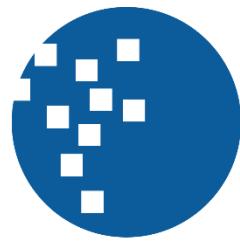


**PENINGKATAN KINERJA MODEL *MACHINE LEARNING*
DALAM *SENTIMENT ANALYSIS CRYPTOCURRENCY*
MELALUI PENDEKATAN *FEATURE SELECTION*
BERBASIS *SWARM INTELLIGENCE***



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

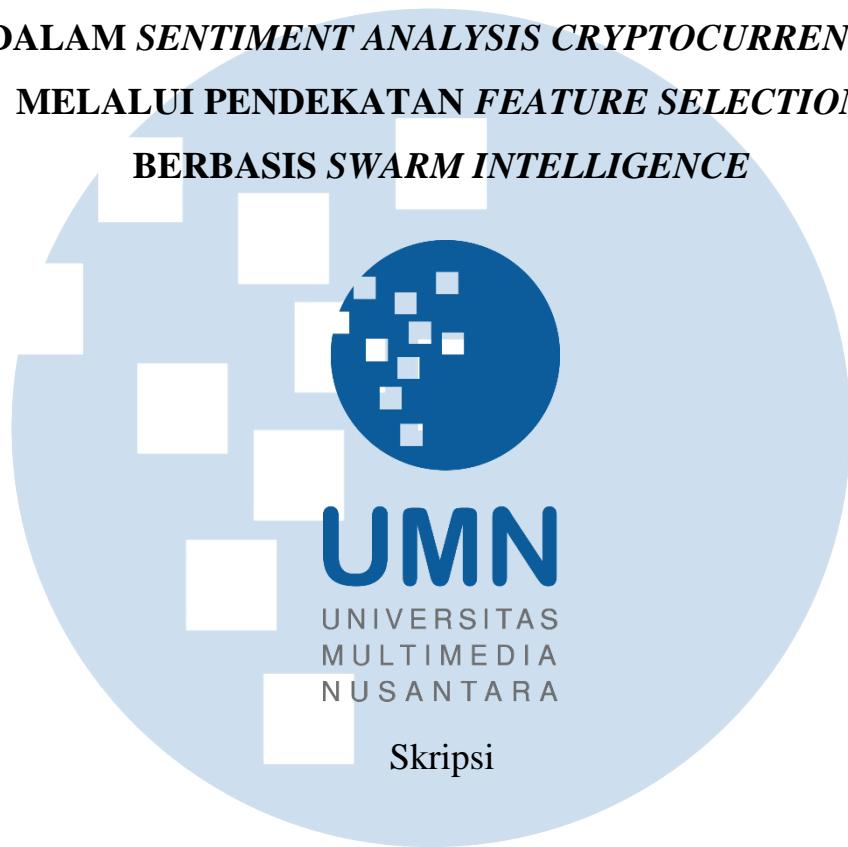
Skripsi

Jason Suhali

00000045379

**PROGRAM STUDI SISTEM INFORMASI
TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**PENINGKATAN KINERJA MODEL MACHINE LEARNING
DALAM SENTIMENT ANALYSIS CRYPTOCURRENCY
MELALUI PENDEKATAN FEATURE SELECTION
BERBASIS SWARM INTELLIGENCE**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom.)

Jason Suhali

00000045379

PROGRAM STUDI SISTEM INFORMASI
TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Jason Suhali
Nomor Induk Mahasiswa : 00000045379
Program studi : Sistem Informasi

Skripsi dengan judul:

Peningkatan Kinerja Model *Machine Learning* Dalam *Sentiment Analysis Cryptocurrency* Melalui Pendekatan *Feature Selection* Berbasis *Swarm Intelligence*

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, Mei 2024



(Jason Suhali)

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

PENINGKATAN KINERJA MODEL MACHINE LEARNING DALAM
SENTIMENT ANALYSIS CRYPTOCURRENCY MELALUI PENDEKATAN
FEATURE SELECTION BERBASIS SWARM INTELLIGENCE

Oleh

Nama : Jason Suhali
NIM : 00000045379
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Kamis, 30 Mei 2024

