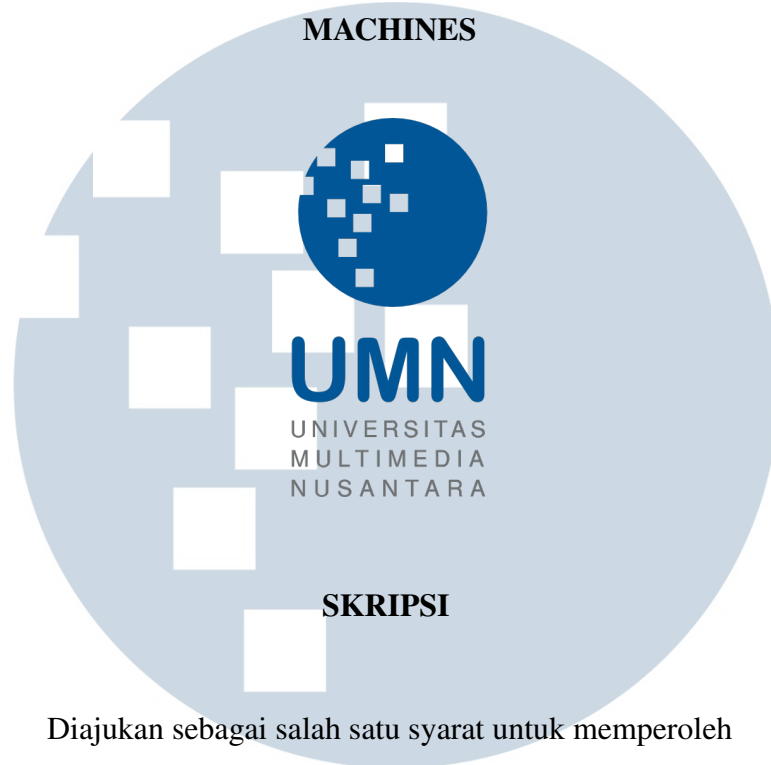


**PREDIKSI TIPE KEPERIBADIAN MBTI BERDASARKAN KONTEN  
MEDIA SOSIAL MENGGUNAKAN ALGORITMA SUPPORT VECTOR  
MACHINES**



**SKRIPSI**

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

**Jessica Valeria**

**00000027629**

UMN

UNIVERSITAS

MULTIMEDIA

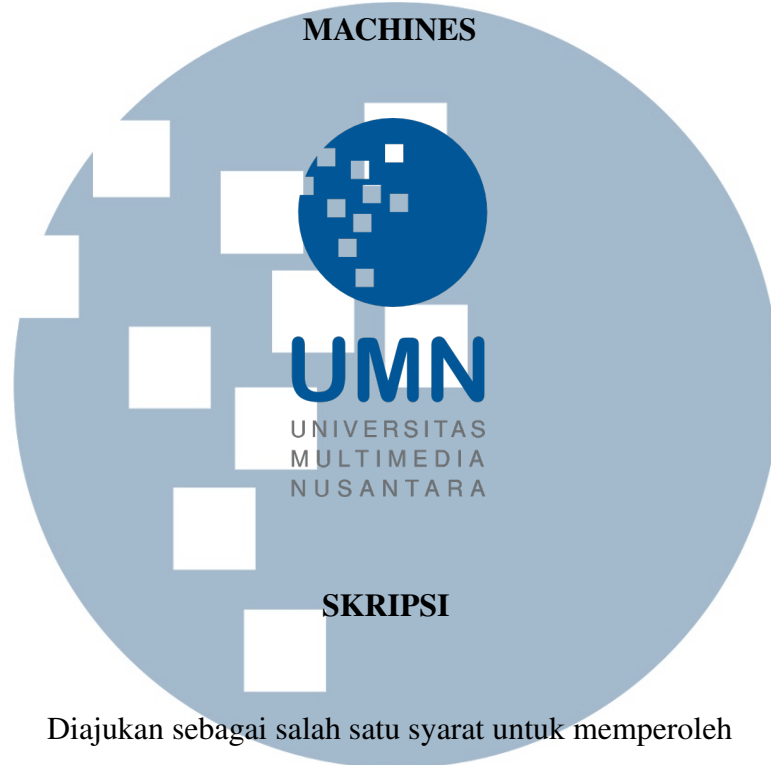
NUSANTARA

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

**TANGERANG**

**2022**

**PREDIKSI TIPE KEPRIBADIAN MBTI BERDASARKAN KONTEN  
MEDIA SOSIAL MENGGUNAKAN ALGORITMA SUPPORT VECTOR  
MACHINES**



**SKRIPSI**

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

**Jessica Valeria**  
**0000027629**

UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

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

**TANGERANG**

**2022**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Jessica Valeria

Nomor Induk Mahasiswa : 00000027629

Program Studi : Informatika

Skripsi dengan judul:

Prediksi Tipe Kepribadian MBTI Berdasarkan Konten Media Sosial Menggunakan Algoritma Support Vector Machines

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, 17 Juni 2022



(Jessica Valeria)

UMM  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

Skripsi dengan judul

### **PREDIKSI TIPE KEPRIBADIAN MBTI BERDASARKAN KONTEN MEDIA SOSIAL MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINES**

oleh

Nama : Jessica Valeria  
NIM : 00000027629  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 30 Juni 2022

Pukul 10.00 s/d 12.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Adhi Kusnadi, S.T, M.Si.)

NIDN: 303037304

Penguji



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

NIDN: 0407128901

Pembimbing



(Alethea Suryadibrata, S.Kom., M.Eng.)

NIDN: 0322099201

Ketua Program Studi Informatika,

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

NIDN: 0818038501

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

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Jessica Valeria  
NIM : 00000027629  
Program Studi : Informatika  
Fakultas : Teknik dan Informatika  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**PREDIKSI TIPE KEPRIBADIAN MBTI BERDASARKAN KONTEN  
MEDIA SOSIAL MENGGUNAKAN ALGORITMA SUPPORT VECTOR  
MACHINES**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 17 Juni 2022  
Yang menyatakan

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Jessica Valeria

**Halaman Persembahan / Motto**

"Learn as if you will live forever, live like you will die tomorrow."

Mahatma Gandhi



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Prediksi Tipe Kepribadian MBTI Berdasarkan Konten Media Sosial Menggunakan Algoritma Support Vector Machines 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. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Alethea Suryadibrata, S.Kom., M.Eng., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
5. Orang tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.
6. Teman-teman yang telah memberikan bantuan, semangat, serta informasi mengenai pengerjaan skripsi.

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

Tangerang, 17 Juni 2022



Jessica Valeria

**PREDIKSI TIPE KEPERIBADIAN MBTI BERDASARKAN KONTEN  
MEDIA SOSIAL MENGGUNAKAN ALGORITMA SUPPORT VECTOR  
MACHINES**

Jessica Valeria

**ABSTRAK**

Perkembangan industri 4.0 mengadakan salah satu inovasi hebat yaitu kecerdasan buatan yang memungkinkan adanya sistem untuk analisis, *modelling*, komputasi, dan pengambilan keputusan. Sistem ini kemudian disebut pembelajaran mesin yang dapat diterapkan ke berbagai bidang, salah satunya yaitu bidang psikologi. Kepribadian merupakan bagian terkompleks seseorang yang menandakan keunikan setiap individu. Menurut teori Myers-Briggs Type Indicator (MBTI), perilaku manusia merupakan perilaku yang teratur dan konsisten. Hal ini menyiratkan bahwa perilaku manusia dapat diprediksi dengan sebuah pola. Dengan pembelajaran mesin, sistem diharapkan dapat mempelajari pola tersebut. Data yang akan digunakan dalam pelatihan model merupakan data konten yang ditulis pengguna dengan label tipe MBTI masing-masing. Dengan algoritma Support Vector Machines, teknik ekstraksi fitur TF-IDF, dan *oversampling* dengan SMOTE, model akan mengolah dan mempelajari data tersebut. Hasil penelitian menunjukkan bahwa model yang diuji dapat menghasilkan tingkat akurasi prediksi sebesar 93,4%. Kesimpulannya, pengimplementasian algoritma Support Vector Machines, teknik ekstraksi fitur TF-IDF, dan *oversampling* dengan SMOTE dalam memprediksi tipe kepribadian MBTI berdasarkan konten telah selesai.

**Kata kunci:** MBTI, Pembelajaran Mesin, SMOTE, Support Vector Machines, TF-IDF

U M 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



# **MBTI Personality Type Prediction Based on Social Media Content Using Support Vector Machines Algorithm**

Jessica Valeria

## **ABSTRACT**

The development of industry 4.0 has brought one of the greatest innovation, which is artificial intelligence. Artificial intelligence allows systems to do analysis, modelling, computing, and decision making. This system is then called machine learning which can be applied to various fields, one of which is the field of psychology. Personality is the most complex part of a human that indicates uniqueness of each individual. According to the Myers-Briggs Type Indicator (MBTI) theory, human behavior is a regular and consistent behavior. This implies that human behavior can be predicted with a pattern. With machine learning, the system is expected to learn the pattern. The data that will be used in the model training is user-written content data with their respective MBTI type labels. With the Support Vector Machines algorithm, TF-IDF feature extraction technique, and oversampling with SMOTE, the model will process and learn the data. The results showed that the tested model could produce a prediction accuracy rate of 93.4%. In conclusion, the implementation of Support Vector Machines algorithm, TF-IDF feature extraction technique, and oversampling with SMOTE in predicting MBTI personality type based on content has been completed.

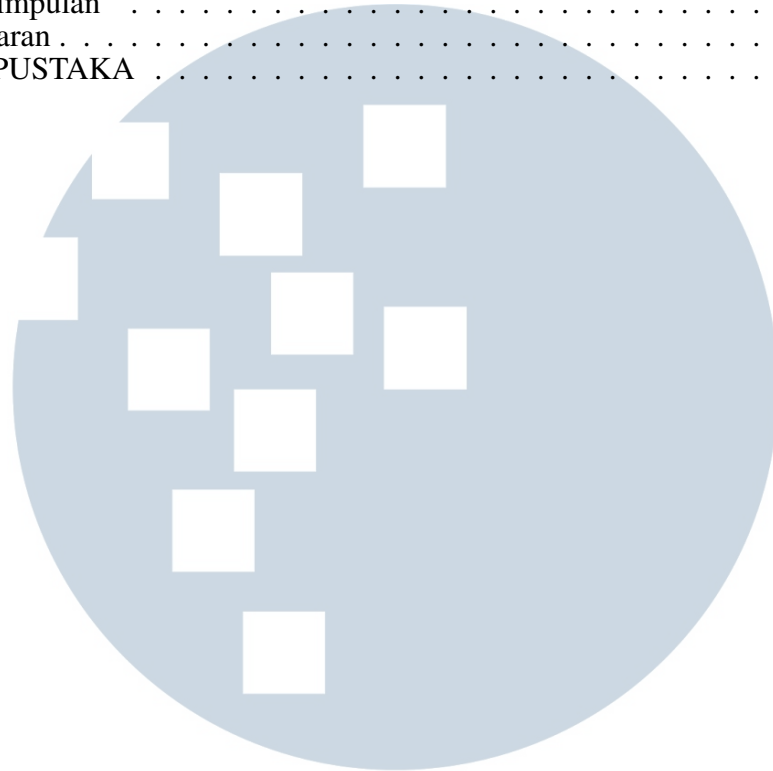
**Keywords:** Machine Learning, MBTI, SMOTE, Support Vector Machines, TF-IDF



## 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 LAMPIRAN . . . . .	xii
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 Myers-Briggs Type Indicator (MBTI) . . . . .	5
2.2 Support Vector Machines . . . . .	6
2.3 Preprocessing . . . . .	9
2.4 TF-IDF . . . . .	10
2.5 Synthetic Minority Oversampling Technique . . . . .	11
2.6 Evaluation Metrics . . . . .	13
BAB 3 METODOLOGI PENELITIAN . . . . .	15
3.1 Metode Penelitian . . . . .	15
3.1.1 Studi Literatur . . . . .	15
3.1.2 Pencarian Dataset . . . . .	15
3.1.3 Perancangan Model . . . . .	15
3.1.4 Implementasi . . . . .	16
3.1.5 Pengujian dan Evaluasi . . . . .	16
3.1.6 Penulisan Laporan . . . . .	17
3.2 Perancangan Program . . . . .	17
3.2.1 Preprocessing . . . . .	18
3.2.2 Feature Extraction . . . . .	20
3.2.3 SMOTE . . . . .	20
3.2.4 Data Splitting . . . . .	21
3.2.5 Modelling . . . . .	22
3.2.6 Evaluation . . . . .	23
BAB 4 HASIL DAN DISKUSI . . . . .	24
4.1 Spesifikasi Sistem . . . . .	24
4.2 Implementasi . . . . .	24
4.2.1 Data Preprocessing . . . . .	25
4.2.2 Implementasi TF-IDF . . . . .	32
4.2.3 Implementasi SMOTE . . . . .	33
4.2.4 Data Splitting . . . . .	33
4.2.5 Pelatihan Model dengan Support Vector Machines . . . . .	34

4.3	Evaluasi . . . . .	34
BAB 5	SIMPULAN DAN SARAN . . . . .	37
5.1	Simpulan . . . . .	37
5.2	Saran . . . . .	37
DAFTAR PUSTAKA	. . . . .	38



UMMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

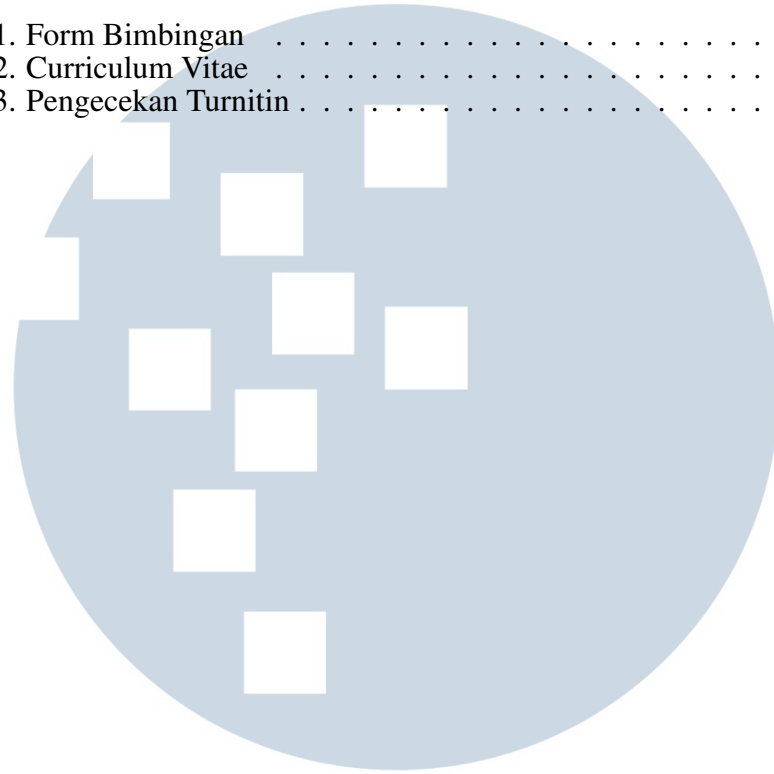
## DAFTAR GAMBAR

Gambar 2.1	<i>Hyperplane</i> 2D terbaik . . . . .	7
Gambar 3.1	<i>Flowchart</i> utama . . . . .	17
Gambar 3.2	<i>Flowchart preprocessing</i> . . . . .	18
Gambar 3.3	<i>Flowchart remove</i> MBTI . . . . .	19
Gambar 3.4	<i>Flowchart feature extraction</i> . . . . .	20
Gambar 3.5	<i>Flowchart SMOTE</i> . . . . .	21
Gambar 3.6	<i>Flowchart modelling</i> . . . . .	22
Gambar 3.7	<i>Flowchart evaluation</i> . . . . .	23
Gambar 4.1	Import library yang diperlukan . . . . .	24
Gambar 4.2	<i>Dataset</i> yang digunakan . . . . .	25
Gambar 4.3	Jumlah data pada setiap kelas . . . . .	25
Gambar 4.4	Mengecek data <i>null</i> . . . . .	26
Gambar 4.5	Menghapus duplikasi data . . . . .	26
Gambar 4.6	Inisiasi <i>stop words</i> . . . . .	27
Gambar 4.7	Implementasi <i>preprocessing</i> . . . . .	27
Gambar 4.8	Hasil penghapusan URL . . . . .	28
Gambar 4.9	Hasil <i>case folding</i> . . . . .	28
Gambar 4.10	Penghapusan <i>stop words</i> . . . . .	29
Gambar 4.11	Contoh sebelum penghapusan simbol . . . . .	29
Gambar 4.12	Contoh setelah penghapusan simbol . . . . .	30
Gambar 4.13	Contoh sebelum penghapusan karakter tunggal . . . . .	30
Gambar 4.14	Contoh setelah penghapusan karakter tunggal . . . . .	30
Gambar 4.15	Hasil implementasi penghapusan beberapa spasi . . . . .	31
Gambar 4.16	Hasil implementasi penghapusan tipe kepribadian MBTI . . . . .	31
Gambar 4.17	Contoh <i>function</i> penghapusan kata yang duplikat . . . . .	32
Gambar 4.18	Hasil implementasi penghapusan kata yang duplikat . . . . .	32
Gambar 4.19	Inisiasi data ke variabel x dan y . . . . .	32
Gambar 4.20	Implementasi TF-IDF . . . . .	33
Gambar 4.21	Implementasi SMOTE . . . . .	33
Gambar 4.22	<i>Data splitting</i> . . . . .	33
Gambar 4.23	Pelatihan model dengan SVM . . . . .	34
Gambar 4.24	Potongan kode untuk memperoleh <i>classification report</i> . . . . .	35
Gambar 4.25	Hasil <i>confusion matrix</i> . . . . .	35

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 LAMPIRAN

Lampiran 1. Form Bimbingan . . . . .	40
Lampiran 2. Curriculum Vitae . . . . .	41
Lampiran 3. Pengecekan Turnitin . . . . .	42



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA