



**PERBANDINGAN KLASIFIKASI *NAIVE BAYES* DENGAN ATAU  
TANPA *PARTICLE SWARM OPTIMIZATION* UNTUK SELEKSI FITUR  
PADA ANALISIS SENTIMENT TERHADAP PERPRES NO 10 TAHUN 2021  
TENTANG INVESTASI MIRAS PADA MEDIA SOSIAL *TWITTER***

**SKRIPSI**

**GILBERT PANANGIAN DOLOKSARIBU**

**1710511081**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA  
2023**



**PERBANDINGAN KLASIFIKASI *NAIVE BAYES* DENGAN ATAU  
TANPA *PARTICLE SWARM OPTIMIZATION* UNTUK SELEKSI FITUR  
PADA ANALISIS SENTIMENT TERHADAP PERPRES NO 10 TAHUN 2021  
TENTANG INVESTASI MIRAS PADA MEDIA SOSIAL *TWITTER***

**SKRIPSI**

**DIAJUKAN SEBAGAI SALAH SATU SYARAT UNTUK  
MEMPEROLEH GELAR SARJANA KOMPUTER**

**GILBERT PANANGIAN DOLOKSARIBU**

**1710511081**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA**

**2023**

## LEMBAR PERSETUJUAN

Dengan ini menyatakan bahwa skripsi / tugas akhir berikut :

Nama : Gilbert Panangian Doloksaribu  
NIM : 1710511081  
Program Studi : S-1 Informatika 2017  
Judul : Perbandingan Klasifikasi Naive Bayes Dengan Atau Tanpa Particle Swarm Optimization Untuk Seleksi Fitur Pada Analisis Sentiment Terhadap Perpres No 10 Tahun 2021 Tentang Investasi Miras Pada Media Sosial Twitter

Sebagai bagian persyaratan yang diperlukan untuk mengikuti ujian sidang Skripsi / Tugas akhir pada program studi S1 Informatika Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Menyetujui,

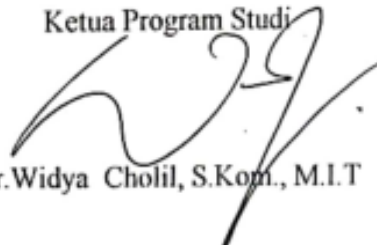
Dosen Pembimbing I



Iin Ernawati S.Kom., M.Si.

Mengetahui,

Ketua Program Studi



Dr. Widya Cholil, S.Kom., M.I.T

Ditetapkan di : Jakarta

Tanggal Pengesahan : 16 Januari 2023

## PERNYATAAN ORISINALITAS

Skripsi ini adalah karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Gilbert Panangian Doloksaribu

NIM : 1710511081

Tanggal : 27 Desember 2022

Bilamana dikemudian hari ditemukan tidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 27 Desember 2022

Yang Menyatakan,

A handwritten signature in black ink is written over a yellow revenue stamp. The stamp features the Garuda Pancasila emblem and the text 'REPUBLIK INDONESIA', '1000', and 'METERAI TEMPEL'. A unique alphanumeric code '73AF2AKX387670918' is printed at the bottom of the stamp.

( Gilbert Panangian Doloksaribu)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK  
KEPENTINGAN AKADEMIS**

---

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Gilbert Panangian Doloksaribu  
NIM : 1710511081  
Fakultas : Ilmu Komputer  
Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

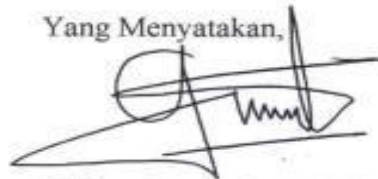
**PERBANDINGAN KLASIFIKASI *NAÏVE BAYES* DENGAN ATAU  
TANPA *PARTICLE SWARM OPTIMIZATION* UNTUK SELEKSI FITUR  
PADA ANALISIS SENTIMENT TERHADAP PERPRES NO 10 TAHUN  
2021 TENTANG INVESTASI MIRAS PADA MEDIA SOSIAL *TWITTER***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pengkalan data (*database*), merawat, dan mempublikasi Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 27 Desember 2022

Yang Menyatakan,



(Gilbert Panangian Doloksaribu)

Pada tanggal : 27 Desember 2022

Yang Menyatakan,

(Gilbert Panangian Doloksaribu)

## LEMBAR PENGESAHAN

Skripsi diajukan oleh :

Nama : Gilbert Panangian Doloksaribu  
NIM : 1710511081  
Program Studi : S-1 Informatika  
Judul Skripsi : Perbandingan Klasifikasi *Naive Bayes* Dengan Atau Tanpa *Particle Swarm Optimization* Untuk Seleksi Fitur Pada Analisis Sentiment Terhadap Perpres No 10 Tahun 2021 Tentang Investasi Miras Pada Media Sosial *Twitter*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Dr. Didit Widiyanto, S.Kom, M.Si.



Ria Astriratma, S.Komp., M.Cs.



Iin Ernawati S.Kom., M.Si.



Dr. Ermatita, M.Kom



Dr. Widya Cholil, S.Kom., M.I.T

Ditetapkan di : Jakarta  
Tanggal Ujian : 20 Januari 2023



## ABSTRAK

Media sosial merupakan salah satu bentuk perkembangan teknologi informasi dan komunikasi. Jejaring sosial adalah ruang di mana Anda dapat berkomunikasi satu sama lain, dan salah satu platform media sosial terbesar adalah Twitter. Presiden Joko Widodo mengeluarkan perpres No 10 tahun 2021 tentang investasi industri minuman keras hal tersebut mengundang pendapat yang pro dan kontra. Penelitian ini dilakukan untuk mengklasifikasi komentar masyarakat terhadap Perpres No 10 tahun 2021 dengan menggunakan metode Klasifikasi *Naive Bayes*, Data tweet yang didapat dengan melakukan crawling pada twitter dengan kata kunci yang terkait dengan topik dilakukan proses perubahan bentuk data yang belum terstruktur menjadi data yang terstruktur lalu dilakukan perbandingan nilai dari evaluasi antar klasifikasi menggunakan *Naive Bayes* dengan klasifikasi *naive bayes* menggunakan *Particel swarm optimization* sebagai seleksi fitur untuk mengurangi atribut yang kurang relevan dalam proses klasifikasi sehingga dapat meningkatkan nilai akurasi dari dataset. Penelitian ini menghasilkan nilai akurasi klasifikasi *naive bayes* sebesar 65% dan klasifikasi *naive bayes* dengan menggunakan *Particel swarm optimization* sebesar 69% dengan menggunakan parameter iterasi sebanyak 100 kali. Pengukuran evaluasi menggunakan metode *Confusion Matrix* dan menunjukkan peningkatan akurasi dan nilai F1-Score dengan menggunakan PSO sebagai fitur seleksi atribut.

**Kata Kunci:** *Twitter, Analisis Sentimen, Pepres No 10 Tahun 2021, Naive Bayes Classifier, dan Particel Swarm Optimization*

## **KATA PENGANTAR**

Segala puji dan syukur penulis panjatkan kehadiran Tuhan Yang Mahakuasa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Proposal Tugas Akhir. Proposal tugas akhir ini disusun sebagai syarat Tugas Akhir Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Dalam penyelesaian karya tulis ilmiah ini tidak lepas dari bantuan banyak pihak yang telah memberikan masukan kepada penulis. Untuk itu penulis mengucapkan terima kasih kepada:

1. Dr. Ermatita, M.Kom., selaku dekan Fakultas Ilmu Komputer
2. Dr.Widya Cholil, S.Kom., M.I.T., selaku Ketua Program Studi Sarjana Jurusan S1 Informatika.
3. Iin Ernawati S.Kom,M.Si.,selaku dosen pembimbing dari pihak jurusan.
4. Orang tua yang telah memberikan dukungan baik secara moril maupun materil.
5. Seluruh pihak yang terlibat dalam kelancaran pembuatan makalah karya ilmiah ini dan yang belum disebutkan di atas, penulis ucapkan terima kasih.

Penulis menyadari bahwa masih banyak kekurangan dari laporan ini, baik dari materi maupun teknik penyajiannya. Oleh karena itu, kritik dan saran yang membangun penulis harapkan.

Jakarta, 28 Oktober 2022



## DAFTAR ISI

Lembar Persetujuan	iii
Pernyataan Orisinalitas	iv
Pernyataan Persetujuan	v
Lembar Persetujuan	vi
Abstrak	vii
Kata Pengantar	viii
Daftar Isi	ix
Daftar Gambar	xii
Daftar Table	xiii
Daftar Simbol	xv
Daftar Lampiran	xvi
BAB I Pendahuluan	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian .....	3
1.4 Ruang Lingkup.....	3
1.5 Luaran Penelitian .....	3
1.6 Sistematika Penulisan .....	3
BAB II Tinjauan Pustaka	5
2.1 Twitter.....	5
2.2 Text Mining .....	6
2.3 Analisis Sentimen .....	7
2.4 Pra Proses Data .....	8
2.4.1 Punctual Removal	8
2.4.2 Penghapusan URL	9
2.4.3 Transform Case	9
2.4.4 Normalisasi	9
2.4.5 Stemming	9
2.4.6 Tokenisasi	9
2.4.7 Stopword Removal	10
2.5 Term Frequency - Inverse Document Frequency .....	10
2.6 Naïve Bayes .....	11
2.7 Particle Swarm Optimization.....	12

2.8	<i>Confusion Matrix</i> .....	16
2.9	Review Penelitian Terdahulu .....	17
BAB III Analisis dan Perancangan		19
3.1	Kerangka Pikir .....	19
3.1.1	Identifikasi Masalah	19
3.1.2	Studi Literatur	19
3.1.3	Pengumpulan Data	20
3.1.4	Pra Proses Data	21
3.1.4.1	Punctual Removal	21
3.1.4.2	Penghapusan URL	21
3.1.4.3	<i>Transform Case</i>	21
3.1.4.4	Normalisasi	21
3.1.4.5	<i>Stemming</i>	21
3.1.4.6	Tokenisasi	21
3.1.4.7	<i>Stopword Removal.</i>	22
3.1.5	Pembobotan Term <i>TF-IDF</i>	22
3.1.6	Seleksi Fitur	23
3.1.7	Klasifikasi Naïve Bayes	24
3.1.8	Evaluasi	24
3.2	Perangkat Penelitian.....	25
3.2.1	Perangkat Keras	25
3.2.2	Perangkat Lunak	25
3.2.3	Jadwal Kegiatan	25
BAB IV Hasil dan Pembahasan		26
4.1	Pengumpulan Data .....	26
4.2	PraProses Data .....	29
4.2.1	Punctual Removal	30
4.2.2	Penghapusan URL	32
4.2.3	Casefolding	33
4.2.4	Normalisasi	34
4.2.5	Stemming	35
4.2.6	Tokenisasi	36
4.2.7	Stopword Removal	37
4.3	Pembobotan Term TF-IDF .....	39

4.4	Seleksi Fitur .....	44
4.5	Klasifikasi Naïve Bayes .....	62
4.6	Evaluasi.....	68
BAB V Penutup		72
5.1	Kesimpulan .....	72
5.2	Saran .....	72
Daftar Pustaka		73

## DAFTAR GAMBAR






Gambar 2. 1 Logo Twitter .....	5
Gambar 2. 2 Proses Text Mining .....	6
Gambar 2. 3 Pra Proses Data .....	8
Gambar 2. 4 Alur Algoritma Particle Swarm Optimization .....	13
Gambar 3. 1 Kerangka Pikir .....	19
Gambar 4. 1 Raw Data <i>Crawling</i> .....	26
Gambar 4. 2 Data dalam Frame .....	26
Gambar 4. 3 Best Cost dari Iterasi – 50 .....	45
Gambar 4. 4 Best Cost dari Iterasi – 100 .....	46
Gambar 4. 5 Best Cost dari Iterasi – 150 .....	46
Gambar 4. 6 Best Cost dari Iterasi – 200 .....	47
Gambar 4. 7 Best Cost dari Iterasi – 250 .....	47

## DAFTAR TABLE

Tabel 2. 1 Confusion Matrix.....	16
Tabel 3. 1 Jadwal Kegiatan.....	25
Tabel 4. 1 Pelabelan Data .....	27
Tabel 4. 2 Hasil Pelabelan Data.....	28
Tabel 4. 3 Standar Kekuatan Kesepakatan .....	29
Tabel 4. 4 Data sebelum Pra Proses.....	30
Tabel 4. 5 Punctual Removal.....	31
Tabel 4. 6 Penghapusan URL .....	32
Tabel 4. 7 Casefolding .....	33
Tabel 4. 8 Normalisasi .....	34
Tabel 4. 9 Stemming.....	35
Tabel 4. 10 Tokenisasi .....	36
Tabel 4. 11 Stopword Removal .....	37
Tabel 4. 12 Data Setelah Praproses.....	38
Tabel 4. 13 Data Sample.....	39
Tabel 4. 14 Hasil TF-IDF.....	40
Tabel 4. 15 Parameter PSO.....	44
Tabel 4. 16 Iterasi dan Total Fitur Terpilih.....	47
Tabel 4. 17 Fitur.....	48
Tabel 4. 18 Partikel.....	49
Tabel 4. 19 Pemetaan Fitur Pada Partikel.....	50
Tabel 4. 20 Inisialisasi Posisi.....	50
Tabel 4. 21 Inisialisasi Kecepatan.....	51
Tabel 4. 22 Sigmoid iterasi-1.....	52
Tabel 4. 23 Hasil Perbandingan $X_1$ dengan Sigmoid Iterasi 1.....	53
Tabel 4. 24 Fitness Iterasi -1.....	54
Tabel 4. 25 Update Kecepatan Iterasi -2.....	54
Tabel 4. 26 Update Posisi Iterasi Ke-2 .....	55
Tabel 4. 27 Hasil Perhitungan Sigmoid Kecepatan Iterasi Ke-2.....	56
Tabel 4. 28 Perbandingan nilai Sigmoid dengan Posisi pada iterasi ke-2.....	57
Tabel 4. 29 Fitness Partikel.....	58
Tabel 4. 30 Update Kecepatan Iterasi ke-3.....	58

Tabel 4. 31 Update Posisi Partikel iterasi ke-3 .....	59
Tabel 4. 32 Sigmoid Kecepatan Iterasi ke -3 .....	60
Tabel 4. 33 Hasil Perbandingan Sigmoid dengan Posisi Iterasi ke -3 .....	61
Tabel 4. 34 Nilai Fitness Iterasi ke-3 .....	61
Tabel 4. 35 Perubahan Nilai Pbest partikel dalam setiap iterasi .....	62
Tabel 4. 36 Pembagian Data Klasifikasi Naïve bayes .....	63
Tabel 4. 37 Probabilitas Sampel Data Latih .....	64
Tabel 4. 38 Data Uji.....	65
Tabel 4. 39 Probabilitas Data Uji.....	66
Tabel 4. 40 Hasil Klasifikasi Naïve Bayes Tanpa PSO .....	67
Tabel 4. 41 Pembagian Data Klasifikasi Naïve Bayes dengan PSO.....	68
Tabel 4. 42 Hasil Klasifikasi Naïve Bayes dengan PSO berdasarkan Iterasi .....	68
Tabel 4. 43 Confusion matrix klasifikasi Naïve Bayes.....	69
Tabel 4. 44 Hasil Evaluasi Klasifikasi Naïve bayes .....	70
Tabel 4. 45 Confusion Matrix Klasifikasi Naïve Bayes dengan PSO .....	70
Tabel 4. 46 Evaluasi PSO + Naïve Bayes.....	71
Tabel 4. 47 Perbandingan Hasil Evaluasi .....	71

## DAFTAR SIMBOL

<b>Simbol Flowchart</b>			
	<b>GAMBAR</b>	<b>NAMA</b>	<b>KETERANGAN</b>
1		<i>Terminal (start, end)</i>	Menggambarkan bagaimana kegiatan dimulai atau kegiatan berakhir.
2		<i>Flow Direction</i>	Menggambarkan hubungan antar simbol yang menyatakan suatu jalannya proses dalam sistem.
3		<i>Process</i>	Menggambarkan deskripsi dari proses yang dijalankan.
4		<i>Document</i>	Menggambarkan bahwa masukkan ( <i>input</i> ) berasal dari sebuah data dokumen yang dapat berupa kertas atau keluaran ( <i>output</i> ) yang dicetak ke kertas.
5		<i>Predefine process</i>	Menggambarkan pelaksanaan dari sebuah proses atau disebut sebagai subprogram.

## DAFTAR LAMPIRAN

Lampiran 1 Fitur Terpilih Iterasi – 50 .....	76
Lampiran 2 Fitur terpilih iterasi -100.....	81
Lampiran 3 Fitur terpilih iterasi -150.....	86
Lampiran 4 Fitur terpilih iterasi -200.....	91
Lampiran 5 Fitur terpilih iterasi -250.....	96
Lampiran 6 Similarity Index.....	101