Pukul 08.00 s.d 10.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang

Rudi Sutomo, S.Kom., M.Si., M.Kom.
NIDN: 0222057501/NIK: 072470

Penguji

Samuel Ady Sanjaya, S.T., M.T.
NIDN: 0305049402/NIK: 075049

Pembimbing

Dinar Ajeng Kristiyanti, S.Kom., M.Kom.
NIDN: 0330128801/NIK: 079159

Ketua Program Studi Sistem Informasi

Ririn Ikana Desanti, S.Kom., M.Kom.
NIDN: 0313058001/NIK: 051314

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan dibawah ini:

Nama : Jason Suhali
Nomor Induk Mahasiswa : 00000045379
Program Studi : Sistem Informasi
Jenjang : S2 / S1 / D3
Judul Karya Ilmiah :

PENINGKATAN KINERJA MODEL MACHINE LEARNING DALAM SENTIMENT ANALYSIS CRYPTOCURRENCY MELALUI PENDEKATAN FEATURE SELECTION BERBASIS SWARM INTELLIGENCE

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, dikarenakan:

Dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) *.

Tangerang, 30 Mei 2024


(Jason Suhali)

* Jika tidak bisa membuktikan LoA jurnal/HKI selama 6 bulan kedepan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa atas terselesaikannya skripsi untuk pemenuhan laporan skripsi. Tujuan dari pembuatan penelitian adalah untuk dapat menganalisis sentimen masyarakat terhadap *cryptocurrency* dengan melakukan optimasi pada model *machine learning* menggunakan teknik seleksi fitur berbasis *swarm intelligence*. Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna, karena itu penulis ingin mengucapkan terima kasih kepada beberapa pihak yang telah memberikan bantuan, bimbingan, dan dukungan selama proses penyusunan laporan skripsi ini. Oleh sebab itu, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Dr. Ninok Leksono, M.A., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, ST, M.Sc., selaku Dekan Fakultas Teknik dan Infomatika Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Dinar Ajeng Kristiyanti, S.Kom, M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Rekan-rekan sejawat lainnya yang tidak bisa disebutkan satu per satu.

Semoga karya ilmiah ini dapat menjadi inspirasi, acuan, dan pedoman bagi para peneliti selanjutnya dalam mengembangkan penelitiannya.

Tangerang, 13 Mei 2024



(Jason Suhali)

UNIVERSITAS
MULTIMEDIA
NUSANTARA

**PENINGKATAN KINERJA MODEL MACHINE LEARNING
DALAM SENTIMENT ANALYSIS CRYPTOCURRENCY
MELALUI PENDEKATAN FEATURE SELECTION
BERBASIS SWARM INTELLIGENCE**

(Jason Suhali)

ABSTRAK

Cryptocurrency merupakan mata uang digital yang dapat dijadikan sebagai alat untuk bertransaksi di seluruh dunia. Pakar ekonomi dunia memprediksi bahwa tren *cryptocurrency* akan mengalami kenaikan dari tahun sebelumnya. Hal tersebut didukung dengan penyetujuan kebijakan Bitcoin Spot Exchange Trade Funds (ETF) oleh Komisi Sekuritas dan Bursa Amerika Serikat. Namun, terdapat masyarakat yang skeptis atau memilih untuk menahan diri pada tren *cryptocurrency* tersebut. Hal tersebut dikarenakan tidak pernah ada pernyataan resmi dari lembaga manapun dan siapa yang mengatur alur perkembangan *cryptocurrency*. Penelitian ini menggunakan *framework* KDD. Penelitian dimulai dengan melakukan teknik *web scrapping* untuk mengumpulkan dataset dari *platform* X. Setelah data didapatkan, dilakukan *preprocessing data* yang meliputi *data cleansing*, *labeling*, pembobotan kata, dan SMOTE. Selanjutnya, dilakukan *modeling* dengan algoritma *machine learning* SVM, KNN, dan Naïve Bayes. Setelah itu, dilakukan optimasi *feature selection* menggunakan PSO, ACO, dan CSO terhadap semua model *machine learning*. Penelitian ini melakukan optimasi model *machine learning* pada peningkatan akurasi, *classification report*, dan *time execution*. Hasil penelitian mengungkapkan bahwa algoritma ACO-SVM memiliki tingkat akurasi tertinggi, yakni 86,95%. Terjadi peningkatan akurasi sebesar 2% dari yang sebelumnya adalah 84,91%. Waktu pemrosesan pemodelan juga mengalami percepatan yang signifikan, dari sebelumnya 5.416 detik menjadi 0,882 detik. Oleh karena itu, penggunaan algoritma optimasi *feature selection* berbasis *swarm intelligence* terbukti dapat meningkatkan kinerja model *machine learning*.

