

Riana

by Jiko jurnal

Submission date: 26-Dec-2022 05:55PM (UTC-0500)

Submission ID: 1983651589

File name: 5260-13847-1-SM.doc (73.16K)

Word count: 3450

Character count: 22422

IMPLEMENTASI INFORMATION GAIN DAN PSO PADA ANALISIS SENTIMEN PENANGANAN COVID-19 MENGGUNAKAN K-NEAREST NEIGHBOR

Riana¹, Muhammad I Mazdadi², Irwan Budiman³, Muliadi⁴, Rudy Herteno⁵

^{1,2,3,4,5}Ilmu Komputer, Fakultas MIPA, Universitas Lambung Mangkurat
Email: ¹1811016220010@mhs.ulm.ac.id, ²mazdadi@ulm.ac.id, ³irwan.budiman@ulm.ac.id,
⁴muliadi@ulm.ac.id, ⁵rudy.herteno@ulm.ac.id

19

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

Awal tahun 2020, virus baru dari Wuhan, China yang dikenal dengan virus corona atau COVID-19, mengguncang dunia (Coronavirus Disease 2019). Berbagai upaya telah dilakukan pemerintah untuk memerangi wabah ini, meskipun peran pemerintah dalam memerangi Covid-19 memiliki banyak kekurangan dan kelebihan. Perlakuan pemerintah Indonesia terhadap virus Covid-19 menjadi salah satu topik yang paling sering dibicarakan di Twitter. Komentar Twitter dapat diperiksa untuk sentimen dan kemudian dikategorikan menggunakan algoritma klasifikasi untuk menentukan mana yang positif dan mana yang negatif. Metode klasifikasi K-Nearest Neighbor, pemilihan fitur Information Gain dengan klasifikasi K-Nearest Neighbor, dan pemilihan fitur Information Gain dan optimasi Particle Swarm Optimization dengan algoritma klasifikasi K-Nearest Neighbor dibandingkan dalam makalah ini. Klasifikasi K-Nearest Neighbor dengan pemilihan fitur Information Gain dan optimasi Particle Swarm Optimization lebih unggul dari model K-Nearest Neighbor tanpa pemilihan fitur dan tanpa optimasi serta lebih unggul dari model K-Nearest Neighbor dengan pemilihan fitur Information Gain yaitu 87,33% dengan nilai K 5.

Kata kunci: Analisis Sentimen, Penanganan Covid-19, K-Nearest Neighbor, Information Gain, Particle Swarm Optimization

IMPLEMENTATION OF INFORMATION GAIN AND PSO IN SENTIMEN ANALYSIS OF COVID-19 HANDLING USING K-NEAREST NEIGHBOR

Abstract

Beginning in 2020, a new virus from Wuhan, China, known as the corona virus or COVID-19, shook the world (Coronavirus Disease 2019). Various measures have been made by the government to combat this epidemic, although the government's response in combating Covid-19 has numerous shortcomings and strengths. The Indonesian government's treatment of the Covid-19 virus is one of the most frequently discussed subjects on Twitter. Twitter comments can be examined for sentiment and then categorised using a classification algorithm to determine which are positive and which are negative. The K-Nearest Neighbor classification method, Information Gain feature selection with K-Nearest Neighbor classification, and Information Gain feature selection and Particle Swarm Optimization optimization with the K-Nearest Neighbor classification algorithm are compared in this paper. The K-Nearest Neighbor classification with Information Gain feature selection and Particle Swarm Optimization optimization is superior to the K-Nearest Neighbor model without feature selection and without optimization and is superior to the K-Nearest Neighbor model with Information Gain feature selection, which is 87.33% with a K value of 5.

Keywords: Sentiment Analysis, Handling Covid-19, K-Nearest Neighbor, Information Gain, Particle Swarm Optimization

1. PENDAHULUAN

Peran pemerintah dalam penanganan Covid-19 menuai banyak pro dan kontra. Hal ini membuat masyarakat umum ingin mengutarakan berbagai opini. Keterbatasan waktu dan ruang membuat masyarakat terkendala menyampaikan opini, adanya jejaring sosial merupakan salah satu tempat yang

cocok untuk mengutarakan berbagai opini, seperti Twitter [1]. Pengguna Twitter kerap memperdebatkan persepsi publik tentang bagaimana pemerintah Indonesia menangani kasus Covid-19. Salah satu cara untuk mengukur opini publik tentang bagaimana pemerintah menangani COVID-19 di Indonesia adalah melalui Twitter [2].

Analisis sentimen merupakan proses memahami dan mengelompokkan apa yang ditemukan dalam sebuah tulisan melalui penggunaan teknik analisis teks. Analisis sentimen bertujuan untuk menilai emosi, persepsi, dan penilaian konsumen atau peneliti terhadap produk, layanan, atau kepribadian seseorang. Keuntungan dari analisis sentimen ini dapat dilakukan secara otomatis, sehingga menghemat waktu dan sumber daya [3].

Penelitian terdahulu yang dilakukan [4] menggunakan metode K-NN pada analisis sentimen masyarakat terhadap larangan Covid 2021 menghasilkan akurasi sebesar 86.67%. Ada atau tidak adanya fitur yang tidak relevan, atau jika bobot fitur tersebut tidak sama dengan signifikansinya terhadap klasifikasi, berdampak signifikan terhadap akurasi algoritma K-NN [5].

Pada penelitian yang berbeda [6] tentang analisis sentimen penggunaan *E-Wallet* saat pandemi menggunakan *Naive Bayes Classifier* dan *Information Gain* menghasilkan akurasi 92%. *Information Gain* adalah salah satu metode terbaik dalam pemilihan fitur. Salah satu teknik yang banyak digunakan untuk seleksi fitur adalah *Information Gain*. Dalam aplikasi seperti klasifikasi teks, analisis data *microarray*, dan analisis data gambar, *Information Gain* adalah pendekatan pemilihan fitur yang paling sederhana berdasarkan karakteristik peringkat. *Information Gain* dapat membantu dalam menurunkan *noise* yang disebabkan oleh fitur yang tidak penting [7].

Selain itu, penelitian yang dilakukan oleh [8] tentang analisis sentimen transportasi online menggunakan algoritma *Support Vector Machine* dengan pembobotan TF-IDF, menghasilkan akurasi 95,46%. Selanjutnya ditambahkan algoritma *Particle Swarm Optimization* menghasilkan akurasi sebesar 96,04%. Perbandingan teknik klasifikasi analisis sentimen pada temuan masalah ini menunjukkan bahwa pendekatan SVM yang dioptimalkan PSO lebih unggul daripada algoritma SVM saja.

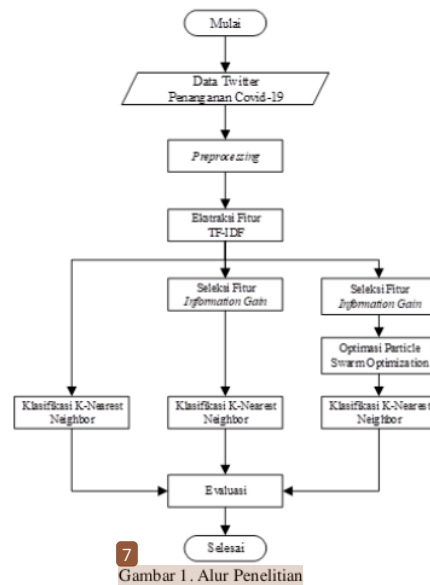
Algoritma *K-Nearest Neighbor* memiliki kekurangan yaitu kurang optimal dalam mengidentifikasi nilai k dan membutuhkan seleksi fitur untuk mencapai hasil terbaik [9]. Kelemahan lainnya termasuk biaya komputasi algoritma yang tinggi dan mudah tertipu dengan fitur yang tidak relevan. Masalah ini dapat diselesaikan dengan pemilihan fitur *Information Gain* dengan menghilangkan fitur-fitur yang tidak relevan dan memberi peringkat pada fitur-fitur penting untuk menghasilkan nilai evaluasi sistem terbaik [10]. Selain itu, metode optimasi menggunakan algoritma *Particle Swarm Optimization* diterapkan untuk meningkatkan klasifikasi K-NN. Dalam proses klasifikasi data mining, PSO digunakan sebagai pengambil keputusan untuk menentukan solusi yang terbaik.

Pada penelitian ini, penulis menggunakan metode *K-Nearest Neighbor* beserta seleksi fitur

Information Gain dan optimasi *Particle Swarm Optimization* untuk menganalisis sentimen pengguna Twitter terhadap penanganan Covid-19.

2. METODE PENELITIAN

Gambar 1 menunjukkan tahapan penelitian. Penelitian ini terdiri dari tahapan pengumpulan data, *preprocessing* data, ekstraksi fitur menggunakan TF-IDF, seleksi fitur menggunakan *Information Gain*, pembagian data, klasifikasi menggunakan *K-Nearest Neighbor*, optimasi menggunakan *Particle Swarm Optimization*, evaluasi menggunakan *confusion matrix*. Pembagian data terdiri dari *Data Train* dan *Data Test*. Pembagian data dibagi menjadi 90:10.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Penelitian ini menggunakan data Twitter tentang opini penanganan Covid-19 oleh pemerintah di Indonesia. Data yang dikumpulkan diperoleh dari Kaggle, setelah itu dilakukan seleksi dan total 1500 tweet dikumpulkan.

2.2. Preprocessing

Preprocessing pada penelitian ini terbagi menjadi beberapa tahapan yaitu:

Pelabelan: Pada proses pelabelan data dibagi menjadi 2 label yaitu: (i) positif dan (ii) negatif.

Cleansing: Proses membersihkan dataset dari link url, *hashtag*, *username*, *symbol*, email, tanda baca, dan angka.

Case Folding: Setiap bentuk kata diubah menjadi huruf kecil atau *lowercase*.

Formalisasi: Kata yang tidak baku diubah menjadi kata baku sesuai kaidah KBBI yang berlaku.

Stemming: Proses ini mengubah kata ke bentuk dasarnya atau menghapus imbuhan pada kata.

Stopword Removal: Proses ini untuk menghapus kata yang tidak memiliki makna penting, seperti “aku”, “agak”, “siapa”.

2.3. TF-IDF

Term Frequency - Inverse Document Frequency (TF-IDF) adalah teknik untuk menghitung pentingnya setiap kata dalam sebuah frasa berdasarkan frekuensi kemunculannya di seluruh kumpulan dokumen. Seberapa sering term tertentu muncul diukur dengan statistik yang disebut *Term Frequency*. Jika nilai TF semakin tinggi, maka semakin tinggi pula bobotnya. IDF menunjukkan seberapa sering term tertentu muncul di setiap dokumen. Semakin besar nilai IDF, semakin rendah TF. Setelah data telah diproses sebelumnya, itu harus dalam format numerik. TF-IDF digunakan untuk mengubah data menjadi numerik. Nilai TF awal setiap kata dalam perhitungan bobot TF-IDF adalah 1 [8]. Rumus berikut dapat digunakan untuk mendapatkan TF-IDF:

$$TFIDF_a = TF_a \times \log\left(\frac{N}{DF_a}\right) \quad (1)$$

Keterangan:

TF_a : banyaknya kemunculan setiap kata pada suatu dokumen a

N : jumlah seluruh dokumen

DF_a : jumlah dokumen keseluruhan yang memuat kata a

[12].

2.4. Information Gain

Information Gain adalah proses pemeringkatan fitur yang sangat sederhana, metode ini banyak digunakan dalam aplikasi berkategori bacaan. **Information Gain** ini mengukur seberapa banyak informasi pada tiap fitur atau term terhadap kategori. **Information Gain** memungkinkan seseorang untuk menentukan nilai setiap fitur yang menunjukkan seberapa efektif fitur itu membentangkan fitur-fitur penting dari kategori sentimen. **Information Gain** dihitung berdasarkan seberapa besar pengaruh sebuah term digunakan untuk klasifikasi [13]. Langkah-langkah dalam perhitungan bobot *information gain* sebagai berikut.

1. Menghitung nilai entropy pada dataset dengan Persamaan 2.

$$Entropy(S) = -\sum_{i=1}^c P(i) \log_2 P(i) \quad (2)$$

Persamaan (2) merupakan jumlah pada atribut target (jumlah kelas klasifikasi), sedangkan $P(i)$ merupakan proporsi sampel pada kelas i dengan sampel data.

2. Menghitung nilai entropy dengan A merupakan atribut, dan v adalah suatu nilai yang mungkin untuk atribut A, serta entropy (S_v) yang merupakan entropy untuk setiap partisi j dengan Persamaan 3.

$$Entropy(S, A) = \sum_{j=1}^v \frac{|D_j|}{|D|} Entropy(S_v) \quad (3)$$

Persamaan (3), $|D|$ adalah jumlah sampel data dan $|D_j|$ adalah jumlah sampel data untuk nilai partisi j. v adalah suatu nilai yang mungkin untuk atribut A.

3. Langkah terakhir untuk mendapatkan nilai bobot information gain adalah menggunakan Persamaan 4.

$$Entropy(S, A) = - Entropy(D) - Entropy(S, A) \quad (4)$$

Persamaan (4) nilai *information gain* diperoleh dari pengurangan nilai persamaan 2 dengan nilai persamaan 3.

Selanjutnya dengan menggunakan bobot *information gain* tersebut setiap kata dirangking sehingga menghasilkan fitur yang terbaik. Bobot tersebut digunakan untuk klasifikasi data uji yang nantinya diambil fitur dengan nilai tertinggi [14].

2.5. Particle Swarm Optimization

Particle Swarm Optimization hanya dengan menyesuaikan beberapa parameter, adalah pendekatan optimasi tercepat dan termudah. Pemilihan atribut, pemilihan fitur, dan menaikkan bobot atribut untuk semua atribut atau variabel yang digunakan adalah semua cara untuk mengoptimalkan dalam *Particle Swarm Optimization* [8]. *Particle Swarm Optimization* seperti segerombolan burung terbang ke suatu wilayah untuk mencari kebutuhan makanannya. Mereka tidak tahu bagaimana menuju ke sana, tetapi mereka tahu seberapa jauh jarak makanannya. Oleh karena itu, mengikuti burung yang terakhir terlihat di dekat makanan adalah metode yang paling efisien [15].

Kelebihan *Particle Swarm Optimization* yaitu konvergensi yang cepat dan sederhana serta mudah diterapkan. Ini mengacu pada proses pencarian solusi berdasarkan jumlah partikel dari *Particle Swarm Optimization*. Populasi dinilai secara acak dan memiliki batas minimum dan maksimum. Partikel-partikel beradaptasi dengan letak terbaiknya (*local best*) dan beradaptasi terhadap letak partikel terbaik dalam kelompok secara keseluruhan (*global best*) saat bergerak melintasi ruang pencarian [8].

2.6. K-Nearest Neighbor

Untuk mengklasifikasikan data baru, metode **K-Nearest Neighbor (K-NN)** menghitung jarak antara setiap bagian data dan tetangga terdekatnya, dan kemudian menetapkan setiap bagian data ke salah satu kategori tersebut. **K-Nearest Neighbor** meningkatkan metode klasifikasi **Nearest Neighbor** [13]. Pendekatan K-NN memiliki keuntungan bekerja dengan baik dengan kumpulan data yang besar, dan

juga kuat terhadap data latih yang *noise*, yaitu data dengan rentang nilai yang luas tetapi dapat mengganggu struktur data saat ini [9]. Jarak pada algoritma *K-Nearest Neighbor* menggunakan *Cosine Similarity* dihitung menggunakan Persamaan 5:

$$\cos(\theta_{QD}) = \frac{\sum_{i=1}^n Q_i D_i}{\sqrt{\sum_{i=1}^n (Q_i)^2} \sqrt{\sum_{i=1}^n (D_i)^2}} \quad (5)$$

Keterangan:

$\cos(\theta_{QD})$: kemiripan Q terhadap dokumen D

Q : data uji

d : data latih

n : jumlah data latih [16].

2.7. Evaluasi

Menghitung kinerja dari model klasifikasi adalah salah satu cara untuk menentukan efektivitas dari sistem yang sedang dibangun. Salah satu metrik yang digunakan untuk menghitung kinerja dan efektivitas sistem adalah *confusion matrix* [12]. *Confusion Matrix* adalah teknik untuk mengevaluasi kinerja algoritma klasifikasi. Penerapan evaluasi dengan *confusion matrix* akan mendapatkan nilai akurasi, nilai akurasi ini adalah persentase dari data yang sudah diklasifikasikan dengan benar oleh suatu algoritma. Pada *confusion matrix* juga dapat dihitung metrik seperti akurasi untuk menilai efektivitas hasil klasifikasi yang diterapkan. Akurasi merupakan proporsi dari hasil klasifikasi *true positive* maupun *true negative* dari semua dokumen menggunakan persamaan 6

$$accuracy = \frac{TN+TP}{TN+FP+FN+TP} \times 100\% \quad (6)$$

[13].

20

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Penelitian ini memanfaatkan dataset Twitter yang memuat pembahasan tentang bagaimana pemerintah Indonesia menangani virus Covid-19. Data yang dikumpulkan berjumlah 1500 *record* sumber dari kaggle, data dapat diunduh pada link <https://www.kaggle.com/dionisiusdh/covid19indonesian-twitter-sentiment>.

3.2. Preprocessing

Tahap pertama dilakukan pelabelan manual berdasarkan karakteristik data yang digunakan. Pada proses pelabelan data dibagi menjadi 2 label yaitu: (i) Negatif, dan (ii) Positif.

Tabel 1. Pelabelan

Text	Label
Aku bingung pemerintah sebenarnya gimana ya? Serius apa ngga nanganin covid? https://twitter.com/ainunnajib/status/1260924890439315456	Negatif

hayuuk Bersatu cegah Covid-19 w/ mengikuti aturan dri Pemerintah. #ShalatIdDirumahAja Positif pic.twitter.com/6bjLhMgBkZ

Tahap kedua yaitu *Cleansing* yang digunakan untuk mengurangi *noise* dengan cara menghilangkan kata – kata yang tidak diperlukan. Kata yang dihilangkan seperti URL, *hashtag*, nama pengguna (@username), angka, email, dan tanda baca. Proses *Cleansing* dapat ditunjukkan pada Tabel 2.

Tabel 2. Cleansing

Sebelum Cleansing	Setelah Cleansing
Aku bingung pemerintah sebenarnya gimana ya? Serius apa ngga nanganin covid? https://twitter.com/ainunnajib/status/1260924890439315456 #ShalatIdDirumahAja pic.twitter.com/6bjLhMgBkZ	Aku bingung pemerintah sebenarnya gimana ya Serius apa ngga nanganin covid
hayuuk Bersatu cegah Covid-19 w/ mengikuti aturan dri Pemerintah. #ShalatIdDirumahAja pic.twitter.com/6bjLhMgBkZ	hayuuk Bersatu cegah Covid w mengikuti aturan dri Pemerintah

Tahap ketiga yaitu *Case Folding* yang digunakan untuk mengubah bentuk kata menjadi kata *lowercase* atau huruf kecil. Proses *Case Folding* dapat ditunjukkan pada Tabel 3.

Tabel 3. Case Folding

Sebelum Case Folding	Setelah Case Folding
Aku bingung pemerintah sebenarnya gimana ya Serius apa ngga nanganin covid	aku bingung pemerintah sebenarnya gimana ya serius apa ngga nanganin covid
hayuuk Bersatu cegah Covid w mengikuti aturan dri Pemerintah	hayuuk bersatu cegah covid w mengikuti aturan dri pemerintah

Tahap keempat yaitu Formalisasi yang digunakan untuk mengubah bentuk kata dari kata singkatan dan kata yang tidak sesuai ejaan KBBI ke bentuk kata baku yang sesuai ejaan KBBI. Pada tahapan formalisasi ini menggunakan kamus bahasa baku yang dibuat khusus berdasarkan data yang digunakan. Proses formalisasi dapat ditunjukkan pada Tabel 4.

Tabel 4. Formalisasi

Sebelum Formalisasi	Setelah Formalisasi
aku bingung pemerintah sebenarnya gimana ya serius apa ngga nanganin covid	aku bingung pemerintah sebenarnya bagaimana ya serius apa tidak tangan covid
hayuuk bersatu cegah covid w mengikuti aturan dri pemerintah	ayo bersatu cegah covid w mengikuti aturan dari pemerintah

Tahap kelima yaitu *Stemming* yang digunakan untuk menghapus imbuhan yang terdapat pada kata atau mengubah kata menjadi bentuk dasarnya. Proses *Stemming* dapat ditunjukkan pada Tabel 5.

Tabel 5. Stemming

Sebelum Stemming	Setelah Stemming
aku bingung pemerintah sebenarnya bagaimana ya serius apa tidak tangan covid	aku bingung pemerintah benar bagaimana ya serius apa tidak tang covid
ayo bersatu cegah covid w mengikuti aturan dari pemerintah	ayo satu cegah covid w ikut atur dari pemerintah

Tahap ²⁴enam yaitu *Stopword Removal* yang digunakan untuk menghilangkan kata yang tidak ²⁶memiliki arti atau tidak baku (*stopword*) yang terdapat didalam *tweet*. Untuk menghapus kata yang tidak memiliki makna penting atau tidak baku maka digunakan kamus yang dibuat berdasarkan data yang digunakan dan kamus umum Bahasa Indonesia yang diakses melalui link github masdevid <https://github.com/masdevid/ID-Stopword>. Proses *Stopword Removal* dan ⁵ditunjukkan pada Tabel 6.

Tabel 6. *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
aku bingung pemerintah benar bagaimana ya serius apa tidak tang covid	bingung pemerintah serius tang covid
ayo satu cegah covid w ikut atur dari pemerintah	cegah covid atur pemerintah

3.2. Ekstraksi Fitur TF-IDF

¹⁵ Proses ekstraksi fitur menggunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF). TF-IDF atau *Term Frequency - Inverse Document Frequency* merupakan sebuah perhitungan yang biasa digunakan untuk menghitung bobot kata pada sebuah dokumen. Pada Tabel 7 dapat dilihat contoh hasil ekstraksi fitur dari data yang sudah di *preprocessing* pada 1500 data dengan 750 setiap labelnya.

Tabel 7. Contoh Hasil Pembobotan TF-IDF

ambil	bantu	...	temu	zona	label
0	0,354994	...	0,265754	0	Negatif
0,463629	0	...	0	0	Negatif
...
0	0,328698	...	0	0	Positif

3.3. Seleksi Fitur dengan *Information Gain*

Pada proses selanjutnya dilakukan penyeleksian fitur menggunakan *Information Gain* yang ¹²implementasikan menggunakan tools RapidMiner. Proses seleksi fitur dengan algoritma *Information Gain* dilakukan menggunakan *threshold* sebesar 0,01 dan menghasilkan 1897 fitur dari 1916 fitur.

3.4. Particle Swarm Optimization

Untuk mengoptimalkan klasifikasi *K-Nearest Neighbor* dan membantu meningkatkan akurasi ³⁰pada model yang dibuat digunakan teknik optimasi dengan algoritma *Particle Swarm Optimization*. Algoritma *Particle Swarm Optimization* diimplementasikan menggunakan tools RapidMiner. Nilai partikel yang digunakan adalah 10 dengan jumlah iterasi 30, untuk nilai parameter lainnya diatur secara *default*.

3.5. Pembagian Data

Sebelum melakukan klasifikasi data terlebih dahulu dibagi menjadi data *training* dan data *testing*, dimana pembagian tersebut menggunakan teknik *random percentage split*. ¹⁰ dalam penelitian ini dilakukan pembagian dataset 90% untuk data *training*

dan 10% untuk data *testing*. Split data dilakukan secara *random by machine*.

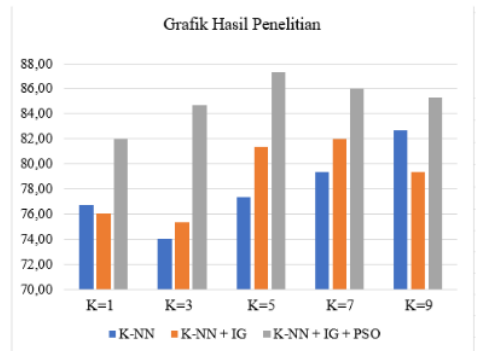
3.6. Perbandingan Performa

Pada klasifikasi *K-Nearest Neighbor*, ¹⁴ parameter K yang dipakai untuk pembuatan model *K-Nearest Neighbor*, model *K-Nearest Neighbor* menggunakan ³⁷ *Information Gain*, dan model *K-Nearest Neighbor* dengan seleksi fitur *Information Gain* dan optimasi *Particle Swarm Optimization* adalah bilangan ganjil dari rentang 1 sampai 10 dan perhitungan jarak menggunakan *Cosine Similarity*.

Tabel 8. Hasil Akurasi

Model	Akurasi				
	K=1	K=3	K=5	K=7	K=9
K-NN	76,76%	74%	77,33%	79,33%	82,67%
<i>Information Gain</i> +K-NN	76%	75,33%	81,33%	82%	79,33%
<i>Information Gain</i> + PSO+ K-NN	82%	84,67%	87,33%	86%	85,33%

Dilihat dari tabel 8 di atas pada model *K-Nearest Neighbor* K=9 dengan akurasi 82,67% lebih tinggi daripada akurasi pada nilai K yang lain. Pada model *K-Nearest Neighbor* dengan seleksi fitur *Information Gain*, K=7 dengan akurasi 82% lebih tinggi daripada akurasi pada nilai K yang lain. Kemudian pada model *K-Nearest Neighbor* dengan seleksi fitur *Information Gain* dan optimasi *Particle Swarm Optimization*, K=5 dengan akurasi 87,33% lebih tinggi daripada akurasi pada nilai K yang lain.



Gambar 2. Perbandingan Akurasi

3.7. Evaluasi dengan Confusion Matrix

Pada tahap evaluasi akan dicari akurasi menggunakan *confusion matrix*. Pada klasifikasi *K-Nearest Neighbor* K=9 dengan akurasi 82,67% terdapat 52 data negatif dan 3 data positif yang di prediksi sebagai data negatif. Kemudian terdapat 23 data negatif dan 72 data positif yang dianggap sebagai prediksi data positif. Adapun penjabaran perhitungan dari akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{52+72}{52+23+3+72} \times 100\% = 82,67\%$$

Pada klasifikasi *K-Nearest Neighbor* dengan seleksi fitur *Information Gain*, $K=7$ dengan akurasi 82% terdapat 56 data negatif dan 8 data positif yang di prediksi sebagai data negatif. Kemudian terdapat 19 data negatif dan 67 data positif yang dianggap sebagai prediksi data positif. Adapun penjabaran perhitungan dari akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{56+67}{56+19+8+67} \times 100\% = 82\%$$

Pada klasifikasi *K-Nearest Neighbor* dengan seleksi fitur *Information Gain* dan optimasi *Particle Swarm Optimization*, $K=5$ dengan akurasi 87,33% terdapat 62 data negatif dan 6 data positif yang di prediksi sebagai data negatif. Kemudian terdapat 13 data negatif dan 69 data positif yang dianggap sebagai prediksi data positif. Adapun penjabaran perhitungan dari akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{62+69}{62+13+6+69} \times 100\% = 87,33\%$$

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, pada klasifikasi *K-Nearest Neighbor* didapatkan akurasi terbaik adalah dengan nilai $K=9$ sebesar 82,67%. Pada klasifikasi *K-Nearest Neighbor* dengan seleksi fitur *Information Gain* didapatkan akurasi terbaik adalah dengan nilai $K=7$ sebesar 82%. Pada klasifikasi *K-Nearest Neighbor* seleksi fitur *Information Gain* dan Optimasi *Particle Swarm Optimization* didapatkan akurasi terbaik adalah dengan nilai $K=5$ sebesar 87,33%. Klasifikasi *K-Nearest Neighbor* dengan seleksi fitur *Information Gain* dan optimasi *Particle Swarm Optimization* lebih unggul daripada klasifikasi *K-Nearest Neighbor* tanpa seleksi fitur dan tanpa optimasi serta lebih unggul dari klasifikasi *K-Nearest Neighbor* dengan seleksi fitur *Information Gain*. Untuk penelitian selanjutnya menggunakan nilai K yang lain pada metode klasifikasi *K-Nearest Neighbor* serta menggunakan metode seleksi fitur dan metode optimasi yang lain agar mendapatkan pengaruh yang signifikan.

5. DAFTAR PUSTAKA

- [1] Syarifuddin, M., 2020. Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn. *Inti Nusa Mandiri*, 15(1), pp.23-28.
- [2] Hidayah, N. and Sahibu, S., 2021. Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(4), pp.820-826.
- [3] Irawan, F.R., Jazuli, A. and Khotimah, T., 2022. Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest

- Neighbors. *JIKO (Jurnal Informatika dan Komputer)*, 5(1), pp.62-68.
- [4] Lestari, D.A. and Mahdiana, D., 2021. Penerapan Algoritma K-Nearest Neighbor pada Twitter untuk Analisis Sentimen Masyarakat Terhadap Larangan Mudik 2021. *Informatik: Jurnal Ilmu Komputer*, 17(2), pp.123-131.
- [5] Desiani, A., 2022. Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati. *Jurnal Sistem Informasi dan Sistem Komputer*, 7(2), pp.104-110.
- [6] Isnanda, A., Umaidah, Y. and Jaman, J.H., 2021. Implementasi Naïve Bayes Classifier Dan Information Gain Pada Analisis Sentimen Penggunaan E-Wallet Saat Pandemi. *Jurnal Teknologi Informatika dan Komputer*, 7(2), pp.144-153.
- [7] Aini, S.H.A., 2018. *Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes* (Doctoral dissertation, Universitas Brawijaya).
- [8] Que, V.K.S., Iriani, A. and Purnomo, H.D., 2020. Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* Vol. 9(2).
- [9] Mahardika, K.W., 2018. *Optimasi K-Nearest Neighbour Menggunakan Particle Swarm Optimization pada Sistem Pakar untuk Monitoring Pengendalian Hama pada Tanaman Jeruk* (Doctoral dissertation, Universitas Brawijaya).
- [10] Putri, F.O. and Indriati, R.C.W., Analisis Sentimen pada Ulasan Pengguna MRT Jakarta Menggunakan Metode Neighbor-Weighted K-Nearest Neighbor dengan Seleksi Fitur Information Gain. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548*, p.964X.
- [11] Utomo, P.E.P., Manaar, M., Khaira, U. and Suratno, T., 2019. Analisis Sentimen Online Review Pengguna Bukalapak Menggunakan Metode Algoritma TF-IDF. *JUSS (Jurnal Sains dan Sistem Informasi)*, 2(2), pp.35-39.
- [12] Al Faraby, S., 2021. Analisis Sentimen Pada Ulasan Produk Kecantikan Menggunakan K-nearest Neighbor Dan Information Gain. *eProceedings of Engineering*, 8(5).
- [13] Rahman, A.F. and Indriati, M., Analisis Sentimen Penggunaan Tol Trans Jawa Periode Mudik Lebaran 2019 dengan Metode K-Nearest Neighbor dan Seleksi Fitur Information Gain. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548*, p.964X.
- [14] Negara, A.B.P., Muhandi, H. and Putri, I.M., 2020. Analisis Sentimen Maskapai Penerbangan

- Menggunakan Metode Naive Bayes dan Seleksi Fitur *Information Gain*. *J. Teknol. Inf. dan Ilmu Komput.*, 7(3).
- [15] Nuris, N., Yulia, E.R. and Solecha, K., 2021. Implementasi Particle Swarm Optimization (PSO) Pada Analysis Sentiment Review Aplikasi Halodoc Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informasi*, 7(1), pp.17-23.
- [16] Akbar, F.R., 2020. *Implementasi Analisis Data Kredit Nasabah Menggunakan Metode K-Nearest Neighbors* (Doctoral dissertation, Institut Teknologi Nasional Malang).

Riana

ORIGINALITY REPORT

32%

SIMILARITY INDEX

30%

INTERNET SOURCES

10%

PUBLICATIONS

12%

STUDENT PAPERS

PRIMARY SOURCES

1	repository.ub.ac.id Internet Source	6%
2	jtiik.ub.ac.id Internet Source	4%
3	ejournal.unkhair.ac.id Internet Source	2%
4	Submitted to Universitas Bina Sarana Informatika Student Paper	2%
5	docplayer.info Internet Source	1%
6	pdfs.semanticscholar.org Internet Source	1%
7	ejournal.unida.gontor.ac.id Internet Source	1%
8	Submitted to Universitas Brawijaya Student Paper	1%
9	123dok.com Internet Source	1%

10	repository.uin-suska.ac.id Internet Source	1 %
11	Submitted to Universitas Islam Lamongan Student Paper	1 %
12	journal.thamrin.ac.id Internet Source	1 %
13	Submitted to Universitas Gunadarma Student Paper	1 %
14	repository.nusamandiri.ac.id Internet Source	1 %
15	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	1 %
16	Muhammad Irfan Saputra, Irwan Budiman, Dwi Kartini, Dodon Turianto Nugrahadi, Mohammad Reza Faisal. "Performance Comparison of Adaptive Neuro Fuzzy Inference System and Support Vector Machine Algorithm in Balanced and Unbalanced Multiclass Data Classification", 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), 2021 Publication	<1 %
17	Submitted to Bellevue Public School Student Paper	<1 %

journal.unipdu.ac.id:8080

18	Internet Source	<1 %
19	repository.mercubuana.ac.id Internet Source	<1 %
20	journal.uii.ac.id Internet Source	<1 %
21	Submitted to Sriwijaya University Student Paper	<1 %
22	ejournal.nusamandiri.ac.id Internet Source	<1 %
23	kabar-terhangat.blogspot.com Internet Source	<1 %
24	technoscience.akprind.ac.id Internet Source	<1 %
25	tunasbangsa.ac.id Internet Source	<1 %
26	Submitted to Universitas Negeri Semarang Student Paper	<1 %
27	Submitted to Universitas Sanata Dharma Student Paper	<1 %
28	ejurnal.ubharajaya.ac.id Internet Source	<1 %
29	eprints.poltekkesjogja.ac.id Internet Source	<1 %

30	j-ptiik.ub.ac.id Internet Source	<1 %
31	repository.telkomuniversity.ac.id Internet Source	<1 %
32	Rakhmi Khalida, Siti Setiawati. "Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain", Journal of Informatic and Information Security, 2020 Publication	<1 %
33	ejournal.uin-suka.ac.id Internet Source	<1 %
34	ejournal.upnvj.ac.id Internet Source	<1 %
35	elib.gstu.by Internet Source	<1 %
36	www.ejournal.ust.ac.id Internet Source	<1 %
37	www.researchgate.net Internet Source	<1 %
38	Titin Winarti, Henny Indriyawati, Vensy Vydia, Febrian Wahyu Christanto. "Performance comparison between naive bayes and k-nearest neighbor algorithm for the classification of Indonesian language articles",	<1 %

39

Rizqi Darmawan, Indra Indra, Asep Surahmat.
"Optimalisasi Support Vector Machine (SVM)
Berbasis Particle Swarm Optimization (PSO)
Pada Analisis Sentimen Terhadap Official
Account Ruang Guru di Twitter", Jurnal Kajian
Ilmiah, 2022

Publication

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On

Riana

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7
