



**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER
MENGENAI KEBIJAKAN SELEKSI SEKOLAH JALUR
ZONASI MENGGUNAKAN METODE KLASIFIKASI *NAÏVE*
BAYES DAN SELEKSI FITUR *PARTICLE SWARM*
*OPTIMIZATION***

SKRIPSI

**Yudhistira
1810511120**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2022**



**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER
MENGENAI KEBIJAKAN SELEKSI SEKOLAH JALUR
ZONASI MENGGUNAKAN METODE KLASIFIKASI *NAÏVE*
BAYES DAN SELEKSI FITUR *PARTICLE SWARM*
*OPTIMIZATION***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Komputer**

Yudhistira

1810511120

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa proposal berikut:

Nama : Yudhistira
NIM : 1810511120
Program Studi : S1 Informatika
Judul : Analisis Sentimen Pada Media Sosial Twitter Mengenai Kebijakan Seleksi Sekolah Jalur Zonasi Menggunakan Metode Klasifikasi *Naïve Bayes* Dan Seleksi Fitur *Particle Swarm Optimization*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Yuni Widiastwi, S.Kom, M.Si.
Penguji 1



Nurul Chamidah, S.Kom, M.Kom.
Penguji 2



Dr. Didit Widiyanto, S.Kom, M.Si.
Pembimbing 1



Mayanda Mega Santoni, S.Kom, M.Kom.
Pembimbing 2



Dr. Ermatita, M.Kom.
Dekan



Desta Sandya Pravista, S.Kom, M.Kom.
Ketua Program Studi

Ditetapkan di : Jakarta
Tanggal pengesahan : 13 Juni 2022



PERNYATAAN ORSINALITAS

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

Nama : Yudhistira

NIM : 1810511120

Tanggal : 13 Juni 2022

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

Jakarta, 13 Juni 2022

Yang Menyatakan,



(Yudhistira)

PERNYATAAN PERSETUJUAN PUBLIKASI

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

Nama : Yudhistira
NIM : 1810511120
Fakultas : Ilmu Komputer
Program Studi : S1-Informatika

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

Analisis Sentimen Pada Media Sosial Twitter Mengenai Kebijakan Seleksi Sekolah Jalur Zonasi Menggunakan Metode Klasifikasi *Naïve Bayes* Dan Seleksi Fitur *Particle Swarm Optimization*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti di Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formalitas, mengelola dalam bentuk pengkalan data (Basis Data), 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 : 13 Juni 2022

Yang Menyatakan,



(Yudhistira)

**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER MENGENAI
KEBIJAKAN SELEKSI SEKOLAH JALUR ZONASI MENGGUNAKAN
METODE KLASIFIKASI *NAÏVE BAYES* DAN SELEKSI FITUR
*PARTICLE SWARM OPTIMIZATION***

Yudhistira

ABSTRAK

Media sosial menjadi wadah untuk menampung opini atau sentimen masyarakat. Contoh aplikasi yang sering digunakan untuk membahas sentimen tersebut adalah Twitter. Pengguna Twitter sering menyampaikan pendapatnya tentang beberapa topik termasuk pendidikan khususnya kebijakan pemerintah mengenai seleksi sekolah jalur zonasi yang dapat dilihat di Twitter. Berdasarkan latar belakang kondisi tersebut, maka dibutuhkan penelitian tentang sentimen masyarakat mengenai kebijakan seleksi sekolah jalur zonasi. Penelitian ini dilakukan dengan menggunakan metode *Machine Learning* yaitu klasifikasi *Naïve Bayes* dan seleksi fitur menggunakan *Particle Swarm Optimization* untuk mengklasifikasikan *tweet* positif atau *tweet* negatif yang masyarakat lontarkan khususnya pelajar dan orang tuanya yang memperebutkan kursi di sekolah-sekolah negeri. Hasil dari penelitian ini membuktikan bahwa klasifikasi *Naïve Bayes* dengan seleksi fitur *Particle Swarm Optimization* mendapatkan nilai model evaluasi pada akurasi lebih besar dibandingkan tanpa menggunakan *Particle Swarm Optimization* dimana akurasi sebesar 78%, presisi sebesar 23%, *recall* sebesar 45%, dan *specificity* sebesar 82%. Sementara itu, nilai model evaluasi klasifikasi *Naïve Bayes* tanpa menggunakan seleksi fitur *Particle Swarm Optimization* lebih kecil, dimana akurasi sebesar 75%, presisi sebesar 37%, *recall* sebesar 28%, dan *specificity* sebesar 87%. Terjadi kenaikan performa pada akurasi dan *recall*, serta terjadi penurunan performa pada presisi dan *specificity*. Nilai evaluasi pada presisi dan *recall* mendapatkan nilai yang rendah, karena data yang tidak seimbang, dimana perbandingan antara label positif dan negatif adalah 1:4.