Kata kunci: *Cryptocurrency, Feature Selection , Machine Learning, Sentiment Analysis, Swarm Intelligence*

UNIVERSITAS
MULTIMEDIA
NUSANTARA

**IMPROVING MACHINE LEARNING MODEL PERFORMANCE
IN CRYPTOCURRENCY SENTIMENT ANALYSIS
THROUGH A FEATURE SELECTION APPROACH**

SWARM INTELLIGENCE BASED

(Jason Suhali)

ABSTRACT (English)

Cryptocurrency is a digital currency that can be used as a tool for transactions throughout the world. World economic experts predict the cryptocurrency trend will increase from the previous year. This is supported by the approval of the Bitcoin Spot Exchange Trade Funds (ETF) policy by the United States Securities and Exchange Commission. However, some people are skeptical or refrain from this cryptocurrency trend. This is because there has never been an official statement from any institution regulating cryptocurrency development flow. This research uses the KDD framework. The research began by carrying out web scrapping techniques to collect datasets from platform X. After the data was obtained, data preprocessing was carried out which included data cleaning, labeling, word weighting, and SMOTE. Next, modeling was carried out using the SVM, KNN, and Naïve Bayes machine learning algorithms. Afterward, all machine learning models were optimized for feature selection using PSO, ACO, and CSO. This research optimizes the machine learning model to increase accuracy, classification report, and time execution. The research results revealed that the ACO-SVM algorithm had the highest level of accuracy, namely 86.95%. There was an increase in accuracy of 2% from the previous 84.91%. Modeling processing time has also experienced a significant acceleration, from the previous 5,416 seconds to 0.882 seconds. Therefore, using a feature selection optimization algorithm based on swarm intelligence has been proven to improve the performance of machine learning models.

Keywords: Cryptocurrency, Feature Selection , Machine Learning, Sentiment Analysis, Swarm Intelligence

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
LEMBAR PERSETUJUAN PUBLIKASI	iii
KARYA ILMIAH MAHASISWA	iii
KATA PENGANTAR	iv
ABSTRAK	v
<i>ABSTRACT (English)</i>	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	7
1.4 Tujuan dan Manfaat Penelitian	7
1.4.1 Tujuan Penelitian	7
1.4.2 Manfaat Penelitian	8
1.5 Sistematika Penulisan	9
BAB II LANDASAN TEORI	10
2.1 Penelitian Terdahulu	10
2.2 Teori tentang Topik Skripsi	12
2.2.1 <i>Cryptocurrency</i>	12
2.2.2 X	14
2.2.3 <i>Web Scraping</i>	14
2.2.4 <i>Sentiment Analysis</i>	15
2.2.5 <i>Text Preprocessing</i>	15
2.2.6 VADER	17
2.2.7 TF-IDF	17
2.2.8 SMOTE	17
2.2.9 <i>Feature Selection</i>	18

2.3 Teori tentang Framework / Algoritma yang digunakan	19
2.3.1 KDD	19
2.3.2 Grid Search	20
2.3.3 Machine Learning	20
2.3.4 SVM	20
2.3.5 Naïve Bayes	22
2.3.6 KNN (K-Nearest Neighbour).....	23
2.3.7 Swarm Intelligence	25
2.3.8 PSO (Particle Swarm Intelligence).....	26
2.3.9 ACO (Ant Colony Optimization)	26
2.3.10 CSO (Cat Swarm Optimization).....	27
2.3.11 Evaluation Metrics	28
2.3.12 Accuracy	29
2.3.13 Precision	29
2.3.14 Recall	29
2.3.15 F1-score	30
2.4 Teori tentang Tools / Software yang digunakan	30
2.4.1 Python.....	30
2.4.2 Google Collab.....	30
2.4.3 Jupyter Lab.....	31
BAB III METODOLOGI PENELITIAN	32
3.1 Gambaran Umum Objek Penelitian	32
3.2 Metode Penelitian	32
3.2.1 Metode Data Mining.....	33
3.2.2 Alur Penelitian	35
3.3 Teknik Pengumpulan Data	38
3.3.1 Populasi dan Sampel.....	39
3.3.2 Periode Pengambilan Data	39
3.4 Variabel Penelitian	40
3.5 Teknik Analisis Data	40
BAB IV ANALISIS DAN HASIL PENELITIAN	42
4.1 Analisa Masalah dan Kebutuhan Penelitian	42

