OPTIMASI METODE K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK ANALISIS SENTIMEN PEMILIHAN PRESIDEN INDONESIA TAHUN 2024-2029

Leonardus Silih Windanu, Anggri Sartika Wiguna, Alexius Endy Budianto
Teknik Informatika, Universitas PGRI Kanjuruhan Malang
Jalan S. Supriadi No.48 Malang, Indonesia
nardusleo303@gmail.com

ABSTRAK

Indonesia merupakan negara ke 5 pengguna Twitter terbanyak didunia dengan jumlah 19,5 juta. Twitter banyak digunakan oleh masyarakat sebagai wadah untuk menuangkan pendapatnya akan suatu peristiwa yang sedang terjadi atau biasanya peristiwa yang sedang viral dengan menggunakan hastag. Belakangan ini isu tentang pemilihan presiden Indonesia yang baru pada tahun 2024-2029 merupakan salah satu topik yang sedang viral, beragam *cuitan* muncul yang tentunya masih teracak atau belum dikategorikan. Agar pendapat tersebut bisa bermanfaat dan berguna, diperlukan beberapa proses sehingga didapatlah informasi yang penting dan berguna dengan analisis sentimen. Pada penelitian ini melakukan optimasi pada metode *K-Nearest Neighbor* (KNN) dengan menggunakan algoritma pengoptimasi *Particle Swarm Optimization* (PSO) untuk analisis sentimen pemilihan presiden Indonesia tahun 2024-2029, dikarenakan metode KNN relatif menghasilkan akurasi yang lebih rendah dibandingkan dengan metode lain, hal ini bertujuan untuk meningkatkan hasil akurasi yang didapatkan. Hasil akurasi menggunakan metode KNN diuji dengan K=1 sampai dengan K=10, lalu dibandingkan dengan K terbaik dari hasil metode KNN yang sudah dioptimasi menggunakan algoritma PSO guna mendapatkan model terbaik. Dari hasil pengujian yang telah dilakukan didapatkan model terbaik yaitu pada pengujian metode KNN yang dioptimasi menggunakan algoritma PSO pada *rasio dataset* 70:30 dengan akurasi 77,3% pada K=7 menggunakan partikel sebanyak 10 dan iterasi 20.

Kata kunci : *K-Nearest Neighbor, Particle Swarm Optimization, Analisis Sentimen*

1. PENDAHULUAN

Indonesia merupakan salah satu negara dengan kemajuan yang sangat pesat dalam bidang teknologi dan informasi, hal itu terlihat dari jumlah pengguna sosial media yang semakin bertambah disetiap harinya, baik itu Facebook, Instagram, Twitter, dan sosial media lainnya. Pengguna Twitter di Indonesia pada tahun 2022 mencapai angka 19,5 juta dan menjadi yang terbanyak ke lima di dunia [1].

Pemilihan presiden Indonesia yang baru akan dilaksanakan pada tahun 2024. Maraknya isu pemilihan presiden ini menjadikan Twitter sabagai wadah untuk sebagian masyarakat Indonesia dalam menuangkan pendapatnya. Karena kebebasan berpendapat di Twitter membuat *tweet* yang dibuat oleh pengguna Twitter menjadi teracak atau belum dikategorikan, sehingga sulit dipahami apa yang menjadi maksud dan tujuan dari *cuitan* tersebut. Banyak pendapat positif maupun negatif yang sebenarnya bisa menjadi acuan atau gambaran seperti apa keinginan dan harapan sebagian masyarakat akan pelaksanaan pemilihan presiden ini, baik itu tentang bakal calon, partai politik, dan isu-isu yang terkait dengan pemilihan presiden. Agar pendapat atau opini tersebut dapat dimanfaatkan dan berguna, diperlukan beberapa proses sehingga didapatkan suatu informasi yang penting melalui analisis sentimen [2].

Dalam pemrosesan analisis sentimen terdapat beberapa metode yang bisa digunakan, salah satunya adalah metode *K-Nearest Neighbor* (KNN). KNN adalah salah satu metode klasifikasi yang sering dipakai dan dinilai efektif dalam melakukan klasifikasi, terutama dengan menggunakan *data training* yang besar [3]. Namun pada kenyataannya metode ini memiliki kelemahan dibandingkan dengan metode yang lainnya, terutama pada tingkat akurasinya yang dikarenakan pemilihan seleksi fitur, nilai k bias, dan mudah tertipu dengan atribut yang tidak relevan [4]. Sehingga perlu ditingkatkan dengan sebuah algoritma pengoptimal untuk menutupi kelemahannya. *Particle swarm optimization* (PSO) merupakah salah satu algoritma yang banyak dipakai dalam memecahkan masalah optimasi dan seleksi fitur. Algoritma ini dirumuskan oleh (Kennedy dan Eberhart pada tahun 1995) dengan proses pemikiran yang terinspirasi dari perilaku sosial hewan, seperti sekelompok burung atau ikan yang berkelompok. PSO berisi tentang suatu perubahan perilaku yang terdiri atas tindakan setiap individu dan besarnya pengaruh kedalam satu kelompok.

