.

Selama penyusunan skripsi ini, penulis memperoleh banyak bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, atas berkat rahmat-Nya yang memberikan penulis kesehatan, niat, dan pengetahuan untuk menyelesaikan skripsi ini.
2. Keluarga penulis, Ayah, Mama, dan adik tersayang, Ridzka, yang senantiasa memberikan semangat, doa, dan dukungannya kepada penulis.
3. Ibu Dr. Lukita Ambarwati, S.Pd., M.Si., selaku Koordinator Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta.
4. Bapak Drs. Sudarwanto, M.Si., DEA. selaku Dosen Pembimbing Akademik (PA), serta Ibu Dr. Dian Handayani, M.Si., dan Ibu Devi Eka Wardani Meganingtyas, S.Pd., M.Si. selaku Dosen Pembimbing I dan Dosen Pembimbing II yang telah meluangkan waktu, ilmu, dan tenaga untuk memberikan bimbingan, kritik, serta saran kepada penulis untuk bisa menyelesaikan skripsi ini dengan baik.
5. Bapak/Ibu dosen pengajar serta seluruh staf dan administrasi di lingkungan Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta.

6. Semua teman-teman mahasiswa/i Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta angkatan 2019, terutama Yasmin Raihan Nabilah, Ghifari Ismail, Dea Febriliana Fajri, Irvan Rizky Kurniawan, Dinda Triana, Nisya Kusuma Putri, serta Wulandini yang selalu memberikan motivasi, dorongan, dan semangat kepada penulis.

Penulis sadar bahwa masih banyak kekurangan yang terdapat pada skripsi ini karena keterbatasan kemampuan penulis. Oleh karena itu, penulis terbuka untuk kritik dan saran yang diberikan dalam rangka menyempurnakan skripsi ini. Penulis berharap skripsi ini dapat bermanfaat bagi pembaca.

Jakarta, Juli 2023
Erstyna Kathlya Prabowo



DAFTAR ISI

DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	6
1.4 Batasan Masalah	6
1.5 Manfaat Penelitian	7
BAB II KAJIAN PUSTAKA	8
2.1 Program Magang dan Studi Independen Bersertifikat (MSIB)	8
2.2 Media Sosial Twitter	9
2.3 Analisis Sentimen	9
2.4 <i>Machine Learning</i>	10
2.5 <i>Text Mining</i>	14
2.5.1 <i>Text Preprocessing</i>	15
2.6 Pembobotan Kata	19
2.6.1 Contoh Pembobotan Kata dengan TF-IDF	21
2.7 Klasifikasi	23
2.7.1 Pencarian Grid (<i>Grid Search</i>)	24
2.7.2 Validasi Silang <i>K-Lipat (K-Fold Cross Validation)</i>	25
2.8 <i>Support Vector Machine</i>	27
2.8.1 Data yang Dapat Dipisahkan secara Linier (<i>Linearly Separable Data</i>)	32
2.8.2 Data yang Tidak Dapat Dipisahkan secara Linier (<i>Nonlinearly Separable Data</i>)	38
2.8.3 Kernel	41
2.8.4 <i>Multiclass Support Vector Machine</i>	45
2.8.5 Keunggulan dan Keterbatasan <i>Support Vector Machine (SVM)</i>	48
2.9 Pengukuran Kinerja Klasifikasi	49
2.9.1 Contoh Pengukuran Kinerja Model Klasifikasi (Kasus Multikelas)	52
BAB III METODOLOGI PENELITIAN	55
3.1 Data dan Sumber Data	55
3.2 Prosedur Penelitian	55

3.3	Diagram Alir	58
BAB IV HASIL DAN PEMBAHASAN		59
4.1	<i>Scraping</i> Data	59
4.2	<i>Preprocessing</i> Data	62
4.2.1	<i>Text Preprocessing</i>	63
4.3	Pengambilan Sampel	69
4.3.1	Pelabelan Sentimen Manual pada Data Sampel	72
4.3.2	Pembobotan Kata	73
4.4	Pembagian Data (<i>Data Splitting</i>)	75
4.5	Pembangunan Model SVM	76
4.6	Pemilihan Model Terbaik	80
4.7	Interpretasi Model	83
4.7.1	Model SVM Biner 1: Klasifikasi Sentimen Negatif-Netral	84
4.7.2	Model SVM Biner 2: Klasifikasi Sentimen Negatif-Positif	88
4.7.3	Model SVM Biner 3: Klasifikasi Sentimen Netral-Positif	92
4.7.4	Klasifikasi Sentimen dengan Model SVM	96
4.8	Analisis Hasil Klasifikasi	97
BAB V KESIMPULAN DAN SARAN		104
5.1	Kesimpulan	104
5.2	Saran	105
DAFTAR PUSTAKA		106
LAMPIRAN		110
DAFTAR RIWAYAT HIDUP		124

DAFTAR TABEL

Tabel 2.1	Contoh Klasifikasi Sentimen Model Dua Arah	10
Tabel 2.2	Contoh Klasifikasi Sentimen Model Tiga Arah	10
Tabel 2.3	Contoh Hasil Proses <i>Cleansing</i>	15
Tabel 2.4	Contoh Hasil Proses <i>Spelling Normalization</i>	16
Tabel 2.5	Contoh Hasil Proses <i>Stemming</i>	16
Tabel 2.6	Contoh Hasil Proses <i>Case Folding</i>	17
Tabel 2.7	Contoh Hasil Proses <i>Tokenization</i>	18
Tabel 2.8	Contoh Hasil Proses <i>Filtering</i>	18
Tabel 2.9	Contoh Perhitungan Pembobotan Kata	22
Tabel 2.10	Contoh <i>Cross Validation</i> dengan $K = 5$	26
Tabel 2.11	Contoh Model yang Dibangun dengan Metode <i>One to Rest</i>	46
Tabel 2.12	Contoh Model yang Dibangun dengan Metode <i>One-to-One</i>	48
Tabel 2.13	<i>Confusion Matrix</i>	50
Tabel 2.14	Kriteria Nilai <i>Accuracy</i>	50
Tabel 2.15	Kriteria Nilai <i>F1 Score</i>	52
Tabel 2.16	Penggunaan Metrik Kinerja Klasifikasi	52
Tabel 2.17	Contoh <i>Confusion Matrix</i>	53
Tabel 2.18	Kesimpulan <i>Confusion Matrix</i>	53
Tabel 4.1	Fitur "Snsrape"	60
Tabel 4.2	Pengambilan Data Topik MSIB pada Twitter	60
Tabel 4.3	Contoh Data Hasil <i>Scraping</i>	61
Tabel 4.4	Data yang Mengandung <i>Tweet</i> Tidak Relevan	63
Tabel 4.5	Hasil Tahap <i>Cleansing</i>	64
Tabel 4.6	Hasil Tahap <i>Spelling Normalization</i>	65
Tabel 4.7	Hasil Tahap <i>Stemming</i>	66
Tabel 4.8	Hasil Tahap <i>Case Folding</i>	66
Tabel 4.9	Hasil Tahap <i>Case Folding</i>	67
Tabel 4.10	Hasil Tahap <i>Filtering</i>	69
Tabel 4.11	Contoh Data Hasil <i>Sampling</i>	71
Tabel 4.12	Hasil Pelabelan Sentimen Secara Manual	72
Tabel 4.13	Hasil Tahap Pembobotan Kata pada Salah Satu <i>Tweets</i>	74
Tabel 4.14	Matriks Pembobotan Kata	75
Tabel 4.15	Pembagian Data Latih dan Data Uji	76
Tabel 4.16	Kandidat <i>Hyperparameter</i>	77
Tabel 4.17	Kandidat Model SVM Terbaik	78
Tabel 4.18	Hasil <i>Stratified K-Fold CV</i> pada Kandidat Model	80
Tabel 4.19	Metrik Kinerja Klasifikasi Model SVM	82
Tabel 4.20	Model SVM Biner Metode <i>One-to-One</i>	83

Tabel 4.21	Tabel Nilai α_i untuk Model SVM Klasifikasi Sentimen Negatif-Netral	87
Tabel 4.22	Tabel Nilai α_i untuk Model SVM Klasifikasi Sentimen Negatif-Positif	91
Tabel 4.23	Tabel Nilai α_i untuk Model SVM Klasifikasi Sentimen Netral-Positif	95
Tabel 4.24	Hasil Klasifikasi dari Fungsi Keputusan	96
Tabel 4.25	Contoh Data yang Diklasifikasikan sebagai Sentimen Positif, Negatif, dan Netral	98
Tabel 4.26	Hasil Klasifikasi Sentimen	102



DAFTAR GAMBAR

Gambar 2.1	Pencarian Grid (<i>Grid Search</i>)	24
Gambar 2.2	Ilustrasi <i>5-Fold Cross Validation</i>	26
Gambar 2.3	Kandidat <i>Hyperplane</i> di \mathbb{R}^2	27
Gambar 2.4	Ilustrasi <i>Hyperplane</i> di \mathbb{R}^2	30
Gambar 2.5	<i>Hard Margin SVM</i>	33
Gambar 2.6	<i>Soft Margin SVM</i>	39
Gambar 2.7	Pemetaan Data ke Ruang Berdimensi Lebih Tinggi (\mathbb{R}^2 ke \mathbb{R}^3)	42
Gambar 2.8	<i>One-To-Rest</i>	46
Gambar 2.9	<i>One-To-One</i>	47
Gambar 4.1	Proporsi Masing-masing Sentimen pada Data Sampel	73
Gambar 4.2	<i>Confusion Matrix</i> Model SVM Setiap Kernel	81
Gambar 4.3	Distribusi Sentimen	97
Gambar 4.4	Jumlah Kata dan Karakter dalam Tweets	99
Gambar 4.5	<i>Wordcloud</i> Sentimen	99
Gambar 4.6	Jumlah <i>Tweets</i> Berdasarkan Tahun dan Bulan	100
Gambar 4.7	Frekuensi <i>Tweets</i> Harian	101