Kata Kunci: *Twitter, Analisis Sentimen, Sekolah Negeri, Zonasi, Naïve Bayes, Particle Swarm Optimization.*

**ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER MENGENAI
KEBIJAKAN SELEKSI SEKOLAH JALUR ZONASI MENGGUNAKAN
METODE KLASIFIKASI NAÏVE BAYES DAN SELEKSI FITUR
*PARTICLE SWARM OPTIMIZATION***

Yudhistira

ABSTRACT

Social media is a place to accommodate people's opinions or sentiments. An example of an application that is often used to discuss such sentiments is Twitter. Twitter users often express their opinions on several topics including education in particular government policies regarding zoning pathway school selection which can be viewed on Twitter. Based on the background of these conditions, research is needed on community sentiment regarding the zoning line school selection policy. This research was conducted using the *Machine Learning* method, namely *the Naïve Bayes* classification and feature selection using *Particle Swarm Optimization* to classify positive *tweets* or negative *tweets* that the community made, especially students and their parents who were fighting for seats in public schools. The results of this study prove that *the Naïve Bayes* classification with *particle swarm Optimization* feature selection gets an evaluation model value at greater accuracy than without using *Particle Swarm Optimization* where accuracy is 78%, precision is 23%, *recall* is 45%, and *specificity* by 82%. Meanwhile, the value of *the Naïve Bayes* classification evaluation mode without using *the Particle Swarm Optimization* feature selection is smaller, where the accuracy is 75%, the precision is 37%, *the recall* is 28%, and *the specificity* is 87%. There is an increase in performance in accuracy and *recall*, as well as a decrease in performance in precision and *specificity*. The evaluation value on precision and *recall* gets a low value, due to unbalanced data, where the ratio between positive and negative labels is 1: 4.

Keywords: Twitter, Sentimen Analysis, Public School, Zoning, Naïve Bayes, Particle Swarm Optimization.

KATA PENGANTAR

Puji serta syukur kepada Tuhan Yang Maha Esa atas nikmat, rahmat dan karunianya sehingga penulis bisa menyelesaikan laporan skripsi. Skripsi ini dikerjakan untuk syarat dari penyelesaian Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Dalam menyelesaikan laporan skripsi tidak luput dari bantuan berbagai pihak yang sudah memberikan saran dan opini kepada penulis. Maka dari itu penulis mengucapkan terima kasih kepada:

1. Orang tua yang sudah memberikan dukungan baik dan positif secara moril maupun materil.
2. Dr. Ermatita, M.Kom., sebagai dekan Fakultas Ilmu Komputer
3. Desta Sandya Pravista, S.Kom, M.Kom., sebagai Ketua Program Studi Sarjana Jurusan S1 Informatika.
4. Dr. Didit Widiyanto, S.Kom, M.Si., sebagai dosen pembimbing 1 dari pihak jurusan, dan Mayanda Mega Santoni, S.Kom, M.Kom., sebagai dosen pembimbing 2 dari pihak jurusan.
5. Semua tim pengajar dan staf pada prodi informatika, karena telah mendidik dan membimbing penulis selama menjalani proses perkuliahan.
6. Semua jajaran Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta yang sudah membantu dalam perizinan maupun administrasi.
7. Sahabat-sahabat dalam UKM UFO Veteran Jakarta yang telah menuntun peneliti menuju kepada kepemilikan karakter positif dan baik dari dalam diri peneliti, serta telah memberikan semangat dan dorongan untuk dapat menyelesaikan kuliah dan skripsi ini.
8. Teman-teman Informatika angkatan 2018 yang telah membantu dan memberi masukan serta telah menemani peneliti dari masa perkuliahan sampai dengan penyusunan skripsi.
9. Semua orang yang membantu dalam kelancaran pembuatan laporan skripsi dan yang tidak disebutkan satu persatu, penulis mengucapkan terima kasih.

Penulis sadar bahwa masih jauh dari kata sempurna untuk laporan skripsi ini, baik dari teknik penyajian maupun materinya. Oleh karena itu, penulis berharap

diberikan kritik dan saran yang membangun. Akhir kata penulis meminta supaya laporan ini dapat bermanfaat untuk semua pihak.

Bekasi, 13 Juni 2022

Yang Menyatakan,

A handwritten signature in black ink, consisting of several vertical and diagonal strokes, appearing to be a stylized representation of the author's name.

Penulis

DAFTAR ISI

COVER	I
JUDUL	II
LEMBAR PENGESAHAN	III
PERNYATAAN ORSINALITAS	IV
PERNYATAAN PERSETUJUAN PUBLIKASI	V
ABSTRAK	VI
ABSTRACT	VII
KATA PENGANTAR	VIII
DAFTAR ISI	X
DAFTAR GAMBAR	XIII
DAFTAR TABEL	XIV
DAFTAR LAMPIRAN	XV
BAB I	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	3
1.5. Ruang Lingkup	4
1.6. Luaran Yang Diharapkan	4
1.7. Sistematika Penulisan	4
BAB II	6
2.1. Analisis Sentimen	6
2.2. Twitter	6
2.3. Twitter API	7
2.4. Text Mining	8

2.5.	Machine Learning	8
2.6.	Naïve Bayes	9
2.7.	Particle Swarm Optimization	11
2.8.	Confusion Matrix	12
2.9.	Review Penelitian Terdahulu	13
BAB III		17
3.1.	Kerangka Pikir	17
3.1.1.	Identifikasi Masalah	18
3.1.2.	Studi Literatur	18
3.1.3.	Akuisisi Data.....	18
3.1.4.	Pelabelan Data.....	18
3.1.5.	Kappa Value.....	19
3.1.6.	Pra Proses Data	20
3.1.6.1.	Case Folding	20
3.1.6.2.	<i>Tokenizing</i>	20
3.1.6.3.	<i>Normalizing</i>	20
3.1.6.4.	<i>Filtering</i>	21
3.1.6.5.	<i>Cleansing</i>	21
3.1.6.6.	<i>Stemming</i>	21
3.1.7.	Pembobotan TF IDF.....	21
3.1.8.	Seleksi Fitur Particle Swarm Optimization.....	22
3.1.9.	Klasifikasi <i>Naïve Bayes</i>	22
3.1.10.	Pengujian dan Evaluasi	23
3.1.11.	Visualisasi	23
3.1.12.	Hasil Klasifikasi	23
3.2.	Perangkat Penelitian.....	23
BAB IV		25
4.1.	Data	25
4.2.	Pelabelan Data.....	25
4.3.	Praproses Data.....	30
4.3.1.	<i>Case Folding</i>	30
4.3.2.	<i>Cleansing</i>	30

4.3.3. <i>Normalizing</i>	31
4.3.4. <i>Tokenizing</i>	31
4.3.5. <i>Filtering</i>	31
4.3.6. <i>Stemming</i>	32
4.4. Pembobotan Kata	33
4.5. Klasifikasi <i>Naïve Bayes</i>	35
4.5.1. Tahap Data <i>Training</i>	35
4.5.2. Tahap Data <i>Testing</i>	38
4.6. Seleksi Fitur Particle Swarm Optimization.....	40
4.7. Klasifikasi <i>Naïve Bayes</i> Berbasis Particle Swarm Optimization.....	41
4.8. Evaluasi	42
4.8.1. Evaluasi <i>Naïve Bayes</i>	42
4.8.2. Evaluasi <i>Naïve Bayes</i> Berbasis Particle Swarm Optimization	44
4.9. Visualisasi	45
4.9.1. Sentimen Positif	45
4.9.2. Sentimen Negatif.....	46
BAB V KESIMPULAN DAN SARAN.....	48
5.1. Kesimpulan	48
5.2. Saran.....	49
DAFTAR PUSTAKA	50
DAFTAR RIWAYAT HIDUP.....	53
LAMPIRAN.....	54

DAFTAR GAMBAR

Gambar 3. 1 Kerangka Pikir.....	17
Gambar 3. 2 Interpretasi Kappa	20
Gambar 4. 1 Label.....	26
Gambar 4. 2 Wordcloud Sentimen Positif	46
Gambar 4. 3 Wordcloud Sentimen Negatif.....	47

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	12
Tabel 2. 2 <i>Resume Literature</i>	14
Tabel 4. 1 Hasil pelabelan data tweet.....	26
Tabel 4. 2 Sampel tweet	26
Tabel 4. 3 Sampel <i>Case Folding</i>	30
Tabel 4. 4 Sampel <i>Cleansing</i>	30
Tabel 4. 5 Sampel <i>Normalizing</i>	31
Tabel 4. 6 Sampel <i>Tokenizing</i>	31
Tabel 4. 7 <i>Filtering</i>	32
Tabel 4. 8 <i>Stemming</i>	32
Tabel 4. 9 Sampel tweet	33
Tabel 4. 10 Perhitungan TF-IDF	34
Tabel 4. 11 Pembagian data	35
Tabel 4. 12 Data <i>Training</i>	35
Tabel 4. 13 Nilai TF-IDF	36
Tabel 4. 14 Hasil peluang <i>Training</i>	37
Tabel 4. 15 Sampel Data <i>Testing</i>	38
Tabel 4. 16 Hasil Pra Proses Sampel Data <i>Testing</i>	38
Tabel 4. 17 Hasil TF-IDF Sampel Data <i>Testing</i>	39
Tabel 4. 18 Hasil Peluang kata Sampel Data <i>Testing</i>	39
Tabel 4. 19 Seleksi Fitur	41
Tabel 4. 20 Akurasi PSO.....	41
Tabel 4. 21 Confusion Matrix NB.....	43
Tabel 4. 22 Confusion Matrix NB PSO	44
Tabel 4. 23 Evaluasi <i>Confusion Matrix</i> NB Tanpa PSO.....	44
Tabel 4. 24 Evaluasi <i>Confusion Matrix</i> NB PSO.....	45

DAFTAR LAMPIRAN

Lampiran 1 Kamus Normalisasi	54
Lampiran 2 <i>Stopword</i> Bahasa Indonesia.....	62
Lampiran 3 Turnitin	67