Pada penelitian yang dilakukan oleh [5] dengan judul "Perbandingan Metode *K-Nearest Neighbor, Naïve Bayes,* dan *Decision Tree* untuk Prediksi Kelayakan Pemberian Kredit" menunjukan bahwa metode *Decision Tree* yang memiliki akurasi terbaik dengan nilai 92,21%, pada urutan kedua metode *Naïve Bayes* dengan akurasi 81,83%, dan metode KNN yang memiliki akurasi terendah diantara ketiga metode dengan nilai 81,82%. Pada penelitian [6] membandingkan kinerja dari dua metode yakni

Support Vector Machine (SVM) dan KNN. Dari nilai akurasi yang didapat menyimpulkan bahwa SVM lebih baik dari pada KNN, dengan nilai akurasi 95% sedangkan KNN pada angka 80%. Dari beberapa penelitian ini dapat disimpulkan bahwa metode KNN relatif menghasilkan akurasi yang lebih rendah dibandingkan dengan metode lain. Sedangkan penelitian oleh [7] melakukan prediksi penyakit ginjal dengan algoritma KNN berbasis PSO justru mendapatkan hasil yang lebih baik setelah menambahkan algoritma PSO dengan mendapatkan akurasi pada angka 97,25%, yang mana sebelum ditambahkan dengan algoritma PSO hanya mendapatkan akurasi sebesar 78,75%. Dari penelitian ini dapat disimpulkan bahwa algoritma PSO dapat mengoptimalkan akurasi pada metode KNN yang mana hal ini yang menjadi kelemahannya.

Berdasarkan penjelasan diatas, penelitian ini berfokus pada optimasi metode KNN dengan algoritma PSO yang bertujuan untuk meningkatkan akurasi yang dihasilkan serta mencari model terbaik dari hasil klasifikasi.

2. TINJAUAN PUSTAKA

Adapun landasan teori yang dipakai sebagai pendukung penelitian ini adalah data mining, analisis sentimen, *text mining*, *pre-processing*, pembobotan *tf-idf, k-nearest neighbor, particle swarm optimization,* dan *confusion matriks.*

2.1. Data Mining

Data mining adalah proses menggali informasi yang bermanfaat dari data besar dan kompleks. Hal ini melibatkan penggunaan teknologi model penalaran, matematika untuk menemukan pola, teknik statistik, keterkaitan, dan tren yang bermanfaat dalam data. Tujuan dari data mining adalah untuk menghasilkan informasi yang bermanfaat dan dapat dimanfaatkan dalam pengambilan keputusan, perencanaan bisnis, atau untuk menemukan wawasan baru yang dapat membantu organisasi dalam meningkatkan efisiensi dan produktivitasnya [7].

2.2. Analisis Sentimen

Analisis sentimen ialah bidang keilmuan yang mempelajari opini, evaluasi, sentimen, sikap, penilaian, dan emosi seseorang terhadap suatu objek seperti barang, orang, peristiwa, organisasi, dan masalah konkrit. Tujuan utama dari analisis sentimen adalah untuk memahami dan menganalisis opini dan sentimen yang diungkapkan oleh orang-orang dalam teks atau media sosial. Analisis sentimen banyak digunakan dalam berbagai aplikasi seperti pemantauan merek, penelitian pasar, politik, dan manajemen reputasi online [8].

2.3. Text Mining

Text mining ialah sebuah ilmu yang berguna untuk mengolah teks dengan menggunakan metode statistik dan komputasi untuk mengidentifikasi pola dan kecenderungan dalam data teks. Tujuannya adalah untuk menghasilkan informasi yang berguna bagi pengguna dari data teks yang besar dan kompleks. Dalam aplikasi praktis, *text mining* dapat digunakan untuk analisis sentimen, analisis topik, dan pengenalan entitas [9].

2.4. Pre-Processing

Pre-processing atau pra-pemrosesan adalah tahap awal dan penting dalam *text mining* yang dilakukan guna mengubah data teks mentah menjadi format yang sesuai dan dapat diolah lebih lanjut. *Pre-processing* bertujuan untuk meningkatkan kualitas data, menghilangkan noise dan mengubah format data agar lebih mudah diproses. Beberapa langkah dalam *pre-processing* teks meliputi *case folding, tokenisasi* dan *frekuensi token, stopword removal,* normalisasi kata, dan *stemming* [10].

Tabel 1. Tahapan pre-processing

No	Tahapan	Penjelasan
1	*Case Folding*	Merubah seluruh kata menjadi *lower case.*
2	*Tokenisasi, Frekuensi Token*	Memisahkan kata menjadi pertoken lalu menghapus url, tanda baca, dan emoji. Menghitung kemunculan kata didalam dokumen.
3	*Stopword Removal*	Menghapus kata yang tidak memiliki makna penting.
4	Normalisasi Kata	Mengembalikan kata singkat dan slang ke kata sebenarnya.
5	*Steeming*	Mengembalikan kata menjadi kata dasar.

2.5. Pembobotan TF-IDF

Tf-Idf (Term Frequency-Inverse Document Frequency) adalah metode pembobotan yang umum digunakan dalam pemrosesan teks untuk memberikan bobot pada setiap kata yang telah diekstraksi. Metode ini memberikan bobot yang lebih tinggi pada kata-kata yang muncul lebih sering dalam suatu dokumen dan bobot yang lebih rendah pada kata-kata yang muncul dalam banyak dokumen [11].

Persamaan *tf-idf* yang dipakai dalam penelitian :

$$Wi,j = tfi,j \log\left(\frac{N+1}{dfi+1}\right) + 1 \quad (1)$$

Penjelasan :

Wi,j : Banyak data ke-i terhadap kata ke-j
Tfi,j : Banyaknya kata i yang dicari pada sebuah data j
N : Total data
dfi : Banyaknya data yang mengandung kata ke-i

2.6. K-Nearest Neighbor

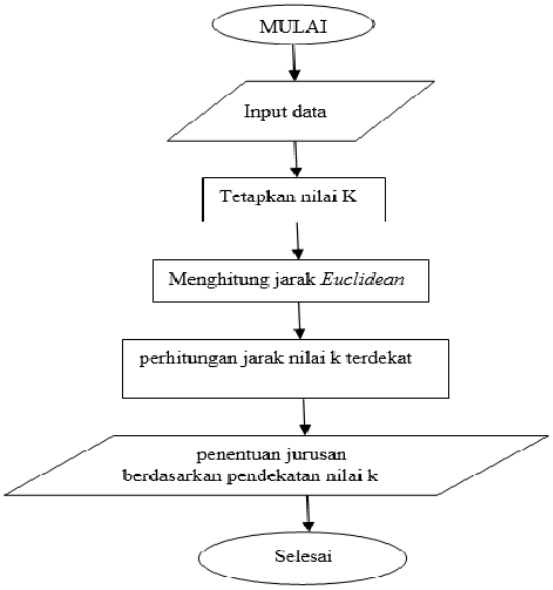
K-Nearest Neighbor (KNN) ialah metode klasifikasi yang banyak dipakai dalam

mengklasifikasikan data, terutama dalam konteks data teks. Inti dari metode ini yaitu untuk mengklasifikasikan objek baru berdasarkan kategori objek yang paling sering muncul di antara k tetangga terdekatnya dalam dataset pelatihan. Jadi, jika ada objek baru yang ingin diklasifikasikan, algoritma KNN akan mencari k tetangga terdekatnya dalam dataset pelatihan dan menggunakan mayoritas kategori dari tetangga-tetangganya untuk memprediksi kategori objek baru tersebut [12].

Berikut persamaan KNN menggunakan *euclidien distance*.

$$d(A,B)= \sqrt{(A1-B1)^2+(A2-B2)^2+\ldots\ldots+|A_i-B_i|^2} \quad (2)$$

Dimana :
d(A,B) : jarak dari dokumen A ke dokumen B
Ai : kata ke i di dokumen A
Bi : kata ke i di dokumen B



Gambar 1. Flowchart metode knn

2.7. Particle Swarm Optimization

Algoritma *Particle Swarm Optimization* (PSO) berdasarkan pada perilaku kelompok serangga seperti lebah, rayap, semut atau burung. PSO menggunakan model matematis dari perilaku kelompok serangga untuk memecahkan masalah optimasi. PSO meniru perilaku kelompok serangga dengan menempatkan sejumlah partikel acak dalam ruang pencarian. Setiap partikel mewakili solusi kandidat untuk masalah optimasi dan memiliki posisi dan kecepatan saat ini. Setiap partikel memperbarui posisinya dan kecepatannya berdasarkan pengalaman sendiri dan pengalaman kelompoknya. Setiap partikel mencari solusi terbaik yang pernah ditemukannya, yang disebut sebagai *pbest*, dan solusi terbaik yang pernah ditemukan oleh seluruh kelompok, yang disebut sebagai gbest. Setiap partikel mengubah kecepatannya berdasarkan *pbest* dan *gbest*, serta menggerakkan dirinya menuju solusi yang lebih baik [13].

Berikut persamaan algoritma PSO. Posisi partikel pada dimensi ruang tertentu dapat dinyatakan dengan :

$$X_i = x_{i1}(t), x_{i2}(t), \ldots, x_{iN}(t) \quad (3)$$

$$V_i(t) = v_{i1}(t), v_{i2}(t), \ldots, v_{iN}(t) \quad (4)$$

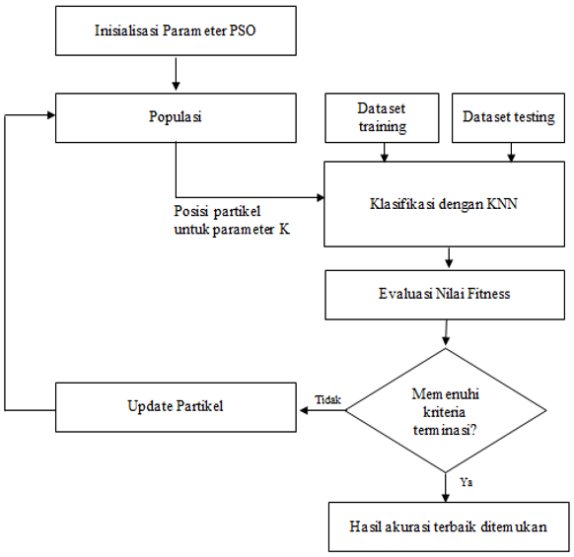
Dengan :
X : Posisi partikel
V : Kecepatan Partikel
i : Indeks partikel
t : Iterasi ke-t
N : Ukuran dimensi ruang

Model matematika untuk mengupdate posisi partikel dalam PSO sebagai berikut :

$$V_i(t) = V_1(t-1) + c_1 r_1 \left[X_{pbest_i} - X_i(t)\right] + c_2 r_2 \left[X_{Gbest_i} - X_i(t)\right] - X_i(t) \quad (5)$$

$$X_i(t) = X_i(t-1) + V_i(t) \quad (6)$$

Keterangan :
Vi(t) :Kecepatan partikel i saat iterasi t
Xi(t) :Posisi partikel i saat iterasi t
c1 dan c2:Learning rates untuk kemampuan individu (cognitive) dan pengaruh sosial (group)
r1 dan r2: Bilangan random yang berdistribusi uniformal dalam interval 0 dan 1
XPbesti : Posisi terbaik partikel i
XGbesti : Posisis terbaik global



Gambar 2. Flowchart algoritma pso

2.8. Confusion Matriks

Confusion matriks adalah metode evaluasi performa untuk model klasifikasi. *Confusion matriks*

menampilkan seberapa banyak data yang diklasifikasikan benar maupun salah oleh model [14]. Dengan menggunakan *confusion matriks*, kita dapat mengevaluasi dan memperbaiki model klasifikasi sehingga dapat menghasilkan prediksi yang lebih akurat.

Berikut persamaan yang dipakai dalam *confusion matriks :*

Tabel 2. Persamaan confusion matriks

		Prediksi	
		Positif	Negatif
Aktual	Positif	A	B
	Negatif	C	D

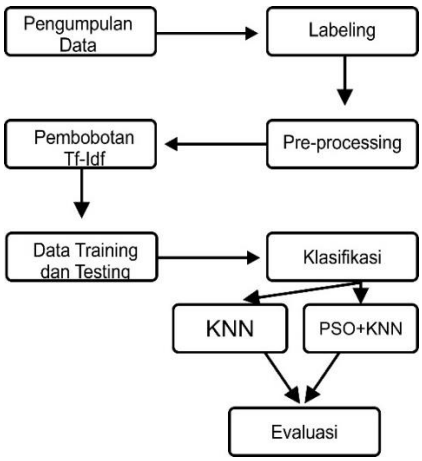
Dimana :
A : Prediksi positif aktual positif
B : Prediksi negatif aktual positif
C : Prediksi positif aktual negatif
D : Prediksi negatif aktual negatif

Menghitung akurasi menggunakan persamaan berikut :

$$\textit{Confusion Matriks} = \frac{A+D}{A+B+C+D} \text{ X } 100\% \quad (6)$$

3. METODE PENELITIAN

Dalam melakukan optimasi metode *k-nearest neighbor* berbasis *particle swarm optimization* untuk analisis sentimen, berikut rancangan penelitian yang dipakai.



Gambar 3. Rancangan penelitian

Dari gambar diatas dapat dilihat rancangan dalam penelitian ini terdiri dari pengumpulan data, labeling, *pre-processing*, pembobotan *tf-idf*, *data testing* dan *training*, klasifikasi, dan evaluasi.

3.1. Pengumpulan Data

Penelitian ini menggunakan data yang didapatkan dari *tweet* di Twitter yang menggunakan bahasa Indonesia. Data diambil menggunakan website *netlytic.org* sebanyak 1000 tweet yang diambil dengan menggunakan kata kunci #pilpres2024, data diambil pada tanggal 15 Januari 2023 dalam bentuk format *csv*. Data diambil dari media sosial Twitter dikarenakan Twitter merupakan salah satu media sosial yang popular dan masih membuka datanya untuk diteliti oleh pihak ketiga [15].

3.2. Labeling

Pada penelitian ini pelabelan dilakukan secara manual oleh peneliti. Hasil pelabelan akan dibagi kedalam dua sentimen yaitu positif dan negatif, peneliti akan membaca dan mengkaji 1000 *tweet* secara teliti satu-persatu lalu akan diberikan label sesuai dengan isi *tweet* lebih mengarah ke positif ataupun negatif.

3.3. Pre-Processing

Tahapan awal dalam proses klasifikasi menggunakan metode KNN berbasis PSO adalah proses *pre-processing. Pre-processing* bertujuan untuk mengolah data mentah menjadi data yang siap digunakan untuk proses klasifikasi.

3.4. Pembobotan TF-IDF

Setelah melalui tahapan *pre-processing*, selanjutnya adalah penghitungan bobot dengan *tf-idf*. Pembobotan *tf-idf* merupakan salah satu metode umum yang dipakai dalam pengolahan teks. Tujuannya adalah untuk mengukur relevansi setiap kata dalam sebuah dokumen berdasarkan seberapa sering kata tersebut muncul di dokumen tersebut dan seberapa umum kata tersebut dalam seluruh korpus dokumen.

3.5. Data Testing dan Training

Data testing dan *data training* dibagi secara otomatis menggunakan metode *splitting data*, dengan 3 perbandingan *rasio dataset*, yaitu 90:10, 80:20, dan 70:30. Pembagian dilakukan secara acak memakai *random state*, sehingga setiap dieksekusi menghasilkan nilai yang stabil.

3.6. Klasifikasi

Klasifikasi merupakan proses analisis data yang membantu untuk mengelompokkan sampel data ke dalam kelas-kelas tertentu dan menemukan hubungan antara atribut-atribut yang terkait [16]. Pada tahapan ini dilakukan proses klasifikasi metode KNN yang dioptimasi dengan algoritma PSO guna meningkatkan akurasi yang dihasilkan dan mencari model terbaik. Dalam pengujian metode KNN menggunakan K=1 sampai dengan K=10, lalu diambil K yang paling optimal dari metode KNN dan K paling optimal dari pengujian menggunakan metode KNN berbasis PSO untuk melihat perbedaan akurasi yang dihasilkan.

3.7. Evaluasi

Dalam penelitian ini menghitung akurasi yang dihasilkan memakai *confusion matriks* untuk melihat keberhasilan sistem melakukan klasifikasi. Tahapan menghitung akurasi yaitu dengan cara menjumlahkan

hasil positif benar dengan negatif salah, lalu dibagi dengan hasil penjumlahan dari positif benar, negatif benar, positif salah, dan negatif salah. Lalu dikalikan dengan 100%.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang dipakai dalam penelitian ini berjumlah 1000. Data merupakan *tweet* berbahasa Indonesia yang diambil menggunakan kata kunci #pilpres2024. Atribut yang dipakai pada tahap awal adalah *description* yang merupakan isi *tweet*.

Tabel 3. Data tweet

Description
Emil Dardak Siap Dukung dan Menangkan AHY #PartaiDemokrat #AHYPimpinPerubahan https://t.co/gTMFX0nXf8
"Sampai hari ini PKB masih bersama Gerindra untuk koalisi Pilpres," kata Muhaimin. https://t.co/BvugLqnsYU
…
@alisyarief mohon maaf sblmnya, jika melihat reel antusias masyarakat terhadap pencapresan pak anies maka sy yakin hanya kecurangan yg mampu mengalahkan pak anies di pilpres 2024 nanti selain daripada takdir Allah

4.2. Labeling

Pada tahapan ini memberikan label dari *tweet* yang akan dibuat kedalam atribut baru yaitu sentimen. Didalam atribut sentimen berisikan label positif dan negatif, dalam hal ini label tersebut akan *diraplace* kedalam bentuk *numerik*, sentimen positif akan *direplace* menjadi angka 0 dan sentimen negatif akan *direplace* menjadi angka 1. Berikut adalah data yang telah melalui tahapan tersebut.

Tabel 4. Data hasil labeling

Description	sentimen
Emil Dardak Siap Dukung dan Menangkan AHY #PartaiDemokrat #AHYPimpinPerubahan https://t.co/gTMFX0nXf8	0
"Sampai hari ini PKB masih bersama Gerindra untuk koalisi Pilpres," kata Muhaimin. https://t.co/BvugLqnsYU	0
…	…
@alisyarief mohon maaf sblmnya, jika melihat reel antusias masyarakat terhadap pencapresan pak anies maka sy yakin hanya kecurangan yg mampu mengalahkan pak anies di pilpres 2024 nanti selain daripada takdir Allah	1

Dari 1000 data yang telah dilakukan pelabelan secara manual oleh peneliti, mendapatkan 728 sentimen bernilai positif dan 272 sentimen bernilai negatif.

4.3. Pre-Processing

Setelah dilakukan pelabelan pada data, berikutnya akan dilakukan *pre-processing* data yang terbagi atas beberapa tahap, yakni *case folding, tokenisasi* dan *frekuensi token, stopword removal,* normalisasi kata, dan *steeming.* Berikut adalah data yang telah melalui tahapan *pre-processing.*

Tabel 5. Data hasil pre-processing

Contoh Data	Emil Dardak Siap Dukung dan Menangkan AHY
Case Folding	emil dardak siap dukung dan menangkan ahy
Tokenisasi, *Frekuensi Token*	['emil', 'dardak', 'siap', 'dukung', 'dan', 'menangkan', 'ahy'] ['emil(1)', 'dardak(1)', 'siap(1)', 'dukung(1)', 'dan(1)', 'menangkan(1)', 'ahy(1)']
Stopword Removal	['emil', 'dardak', 'dukung', 'menangkan', 'ahy']
Normalisasi kata	['emil', 'dardak', 'dukung', 'menangkan', 'ahy']
Stemming	['emil', 'dardak', 'dukung', 'menang', 'ahy']

4.4. Pembobotan TF-IDF

Setelah data melalui tahapan *pre-processing*, selanjutnya masuk ketahap pembobotan menggunakan *tf-idf* yang berguna untuk melihat sejauh mana tingkat relevan sebuah kata didalam dokumen. Berikut hasil dari pembobotan menggunakan *tf-idf.*

Tabel 6. Hasil pembobotan tf-idf

	Amin	Ab	…	Yuk	Yunan
1	0.0	0.0		0.0	0.0
2	0.209	0.0	…	0.0	0.0
3	0.0	0.0	…	0.0	0.0
4	0.0	0.0	…	0.0	0.0
…	…	…	…	…	…
999	0.0	0.0	…	0.0	0.198

4.5. Klasifikasi KNN

Setelah melalui beberapa tahapan sampai data bersih, selanjutnya masuk ketahapan klasifikasi. Tahapan klasifikasi akan terbagi kedalam 2 model, yang pertama menggunakan metode KNN dan kedua menggunakan metode KNN berbasis PSO pada tahapan berikutnya. Proses ini dilakukan guna meningkatkan performa dari metode KNN dan untuk mencari model terbaik dari penelitian. Pengujian akan menggunakan 3 *rasio dataset*, yaitu 90:10, 80:20, dan 70:30.

Tahapan awal yang dilakukan dalam pengujian menggunakan *python* adalah *import KNeighborsClassifier* dari *sklearn*. Selanjutnya *import accuracy score* dari *sklearn* untuk mengukur akurasi yang diperoleh menggunakan metode terhadap *3 rasio dataset* pengujian. Pengujian dilakukan dengan nilai K=1 sampai dengan K=10, lalu diambil nilai K yang paling optimal. Berikut akurasi yang dihasilkan menggunakan algoritma KNN.

Tabel 7. Akurasi metode knn (%)

K	Dataset		
	90:10	80:20	70:30
1	68	73,5	72,67
2	74	75,5	74,67
3	69	72,5	71,34
4	71	74	73,3
5	71	72	73,67
6	71	72,5	74
7	75	72,5	75
8	74	73,5	75,3
9	73	73	75
10	74	73,5	75,67

Dari tabel diatas dapat disimpulkan bahwa akurasi terbaik dari pengujian *dataset* 90:10 adalah 75% pada K=7, *dataset* 80:20 adalah 75,5% pada K=2, dan *dataset* 70:30, dengan akurasi 75,67% pada K=10.

4.6. Klasifikasi KNN berbasis PSO

Pada pengujian ini yaitu menerapakan algoritma PSO pada KNN untuk mengoptimasi nilai K pada metode KNN. Langkah pertama adalah *import pyswarms as ps* selanjutnya *import single_obj as fx* dari *pyswarms.utils.functions*, *library* ini lah yang akan menghitung proses optimasi KNN berbasis PSO.

Penelitian oleh [17] melakukan analisis sentimen dengan metode KNN berbasis PSO. Pada penelitian tersebut mendapatkan akurasi terbaik menggunakan *iterasi* sebanyak 20 dan *partikel* berjumlah 10, komponen sosial pertama (c1 = 1), komponen sosial kedua (c2 = 1), dan *interia weight* (w = 1) sehingga parameter tersebut dipakai didalam penelitian ini, akan tetapi untuk *partikel* akan diuji menggunakan kombinasi sebanyak 5, 10, 15, 20, dan 25 guna mendapatkan hasil yang lebih bervariasi untuk mendapatkan K yang paling optimal dalam penelitian. Berikut akurasi yang dihasilkan dari pengujian.

Tabel 8. Akurasi metode knn+pso dataset 90:10 (%)

Partikel	W	C1	C2	Best Cost	K	Akurasi
5	1	1	1	0,24	4	72
10	1	1	1	0,24	7	76
15	1	1	1	0,24	11	74
20	1	1	1	0,24	14	74
25	1	1	1	0,24	18	73

Berdasarkan tabel diatas maka dapat diangkat kesimpulan pada pengujian metode KNN berbasis PSO pada *dataset* 90:10 mendapatkan model paling optimal dengan akurasi 76% pada K=7 dan *best cost*=0,24 dengan pengujian menggunakan partikel 10 dan iterasi 20.

Tabel 9. Akurasi metode knn+pso dataset 80:20 (%)

Partikel	W	C1	C2	Best Cost	K	Akurasi
5	1	1	1	0,24	6	73
10	1	1	1	0,24	10	76
15	1	1	1	0,24	11	76
20	1	1	1	0,24	15	75,5
25	1	1	1	0,24	25	75

Berdasarkan tabel diatas maka dapat diangkat kesimpulan pada pengujian metode KNN berbasis PSO pada *dataset* 80:20 mendapatkan model paling optimal dengan akurasi 76% pada K=10 dan *best cost*=0,24 dengan pengujian menggunakan partikel 10 dan iterasi 20 dan pada K=11 dan *best cost*=0,24 dengan pengujian menggunakan partikel 15 dan iterasi 20.

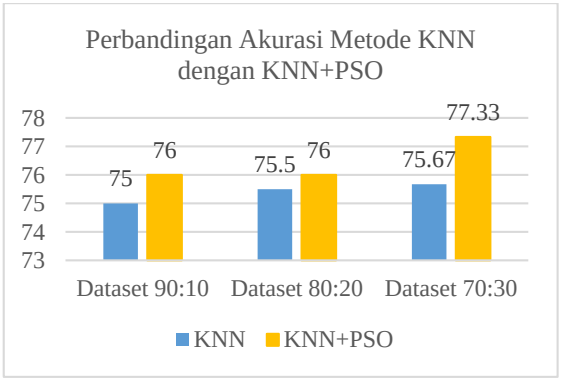
Tabel 10. Akurasi metode knn+pso dataset 70:30 (%)

Partikel	W	C1	C2	Best Cost	K	Akurasi
5	1	1	1	0,27	3	72,67
10	1	1	1	0,27	7	77,3
15	1	1	1	0,27	10	77
20	1	1	1	0,27	15	75
25	1	1	1	0,27	17	75

Berdasarkan tabel diatas maka dapat diangkat kesimpulan pada pengujian metode KNN berbasis PSO pada *dataset* 70:30 mendapatkan model paling optimal dengan akurasi 77,3% pada K=7 dan *best cost*=0,27 dengan pengujian menggunakan partikel 10 dan iterasi 20.

4.7. Rangkuman Penelitian

Setelah melalui 2 tahapan klasifikasi dan mendapatkan akurasi terbaik dengan menggunakan 3 *rasio dataset*, berikut rangkuman dari perbandingan metode KNN dan metode KNN berbasis PSO.

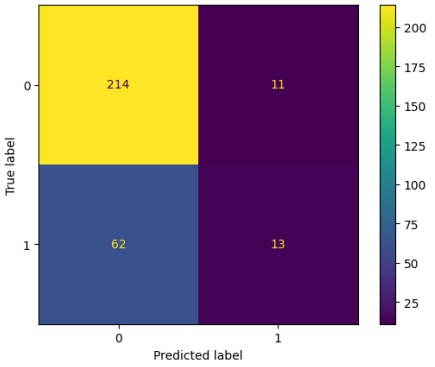


Gambar 4. Perbandingan akurasi metode knn dengan knn+pso (%)

Dari gambar diatas dapat dilihat bahwa algoritma PSO selalu berhasil meningkatkan akurasi awal yang dihasilkan oleh metode KNN.

4.8. Evaluasi KNN

Setelah melalui tahapan pengujian, masuk ketahap evaluasi model menggunakan *confusion matriks*. Klasifikasi menggunakan metode KNN, akurasi terbaik didapatkan pada dataset 70:30, sebagai berikut.



Gambar 5. Confusion matriks metode knn

Nilai akurasi dihitung dengan persamaan berikut:

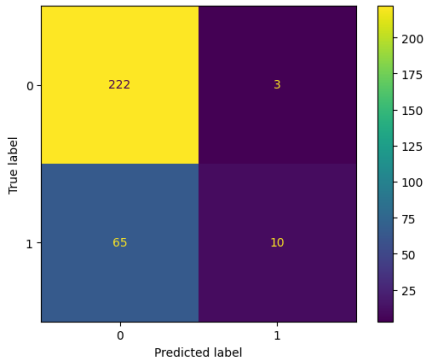
$$Confusion\ Matriks = \frac{A+D}{A+B+C+D} \text{ X } 100\%$$

$$Confusion\ Matriks = \frac{214+13}{214+11+62+13} \text{ X } 100\%$$

$$Confusion\ Matriks = 75{,}67\%$$

4.9. Evaluasi KNN berbasis PSO

Klasifikasi menggunakan metode KNN berbasis PSO, akurasi terbaik didapatkan pada *dataset* 70:30, sebagai berikut.



Gambar 6. Confusion matriks metode knn+pso

Nilai akurasi dihitung dengan persamaan berikut:

$$Confusion\ Matriks = \frac{A+D}{A+B+C+D} \text{ X } 100\%$$

$$Confusion\ Matriks = \frac{222+10}{222+3+65+10} \text{ X } 100\%$$

$$Confusion\ Matriks = 77{,}3\%$$

5. KESIMPULAN DAN SARAN

Dari hasil peneltian yang dilakukan, dapat disimpulkan bahwa pada penelitian ini, penerapan Optimasi metode *K-Nearest Neighbor* menggunakan algoritma *Particle Swarm Optimization* pada analisis sentimen pemilihan presiden Indonesia tahun 2024-2029 selalu berhasil meningkatkan akurasi awal yang dihasilkan oleh metode KNN. Model terbaik yang didapatkan dalam penelitian ini yaitu pada metode *K-Nearest Neighbor* yang dioptimasi menggunakan algoritma *Particle Swarm Optimization* dengan *rasio dataset* 70:30, mendapatkan akurasi sebesar 77,3% pada K=7 dan *best cost*=0,27 dengan pengujian menggunakan partikel 10 dan iterasi 20. Adapun saran dari penelitian ini sebagai berikut: Pada penelitian berikutnya diharapkan untuk memperluas penelitian ini dengan mencoba menggunakan algoritma lain yang dapat meningkatkan akurasi dalam melakukan analisis sentimen. Pada penelitian berikutnya diharapkan dapat melakukan optimasi pada seleksi fitur lainnya untuk meningkatkan hasil akurasi pada metode *K-Nearest Neighbor.*

DAFTAR PUSTAKA

[1] P. Mimboro, "Sentimen Twitter terhadap PILKADA kota Medan menggunakan metode Naive Bayes," *Jnanaloka*, vol. 3, no. 1, pp. 27–32, 2022, doi: 10.36802/jnanaloka.2022.v3-no1-27-32.

[2] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.

[3] N. Khotimah and D. Istiawan, "Perbandingan Algoritma C4. 5, Naïve Bayes dan K-Nearest Neighbour untuk Prediksi Lahan Kritis di Kabupaten Pemalang," in *Prosiding University Research Colloquium*, 2018, pp. 41–50.

[4] A. K. B. Ginting, M. S. Lydia, and E. M. Zamzami, "Peningkatan Akurasi Metode K-Nearest Neighbor dengan Seleksi Fitur Symmetrical Uncertainty," *J. Media Inform. Budidarma*, vol. 5, no. 4, pp. 1714–1719, 2021.

[5] S. Wahyuningsih and D. Retno Utari, "Perbandingan Metode K-Nearest Neighbor, Naïve Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit," *Konf. Nas. Sist. Inf. 2018*, pp. 8–9, 2018.

[6] A. Budianto, R. Ariyuana, and D. Maryono, "Perbandingan K-Nearest Neighbor (Knn) Dan Support Vector Machine (Svm) Dalam Pengenalan Karakter Plat Kendaraan Bermotor," *J. Ilm. Pendidik. Tek. dan Kejuru.*, vol. 11, no. 1, p. 27, 2019, doi: 10.20961/jiptek.v11i1.18018.

[7] W. Yunus, "Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik," *J. Tek. Elektro CosPhi*, vol. 2, no. 2, pp. 51–55, 2018, [Online]. Available: https://cosphijournal.unisan.ac.id/index.php/cosphihome/article/view/43%0Ahttps://cosphijournal.unisan.ac.id/index.php/cosphihome/article/download/43/20

[8] R. Arief and K. Imanuel, "ANALISIS SENTIMEN TOPIK VIRAL DESA PENARI PADA MEDIA SOSIAL TWITTER DENGAN METODE LEXICON BASED Universitas Gunadarma 1, 2 Jalan Margonda Raya No 100 Depok Jawa Barat 16424 Sur-el : rifiana@staff.gunadarma.ac.id 1 , karel4404@gmail.com 2," *J. Ilm. MATRIK*, vol. 21, no. 3, pp. 242–250, 2019, [Online]. Available: https://journal.binadarma.ac.id/index.php/jurnalmatrik/article/view/727

[9] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022.

[10] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 406–414, 2021.

[11] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *INSYST J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019.

[12] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.

[13] V. K. Sitanayah Que, A. Iriani, and H. Dwi Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization)," *J. Nas. Tek. Elektro dan Teknol. Inf.* , vol. 9, no. 2, pp. 162–170, 2020, [Online]. Available: www.tripadvisor.com,

[14] M. A. Imron, "Peningkatan Akurasi Algoritma K-Nearest Neighbor Menggunakan Normalisasi Z-Score Dan Particle Swarm Optimization Untuk Prediksi Customer Churn," 2020.

[15] S. R. I. Rezeki, "Penggunaan sosial media twitter dalam komunikasi organisasi (studi kasus pemerintah provinsi dki jakarta dalam penanganan covid-19)," *J. Islam. Law Stud.*, vol. 04, no. 02, pp. 63–78, 2020.

[16] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, 2018.

[17] D. Pajri, Y. Umaidah, and T. N. Padilah, "K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 242–253, 2020, doi: 10.28932/jutisi.v6i2.2658.