4.1.1	Proses Bisnis.....	42
4.1.2	Spesifikasi Sistem.....	42
4.2	Data Selection	43
4.3	Data Preprocessing.....	46
4.3.1	<i>Remove Duplicate</i>	47
4.3.2	<i>Remove URLs</i>	47
4.3.3	<i>Remove Hashtags</i>	48
4.3.4	<i>Remove Special Characters, Numbers, and Exclude Punctuation</i>	49
4.3.5	<i>Convert to Lower Case</i>	50
4.3.6	<i>Remove Stop Words</i>	51
4.3.7	<i>Lemmatization of Words</i>	52
4.3.8	<i>Remove Words with Length < 3</i>	53
4.3.9	<i>Join Words</i>	53
4.4	Data Transformation.....	54
4.4.1	<i>Labeling Data Using VADER</i>	54
4.4.2	TF-IDF	56
4.4.3	SMOTE	59
4.5	Data Mining	62
4.5.1	<i>Grid Search</i>	62
4.5.2	<i>Machine Learning</i>	67
4.5.3	Optimasi Feature Selection	75
4.6	Evaluation	97
4.6.1	<i>Hasil Classification Report</i>	98
4.6.2	Uji Statistik Shapiro Test	99
4.6.3	Uji Statistik T-Test	100
4.7	Hasil dan Diskusi	101
BAB V	SIMPULAN DAN SARAN	104
5.1	Simpulan	104
5.2	Saran	105
DAFTAR PUSTAKA		107
LAMPIRAN		118

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	10
Tabel 3. 1 Perbandingan Metode <i>Framework Data Mining</i>	33
Tabel 3. 2 Daftar Atribut Data.....	38
Tabel 4. 1 Perbandingan <i>Before-After Remove URLs</i>	48
Tabel 4. 2 Perbandingan <i>Before-After Remove Hashtags</i>	49
Tabel 4. 3 Perbandingan Before-After Remove Special Character, Numbers, and Excluding Punctuation	50
Tabel 4. 4 Perbandingan Before-After Remove Stop Words.....	51
Tabel 4. 5 Perbandingan <i>Before-After Lemmatization of Words</i>	52
Tabel 4. 6 Perbandingan Classification Report Setiap Alogrithma.....	99
Tabel 4. 7 Hasil Uji Statistik Shapiro Test.....	100
Tabel 4. 8 Hasil Uji Statistik T-test	100
Tabel 4. 9 Hasil Model <i>Machine Learning</i>	101
Tabel 4. 10 Hasil Algoritma Optimasi <i>Feature Selection</i> Berbasis <i>Swarm Intelligence</i>	102

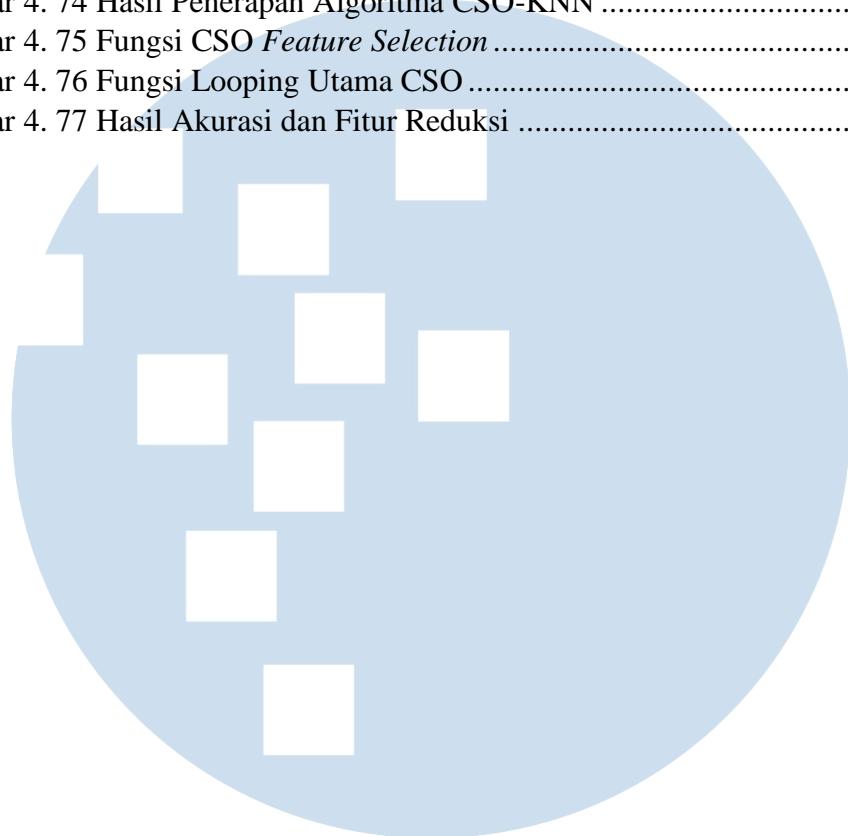


DAFTAR GAMBAR

Gambar 1. 1 Tren Harga <i>Cryptocurrency (Bitcoin)</i> Vs Instrumen Investasi Lainnya.....	2
Gambar 1. 2 Tren Tagar Bitcoin pada Twitter (X) Periode 2022.....	4
Gambar 2. 1 Logo X	14
Gambar 2. 2 Proses Support Vector Machine	21
Gambar 2. 3 Linear Support Vector Machine	22
Gambar 2. 4 Particle Swarm Optimization	26
Gambar 2. 5 Ant Colony Optimization.....	27
Gambar 2. 6 Cat Swarm Optimization.....	28
Gambar 2. 7 Logo Google Colab.....	31
Gambar 3. 1 Alur Penelitian.....	35
Gambar 4. 1 Contoh Pembahasan <i>Cryptocurrency</i>	43
Gambar 4. 2 Twitter/X Auth Token.....	44
Gambar 4. 3 Import Library	44
Gambar 4. 4 Crawl Data.....	45
Gambar 4. 5 Reading Data to Dataframe	45
Gambar 4. 6 File CSV <i>Scrapping Data Tweets X</i>	46
Gambar 4. 7 Remove Duplicate Code	47
Gambar 4. 8 Remove URLs Code	47
Gambar 4. 9 Remove Hashtags Code	48
Gambar 4. 10 Remove Special Characters, Numbers, and Excluding Punctuation Code	49
Gambar 4. 11 Convert To Lower Case Code	50
Gambar 4. 12 Remove Stop Words Code	51
Gambar 4. 13 Lemmatization of Words Code	52
Gambar 4. 14 Remove Words with Length < 3 Code	53
Gambar 4. 15 Join Words Code.....	53
Gambar 4. 16 Hasil Data Preprocessing	54
Gambar 4. 17 Penerapan Labeling VADER	55
Gambar 4. 18 Persebaran Sentimen Positif dan Negatif	56
Gambar 4. 19 Ekstrasi Fitur Menggunakan TF-IDF.....	57
Gambar 4. 20 Mapping Sentiment ke Nilai Numerik	57
Gambar 4. 21 Menggabungkan Fitur TF-IDF dengan Skor Sentimen	58
Gambar 4. 22 Export File XLSX Final.....	58
Gambar 4. 23 Output Export File	59
Gambar 4. 24 Output File Excel.....	59
Gambar 4. 25 Instalasi Library SMOTE	60
Gambar 4. 26 Load Dataset.....	60
Gambar 4. 27 Oversampling SMOTE	61
Gambar 4. 28 Before and After SMOTE.....	62
Gambar 4. 29 Implementasi Grid Search-SVM	63

Gambar 4. 30 Implementasi Grid Search-Naive Bayes	64
Gambar 4. 31 Hasil Implementasi Grid Search-Naive Bayes	66
Gambar 4. 32 Impelementasi <i>Grid Search-KNN</i>	66
Gambar 4. 33 SVM Machine Learning.....	67
Gambar 4. 34 <i>Classification Report</i> SVM	68
Gambar 4. 35 Konversi Matriks Menggunakan Fungsi ToArray.....	69
Gambar 4. 36 <i>Train</i> Algoritma Naive Bayes	69
Gambar 4. 37 Prediksi dan Evaluasi Model <i>Naive Bayes</i>	70
Gambar 4. 38 Print Hasil Eksekusi Model <i>Naive Bayes</i>	71
Gambar 4. 39 Hasil Prediksi Multinomial Naive Bayes	71
Gambar 4. 40 Hasil Prediksi Gaussian Naive Bayes	72
Gambar 4. 41 Hasil Prediksi Bernouli Naive Bayes.....	72
Gambar 4. 42 Skrip Kode Algoritma KNN.....	73
Gambar 4. 43 Hasil Prediksi KNN	74
Gambar 4. 44 Fungsi <i>Fitness Function</i> PSO-SVM	75
Gambar 4. 45 Konfigurasi PSO <i>Feature Selection</i>	75
Gambar 4. 46 Inisialisasi dan Eksekusi PSO	76
Gambar 4. 47 Pencetakan Hasil Optimasi PSO.....	76
Gambar 4. 48 Implementasi PSO-SVM.....	77
Gambar 4. 49 Fitness Function PSO-KNN	78
Gambar 4. 50 Parameter dan Inisialisasi PSO-KNN	78
Gambar 4. 51 Optimasi PSO-KNN.....	79
Gambar 4. 52 Pemilihan Fitur PSO-KNN.....	79
Gambar 4. 53 Hasil Prediksi PSO-KNN	79
Gambar 4. 54 Fitness Function PSO- <i>Naive Bayes</i>	80
Gambar 4. 55 Dimensi Partikel PSO- <i>Naive Bayes</i>	81
Gambar 4. 56 Implementasi Seleksi Fitur PSO-Naive Bayes	81
Gambar 4. 57 Laporan Hasil Implementasi PSO-Naive Bayes.....	82
Gambar 4. 58 Fungsi <i>Evaluate Feature</i> dan <i>Update Feromon</i>	83
Gambar 4. 59 Inisialisasi dan Parameter ACO.....	83
Gambar 4. 60 Implementasi ACO	84
Gambar 4. 61 Laporan Hasil Implementasi ACO-SVM	84
Gambar 4. 62 Fungsi Evaluate Features dan Update Pheromone	85
Gambar 4. 63 Implementasi ACO	86
Gambar 4. 64 Output Pemodelan ACO-KNN	87
Gambar 4. 65 Fungsi <i>Evaluate Features</i> ACO	88
Gambar 4. 66 Fungsi Update Pheromone	88
Gambar 4. 67 Fungsi <i>Ant Colony Optimization</i>	89
Gambar 4. 68 Laporan Klasifikasi dan Fitur Reduksi ACO-Naive Bayes	90
Gambar 4. 69 Inisialisasi Populasi dan Parameter CSO	91
Gambar 4. 70 Fungsi Loop Utama dalam CSO-SVM	91
Gambar 4. 71 Mutasi Algoritma dan Hasil Akhir	92
Gambar 4. 72 Hasil Evaluasi CSO-SVM.....	93

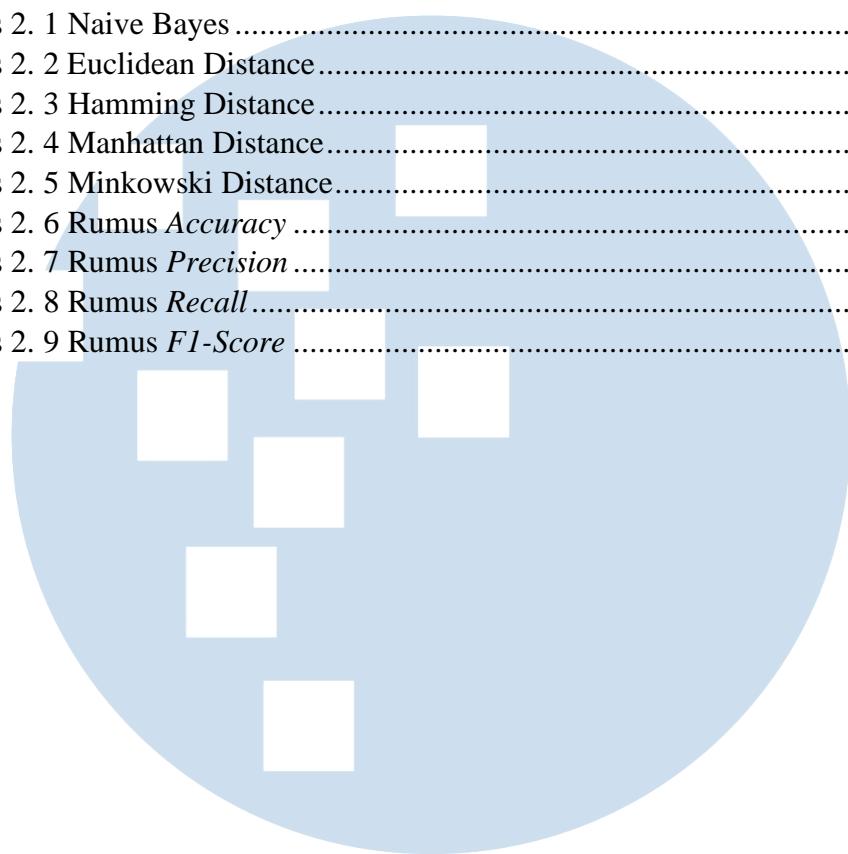
Gambar 4. 73 Implementasi Algoritma CSO-KNN	94
Gambar 4. 74 Hasil Penerapan Algoritma CSO-KNN	95
Gambar 4. 75 Fungsi CSO <i>Feature Selection</i>	96
Gambar 4. 76 Fungsi Looping Utama CSO	96
Gambar 4. 77 Hasil Akurasi dan Fitur Reduksi	97



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

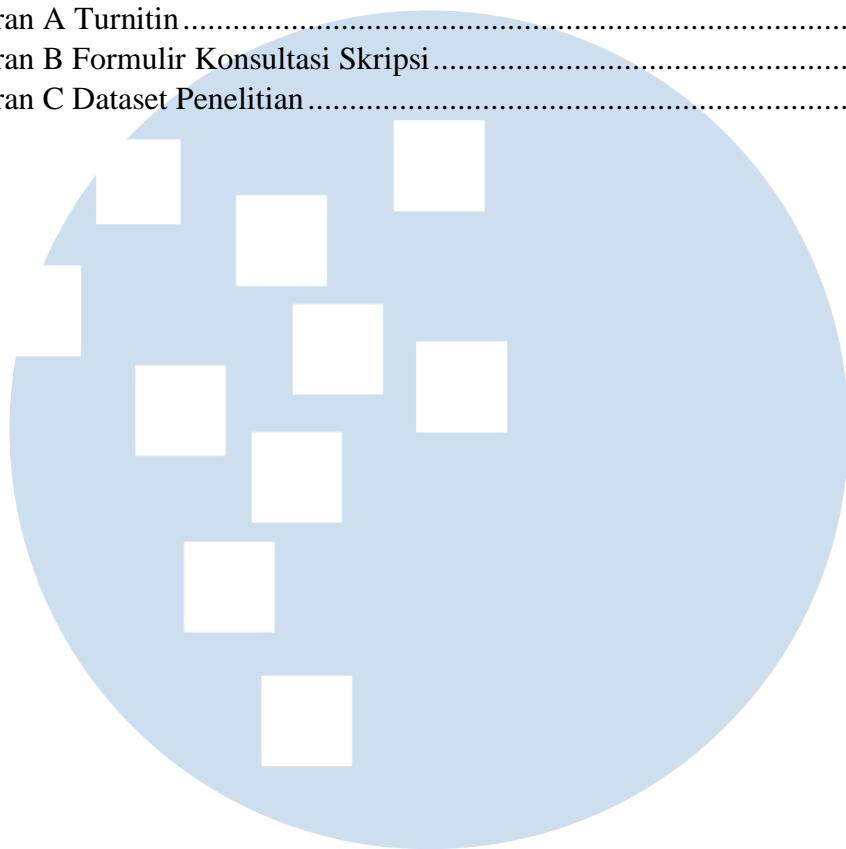
Rumus 2. 1 Naive Bayes	23
Rumus 2. 2 Euclidean Distance	24
Rumus 2. 3 Hamming Distance	25
Rumus 2. 4 Manhattan Distance	25
Rumus 2. 5 Minkowski Distance	25
Rumus 2. 6 Rumus <i>Accuracy</i>	29
Rumus 2. 7 Rumus <i>Precision</i>	29
Rumus 2. 8 Rumus <i>Recall</i>	30
Rumus 2. 9 Rumus <i>F1-Score</i>	30



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Turnitin	118
Lampiran B Formulir Konsultasi Skripsi	130
Lampiran C Dataset Penelitian	132



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA