

# KLASIFIKASI ARTIKEL BERITA MENGGUNAKAN METODE K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION

Afdhalul Ihsan<sup>1</sup>, Ednawati Rainarli<sup>2</sup>

Teknik Informatika-Universitas Komputer Indonesia

Jalan Dipatiukur No.112-116 Bandung, 40312

Email : afdhalulhsan@email.unikom.ac.id,ednawati.rainarli@email.unikom.ac.id

## ABSTRAK

Penelitian ini akan melakukan pengklasifikasian berita dengan mengimplementasikan algoritma *K-Nearest Neighbor* berbasis *Particle swarm optimization*. Pada penelitian ini *Particle Swarm Optimization* digunakan sebagai metode untuk melakukan seleksi fitur. Penelitian ini menggunakan dataset sebanyak 250 dokumen berita, 25 dokumen sebagai data uji dan 225 dokumen sebagai data latih. Tahapan pada penelitian ini terbagi menjadi 2 yaitu proses pelatihan dan pengujian. Berdasarkan hasil dari penelitian menunjukkan bahwa penggunaan *Particle Swarm Optimization* sebagai seleksi fitur dapat memberikan hasil akurasi yang lebih baik dibandingkan dengan tanpa menggunakan *particle swarm optimization*. Hasil akurasi yang didapat setelah melakukan seleksi fitur menggunakan *particle swarm optimization* sebesar 80% dengan nilai  $K=9$ , banyak partikel yang dibangkitkan sebanyak 50, nilai  $c1$  dan  $c2 = 1$  dan total iterasi sebanyak 100. Pada penelitian ini metode *particle swarm optimization* belum memenuhi kondisi henti yang ditentukan sehingga hasil fitur yang dipilih belum optimal. Walaupun proses pada *particle swarm optimization* belum memenuhi kondisi henti, akan tetapi *particle swarm optimization* dapat meningkatkan akurasi pada proses pengklasifikasian sebesar 20% dibandingkan dengan tanpa menggunakan *particle swarm optimization* yang hanya mendapat akurasi tertinggi sebesar 60%.

**Kata Kunci** : *K-Nearest Neighbor*, Seleksi fitur, *Particle Swarm Optimization*, klasifikasi, Dataset

## 1. PENDAHULUAN

Berita merupakan informasi mengenai suatu kejadian yang memuat fakta didalamnya. Saat ini untuk mendapatkan suatu berita sudah sangat mudah, contohnya melalui portal berita online. Banyaknya artikel berita dan juga topik yang dibahas pada berita tersebut terkadang membuat sulitnya untuk mendapatkan suatu berita dengan informasi atau topik pembahasan yang diinginkan. Maka dari itu dibutuhkanlah suatu sistem yang dapat

mengidentifikasi artikel berita tersebut. Untuk melakukan identifikasi pada artikel berita dapat menggunakan metode klasifikasi.

Penelitian yang sudah membahas mengenai klasifikasi dokumen berita, seperti pada penelitian Andi Ahmad Irfa dan kawan-kawan [1] yang menggunakan metode *k-Nearest Neighbor* dalam melakukan klasifikasi dokumen artikel berita [1]. Pada penelitian oleh Andi Ahmad Irfa dan teman-temannya, didapatkan hasil pada penelitian yang dilakukan nilai  $f$ -measure sebesar 69,9% [1]. Penelitian mengenai klasifikasi dokumen menggunakan metode *k-Nearest Neighbor* juga pernah dilakukan oleh Claudio Fresta Suharno dan teman-temannya [2]. Pada penelitian yang dilakukan Claudio dijelaskan perbandingan klasifikasi dokumen menggunakan *k-Nearest Neighbor* dengan seleksi fitur dan tanpa seleksi fitur [2]. Pada penelitian Claudio menjelaskan bahwa penggunaan seleksi fitur dalam tahapan pengklasifikasian memberikan hasil akurasi sistem yang lebih baik dibandingkan dengan tanpa seleksi fitur.

Seleksi fitur merupakan tahap untuk memilih fitur yang paling penting dalam suatu data atau dokumen [2]. Tujuan dari seleksi fitur adalah untuk meningkatkan performa dari klasifikasi dokumen dengan menghilangkan fitur yang dianggap tidak relevan dalam klasifikasi untuk mengurangi dimensi dari himpunan fitur [2]. Seleksi fitur merupakan tahapan yang sangat penting dalam mengoptimalkan kinerja dari metode klasifikasi [3].

Metode untuk melakukan seleksi fitur saat ini sudah banyak dan beragam. Mehdi Hosseinzadeh Aghdam dan Setareh Heidari melakukan penelitian mengenai perbandingan 4 metode seleksi fitur, dimana pada penelitian ini membandingkan metode *Particle Swarm Optimization*, *Information Gain*, *Chi-Square*, dan *Genetic Algorithm*. Pada penelitian Mehdi *K-Nearest Neighbor* digunakan sebagai metode klasifikasi. Pada penelitian Mehdi didapatkan hasil bahwa penggunaan *Particle Swarm Optimization* sebagai seleksi fitur dapat memberikan hasil yang lebih baik dibandingkan metode lainnya yang digunakan [4]. Berdasarkan hal itu pada penelitian ini akan digunakan *Particle Swarm Optimization* dan *K-Nearest Neighbor* untuk

melakukan pengklasifikasian topik artikel dari suatu berita.

## 2. ISI PENELITIAN

### 2.1. Deskripsi Sistem

Dalam perancangan sistem yang akan dibangun merupakan sebuah sistem yang dapat melakukan pengklasifikasian topik artikel dari suatu berita. Sistem yang dibangun akan menerapkan metode *K-Nearest Neighbor* dan *Particle Swarm Optimization*. Terdapat dua tahap pada sistem yang akan dibangun, yaitu tahap pelatihan dan tahap pengujian. Tahap penelitian merupakan tahap pencarian fitur dengan menggunakan *Particle Swarm Optimization* dimana hasil dari tahap pelatihan ini merupakan fitur list kata yang dipilih berdasarkan pencarian *Particle Swarm Optimization*. Hasil dari tahap pelatihan akan digunakan pada tahap pengujian dimana pada tahap pengujian merupakan tahap pengklasifikasian artikel berita dengan menggunakan *Particle Swarm Optimization* dan *K-Nearest Neighbor* dan pada tahap pengujian juga akan mengukur akurasi yang dihasilkan oleh sistem.

### 2.2. Klasifikasi Dokumen

Klasifikasi dokumen merupakan salah satu bidang penelitian dalam perolehan informasi yang mengembangkan metode untuk menentukan atau mengkategorikan suatu dokumen ke dalam kelompok yang telah dikenal sebelumnya secara otomatis [5]. Klasifikasi dokumen bertujuan untuk mengkategorikan dokumen yang tidak terstruktur ke dalam kelompok-kelompok yang menggambarkan isi dari dokumen tersebut [5].

### 2.3. K-Nearest Neighbor

*K-Nearest Neighbor* merupakan salah satu metode yang dapat digunakan untuk melakukan pengklasifikasian. *K-Nearest Neighbor* merupakan metode *supervised learning* yang berarti adanya data latih dalam proses klasifikasi, dan tujuan dari metode ini adalah untuk memprediksi atau memetakan suatu data berdasarkan data latih yang sudah ada sebelumnya [6]. *K-Nearest Neighbor* merupakan teknik klasifikasi yang melakukan proses pengklasifikasian dengan cara menghitung jarak dari data yang akan di uji terhadap data latih [6]. Nilai dari jarak yang dihitung akan digunakan sebagai nilai kedekatan atau kemiripan antara data uji dan data latih. Pengukuran jarak kedekatan atau kemiripan pada metode *K-Nearest Neighbor* dihitung dengan menggunakan persamaan *euclidean distance* [6]. Rumus persamaan *euclidean distance* adalah sebagai berikut:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

Dimana  $x$  dan  $y$  merupakan titik pada ruang vektor  $n$  dimensi sedangkan  $x_i$  dan  $y_i$  merupakan besaran skalar untuk dimensi ke  $i$  dalam ruang vektor  $n$  dimensi.

### 2.4. Particle Swarm Optimization

PSO merupakan algoritma pencarian solusi berbasis populasi yang berdasar pada pola perilaku burung-burung, lebah atau sekumpulan ikan yang bergerak dalam kelompok [7]. *Particle Swarm Optimization* memiliki kelebihan yaitu mudah diterapkan, konvergensi yang cepat dan sederhana [8]. Kata partikel dalam *Particle Swarm Optimization* menunjukkan seekor burung dalam suatu kawanan, ketika salah satu burung dalam suatu kawanan menemukan jalan yang pendek terhadap sumber makanan maka sisa kelompok yang lain juga akan segera mengikuti jalan yang ditemukan burung tersebut meskipun jaraknya berjauhan [8]. Dalam *particle swarm optimization* setiap individu direpresentasikan dengan sebuah vektor dalam sebuah ruang pencarian solusi multidimensi [7].

### 2.5. Seleksi Fitur

Seleksi fitur merupakan proses mengurangi fitur dari  $m$  fitur menjadi sebanyak  $n$  fitur dengan cara menghilangkan fitur-fitur yang tidak informatif atau dengan cara memilih fitur yang paling informatif [9]. Beberapa metode dalam seleksi fitur adalah *filter based*, *wrapper based*, dan *hybrid*.

### 2.6. Term weighting Tf-Idf

*Term Weighting Tf-Idf* merupakan salah satu metode ekstraksi fitur yang banyak digunakan dalam proses klasifikasi dokumen [10]. Ekstraksi fitur merupakan tahapan untuk mengubah teks maupun dokumen kedalam suatu bentuk yang lebih representatif contohnya vektor [10]. Metode Tf-Idf menggabungkan 2 konsep dalam melakukan perhitungan bobot terhadap suatu *term*(kata), yaitu dengan menghitung frekuensi kemunculan suatu kata dalam satu dokumen dan *inverse* dari dokumen yang mengandung *term*(kata) tersebut.

Untuk menghitung bobot dari suatu *term*(kata) tersebut dapat menggunakan rumus:

$$W_i = T_{f_{ij}} * idf_i \quad (2.2)$$

*Term Frequency (Tf)* merupakan frekuensi kemunculan suatu *term* (kata) dalam suatu dokumen. Untuk menghitung *Tf* dapat menggunakan rumus:

$$Tf = \frac{n}{\sum n_k} \quad (2.3)$$

Dimana  $n$  merupakan banyaknya kemunculan *term* yang akan diberi bobot dalam suatu dokumen dan  $n_k$  merupakan banyak kata dalam dokumen. *Inverse Document Frequency (Idf)* merupakan pengurangan dominansi *term* (kata) dalam suatu dokumen. Perhitungan *Idf* ini diperlukan karena pada setiap dokumen *term* (kata) yang sering muncul dalam banyak dokumen akan dianggap sebagai *term* yang umum [10]. Untuk menghitung *Idf* dapat menggunakan rumus:

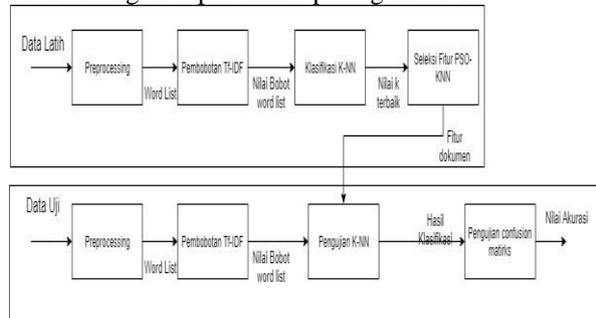
$$Idf = \log \frac{n}{df} \quad (2.4)$$

Dimana :

$Idf$  = inverse document frequency  
 $n$  = jumlah dokumen  
 $df$  = jumlah dokumen yang memiliki kata (*term*)

## 2.7. Analisis Sistem

Sistem yang akan dibangun memiliki 2 tahapan yaitu tahapan pelatihan dan pengujian dimana pada setiap tahapan tersebut terbagi menjadi beberapa tahapan. Gambaran umum dari sistem yang akan dibangun dapat dilihat pada gambar 1.



**Gambar 1** Gambaran Umum Sistem

1. Tahapan pertama yang dilakukan adalah proses pelatihan. Dalam proses pelatihan data latih akan melalui 4 proses utama yaitu proses preprocessing, pembobotan TF-IDF, klasifikasi K-NN, dan Seleksi fitur PSO-KNN. Proses preprocessing meliputi case folding, filtering, tokenizing, dan stopwords removal. Pada tahap pembobotan TF-IDF akan dilakukan pembobotan pada tiap kata yang ada pada dokumen. Pada tahap klasifikasi KNN akan dilakukan proses pengklasifikasian dengan metode k-nearest neighbor untuk mencari nilai k terbaik. Nilai k terbaik yang didapatkan akan digunakan sebagai parameter dalam seleksi fitur menggunakan PSO-KNN. Setelah mendapatkan nilai k terbaik akan dilakukan proses seleksi fitur untuk mencari fitur dari dokumen data latih yang akan digunakan sebagai fitur pada saat proses pengujian.
2. Tahapan berikutnya yang dilakukan adalah proses pengujian. Dalam proses pengujian akan melalui 4 proses utama yaitu proses pre-processing, pembobotan, klasifikasi KNN tahap pengujian, dan pengujian confusion matrix. Hasil dari tahap pengujian ini adalah dokumen uji yang sudah dikelompokkan menggunakan metode k-Nearest Neighbor. Setelah dilakukan pengujian klasifikasi k-Nearest Neighbor dilakukan pengujian sistem dengan menggunakan confusion matrix, hasil dari pengujian tersebut adalah akurasi dari metode k-Nearest Neighbor dan Particle Swarm Optimization.

## 2.8. Analisis Data Masukan

Data masukan yang akan digunakan dalam sistem yang akan dibangun merupakan data artikel berita yang memiliki format \*.txt. Banyak data yang akan digunakan sebanyak 250 dokumen 25 dokumen digunakan sebagai data latih dan 225 dokumen sebagai data uji. Pada data masukan akan

diklasifikasikan kedalam 5 kelas yaitu travel, techno, olahraga, kesehatan, otomotif.

## 2.9. Analisis Tahap Pelatihan

Analisis tahap pelatihan merupakan tahapan untuk mencari fitur yang informatif yang akan digunakan pada proses pengklasifikasian artikel berita. Pada tahap ini terdiri dari tahap *pre-processing*, pembobotan *Tf-Idf*, Klasifikasi KNN, dan seleksi fitur PSO-KNN.

### 2.9.1. Tahap Preprocessing

Tahap preprocessing merupakan tahap awal dalam melakukan proses klasifikasi. Tahap ini merupakan tahap yang penting dalam proses klasifikasi. Pada tahap pre-processing terdapat 4 tahap yaitu *case folding*, *filtering*, *tokenizing*, dan *stopword removal*.

#### a) Case folding

*Case folding* merupakan tahapan untuk menyeragamkan bentuk karakter dalam suatu dokumen. Penyeragam dapat dilakukan kedalam bentuk *lower case* maupun *upper case*. Contoh pada tahapan *case folding* dapat dilihat pada gambar 2.

Teks input:

Program Studi Teknik Informatik merupakan salah satu jurusan di UNIKOM

Hasil :

program studi teknik informatik merupakan salah satu jurusan di unikom

**Gambar 2** Case Folding

#### b) Filtering

Tahap *filtering* merupakan tahapan untuk menghilangkan noise yang ada pada dokumen. Noise disini dapat berupa tanda baca maupun angka. Pada penelitian terdapat aturan dalam proses *filtering* yaitu:

1. Mengganti karakter delimiter dengan spasi.
  2. Penghapusan karakter number pada tahap ini setiap karakter angka(0-9) akan di hapus.
  3. Penghapusan spasi akibat spasi yang berlebihan dari proses penggantian karakter delimiter yang diubah menjadi spasi, sehingga hanya akan menyisakan 1 spasi yang memisahkan tiap kata
- Contoh tahap *filtering* dapat dilihat pada gambar 3.

Teks Input:

nama saya afdhalul ihsan, dan umur saya 21 tahun

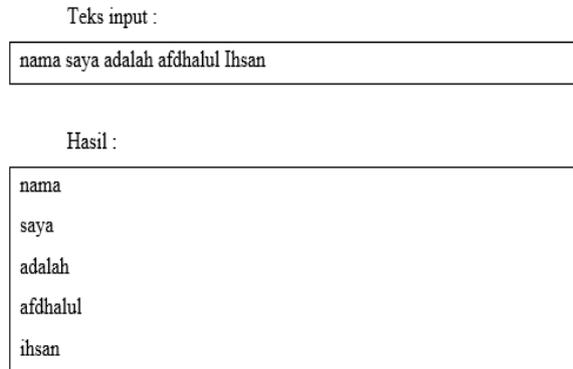
Hasil:

nama saya afdhalul ihsan umur saya tahun

**Gambar 3** Filtering

c) *Tokenizing*

*Tokenizing* merupakan tahapan memisahkan dokumen menjadi term (kata) dengan menggunakan pemisah spasi ( ). Contoh tahap *tokenizing* dapat dilihat pada gambar 4.



**Gambar 4** Tokenizing

d) *Stopword removal*

*Stopword removal* merupakan tahap menghilangkan kata – kata yang dianggap tidak relevan(penting) berdasarkan *list stopwords*. Pada penelitian ini *list stopwords* yang digunakan menggunakan *list stopwords* yang terdapat pada *library python Sastrawy*.

**2.9.2. Tahap Pembobotan Tf-Idf**

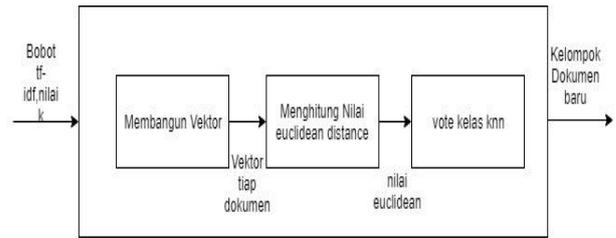
Tahap pembobotan Tf-idf digunakan untuk memberikan bobot nilai terhadap setiap *term* (kata) yang terdapat dalam kumpulan dokumen. Perhitungan bobot dilakukan berdasarkan perhitungan *Tf* dan *Idf*. Perhitungan *Tf* dilakukan dengan menggunakan persamaan 2.3. Perhitungan *Idf* dihitung menggunakan persamaan 2.4. Perhitungan bobot(*w*) dihitung dengan menggunakan persamaan 2.2. Adapaun tahapan pembobotan *Tf-Idf* dapat dilihat pada gambar 5.



**Gambar 5** Pembobotan Tf-Idf

**2.9.3. Tahap Klasifikasi KNN**

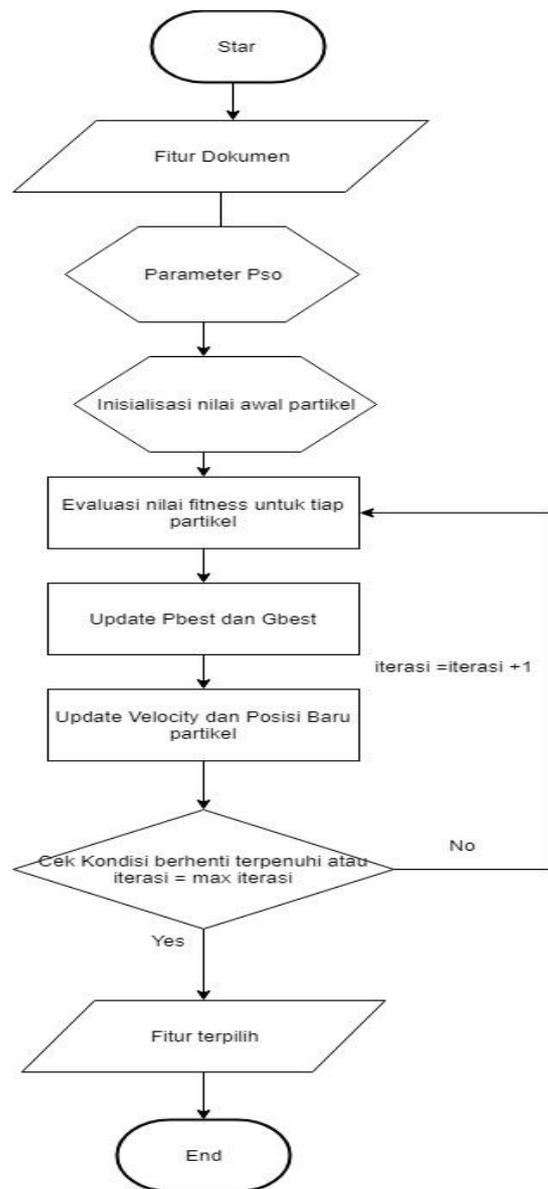
Tahap Klasifikasi KNN adalah tahap melakukan pengklasifikasian dokumen dengan menggunakan KNN dan semua fitur. Pada tahap ini bertujuan untuk mencari nilai *k* yang dapat memberikan nilai akurasi terbaik. Nilai *k* yang didapat akan digunakan sebagai parameter dalam seleksi fitur PSO-KNN. Tahapan Klasifikasi KNN dapat dilihat pada gambar 6.



**Gambar 6** Tahap Klasifikasi KNN

**2.9.4. Tahap Seleksi Fitur PSO-KNN**

Tahap seleksi PSO-KNN merupakan tahap mencari fitur yang dianggap relevan dalam proses klasifikasi. Pada penelitian ini seleksi PSO-KNN akan membuang fitur sebanyak 25% dan 50%. Adapun tahapan Seleksi fitur PSO-KNN dapat dilihat pada gambar 7.



**Gambar 7** Seleksi Fitur PSO KNN

Pada tahap seleksi fitur pso knn data masukan yang akan diproses merupakan list kata yang terdapat pada kumpulan dokumen. Seleksi fitur dengan menggunakan *particle swarm optimization* akan membuang sebanyak n kata berdasarkan hasil pencarian dalam ruang solusi *particle swarm optimization*. Adapun tahapan dalam pencarian solusi particle swarm optimization adalah:

a) Bangkitkan Parameter PSO

Pada tahap ini akan dibangkitkan nilai-nilai yang menjadi paramer dalam pencarian solusi menggunakan PSO. Parameter yang akan dibangkitkan meliputi nilai C1 dan C2, banyak partikel yang akan digunakan dalam tiap proses iterasi, kondisi berhenti yang akan digunakan dan banyak iterasi.

b) Inisialisasi nilai awal partikel

Pada tahap ini akan diinisialisasi nilai awal partikel secara random, dimana akan diinisialisai setiap nilai awal untuk tiap partikel yang dibangkitkan. Partikel disini merupakan representasi dari fitur pada dokumen sebanyak n dimana n merupakan banyaknya fitur yang terdapat pada dataset [9].

c) Evaluasi nilai fitness untuk tiap partikel

Pada tahap ini partikel-partikel yang telah diinisialisai nilainya akan dievaluasi dengan menggunakan fungsi *objective*. Dimana pada tahap evaluasi ini akan menguji apakah fitur yang telah dipilih oleh tiap partikel akan dievaluasi menggunakan fungsi *objective*. Untuk mengevaluasi tiap partikel dapat dihitung menggunakan rumus [9]:

$$Fitness(x_i) = ErrorRate \quad (2.5)$$

$$ErrorRate = \frac{FP+FN}{TP+TN+FP+FN} \quad (2.6)$$

d) Update Pbest dan Gbest

Pada tahap *Update pbest* dan *gbest* adalah mengupdate nilai *pbest* dan *gbest* dengan membandingkan *pbest* saat ini dan *pbest* pada iterasi sebelumnya. *Pbest* merupakan posisi terbaik tiap partikel pada proses iterasi, sedangkan *gbest* merupakan posisi *pbest* terbaik dalam tiap iterasi.

e) Update Velocity

Update *velocity* merupakan tahapan untuk menghitung *velocity* dari tiap nilai pada partikel. Nilai *Velocity* merupakan nilai yang akan digunakan untuk menghitung posisi partikel baru. Untuk menghitung *velocity* dapat menggunakan rumus:

$$V_{id}^{t+1} = V_{id}^t + c1 * r1 * (pbest_{id} - x_{id}^t) + c2 * r2 * (gbest_{id} - x_{id}^t) \quad (2.7)$$

f) Update Posisi Partikel

Tahap ini merupakan tahap untuk menghitung perpindahan dari tiap partikel. Untuk menghitung perpindahan tiap partikel dapat menggunakan rumus:

$$X_i(t+1) = V_i(t+1) + X_i(t) \quad (2.8)$$

g) Cek Kondisi Berhenti

Dalam pencarian solusi dengan menggunakan PSO iterasi akan berhenti ketika proses dalam iterasi sudah memenuhi kondisi henti atau sudah mencapai batas iterasi yang ditentukan [9]. Apabila proses dalam suatu iterasi sudah memenuhi kondisi henti maka hentikan proses proses, dan jika belum memenuhi kondisi henti ulangi proses mulai dari evaluasi partikel.

## 2.10. Analisis Tahap Pengujian

Tahap pengujian adalah tahap pengklasifikasian dan pengukuran akurasi sistem klasifikasi dengan menggunakan PSO-KNN. Pada tahap pengujian data uji dan data latih akan melewati tahapan yang sama seperti pada tahap pelatihan akan tetapi pada saat klasifikasi KNN vektor yang dibangun akan sebanyak fitur yang dipilih berdasarkan hasil dari seleksi fitur PSO. Pengujian dilakukan dengan menguji data dokumen berita yang belum diketahui kategori topiknya menggunakan metode k – Nearest Neighbor dan Particle Swarm Optimization . Data uji yang digunakan sebanyak 25 dokumen berita, sedangkan data latih yang digunakan adalah dokumen berita sebanyak 225 dokumen. Pada penelitian ini akan dilakukan beberapa cara dalam menguji performansi proses klasifikasi dari sistem yang dibangun, skenario pengujian yang akan dilakukan adalah sebagai berikut:

1. Pengujian metode k – Nearest Neighbor dengan nilai k yang akan digunakan adalah 1 sampai 10
2. Pengujian metode seleksi fitur, pada tahap ini dilakukan proses seleksi fitur dimana fitur yang akan dibuang sebanyak 25% dan 50% Jumlah partikel yang dibangkitkan dalam proses pencarian seleksi fitur menggunakan Particle Swarm Optimization adalah 15, 30, 50.
3. Pengujian klasifikasi dengan metode k – Nearest Neighbor dengan seleksi fitur Particle Swarm Optimization, pada pengujian ini semua fitur yang telah terpilih berdasarkan pengujian seleksi fitur akan diuji dengan menggunakan k dari 1 sampai 10.

a) Pengujian KNN

Pada pengujian KNN akan diukur akurasi yang dihasilkan sistem dalam melakukan klasifikasi dokumen tanpa seleksi fitur dengan menggunakan nilai k dari 1-10. Hasil dari akurasi yang didapat untuk tiap nilai k dapat dilihat pada tabel berikut:

**Tabel 1** Hasil Pengujian KNN

Nilai K	Accuracy
1	60%
2	60%
3	60%
4	44%

Nilai K	Accuracy
5	52%
6	56%
7	56%
8	56%
9	60%
10	60%

Dari tabel 1 memperlihatkan hasil pengujian dengan metode *k-Nearest Neighbor* dan menghasilkan akurasi terbesar sebesar 60% pada  $k = 1, k=2, k=3, k=9$  dan  $k=10$ . Hal ini menunjukkan bahwa penggunaan nilai  $k$  yang digunakan mempengaruhi tingkat akurasi yang dihasilkan, semakin besar nilai  $k$  jika sudah menemukan hasil yang optimum cenderung menurun jika menggunakan  $k$  yang semakin besar [1]. Pada tabel diatas sudah mendapatkan hasil akurasi yang optimum pada  $k=1,2$ , dan  $3$ , dan semakin besar  $k$  yang digunakan akurasi cenderung menurun dikarenakan semakin bertambahnya nilai  $k$  semakin bertambah juga data yang tidak memiliki kesamaan terhadap data uji [1].

b) Pengujian Seleksi Fitur PSO

Pada pengujian seleksi fitur akan diuji sistem dalam melakukan proses seleksi fitur dan mengukur nilai fitness yang didapatkan dari tiap proses pengujian. Hasil nilai fitness yang didapat dari tiap pengujian dapat dilihat pada tabel 2.

**Tabel 2** Nilai fitness Pengujian PSO

Partikel	Fitness	
	25% dibuang	50% dibuang
15	0,64	0,68
30	0,76	0,72
50	0,76	0,76

Berdasarkan tabel 2 dapat dilihat pengaruh jumlah partikel yang dibangkitkan mempengaruhi hasil nilai fitness dari optimasi melakukan pencarian fitur yang dipilih. Semakin banyak partikel yang dibangkitkan maka dapat menghasilkan fitur terpilih yang lebih optimal, dan nilai fitness yang lebih besar. Banyaknya partikel yang terlibat dalam proses pencarian fitur menggunakan *particle swarm optimization* memberikan banyak pilihan dalam tiap iterasi yang dilakukan dan dari banyak pilihan tersebut kemudian akan dipilih partikel yang memberikan nilai *fitness* terbaik [11].

c) Pengujian PSO KNN

Tahap pengujian PSO-KNN merupakan tahap untuk mengukur akurasi dari fitur yang dipilih berdasarkan seleksi fitur PSO. Pada pengujian PSO-KNN nilai  $K$  yang digunakan adalah 1-10. Hasil dari

tiap pengujian PSO-KNN dapat dilihat pada tabel berikut:

**Tabel 3** Pengujian PSO-KNN 15 Partikel

Nilai K	15 Partikel	
	25% dibuang	50% dibuang
	$f=0,64$	$f=0,68$
1	60%	56%
2	60%	56%
3	60%	56%
4	48%	60%
5	56%	56%
6	68%	56%
7	60%	56%
8	56%	60%
9	56%	56%
10	64%	68%

**Tabel 4** Pengujian PSO-KNN 30 Partikel

Nilai K	30 Partikel	
	25% dibuang	50% dibuang
	$f=0,76$	$f=0,72$
1	56%	40%
2	56%	40%
3	32%	40%
4	40%	52%
5	44%	64%
6	52%	68%
7	60%	56%
8	76%	60%
9	60%	72%
10	76%	72%

**Tabel 5** Pengujian PSO-KNN 50 Partikel

Nilai K	50 Partikel	
	25% dibuang	50% dibuang
	$f=0,76$	$f=0,76$
1	52%	56%
2	52%	56%
3	44%	56%
4	48%	56%

Nilai K	50 Partikel	
	25% dibuang	50% dibuang
	$f=0,76$	$f=0,76$
5	52%	52%
6	56%	52%
7	68%	56%
8	72%	56%
9	80%	80%
10	76%	76%

Berdasarkan tabel 3-5 dapat dilihat semakin tinggi nilai fitness yang dihasilkan dari pemilihan fitur *particle swarm optimization* maka dapat menghasilkan nilai akurasi tertinggi yang lebih baik. Pada tabel 4.9 akurasi tertinggi diperoleh pada pengujian PSO-KNN dengan menggunakan  $K=9$  yaitu sebesar 80% dimana hasil ini diperoleh dari fitur yang dibuang sebanyak 25% dan 50% dan banyak partikel dibangkitkan sebanyak 50 dengan nilai fitness yang didapat sebesar 0,76. Pada pengujian partikel yang dibangkitkan sebanyak 15 akurasi hanya mencapai 68% untuk dengan penggunaan  $K=6$  untuk fitur yang dibuang sebanyak 25% dan  $K=10$  untuk fitur yang dibuang sebanyak 50% dengan nilai fitness 0,64 dan 0,68. Pada pengujian partikel yang dibangkitkan sebanyak 30 akurasi tertinggi diperoleh pada penggunaan  $K=10$  dan  $K=9$  untuk 25% fitur dibuang dengan menghasilkan akurasi sebesar 76% dengan nilai fitness sebesar 0,76, sedangkan pada pengujian 50% fitur dibuang terjadi penurunan tingkat akurasi dimana akurasi tertinggi yang diperoleh sebesar 72% untuk penggunaan  $K=9$  dan  $K=10$  dengan nilai fitness 0,72.

### 3. PENUTUP

#### 3.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa penerapan metode K-Nearest Neighbor berbasis *particle swarm optimization* sebagai seleksi fitur memiliki akurasi yang lebih baik dibandingkan dengan hanya menggunakan metode K-Nearest neighbor. Akan tetapi fitur yang dihasilkan dari proses seleksi fitur belum optimal dikarenakan *particle swarm optimization* belum memenuhi kondisi henti yang ditentukan. Pengujian pengklasifikasian dokumen berita dengan menggunakan 250 dokumen, dimana 225 dokumen digunakan sebagai data latih dan 25 dokumen sebagai data uji, menghasilkan akurasi tertinggi sebesar 80% dengan menggunakan  $K=9$  dan banyak partikel yang dibangkitkan sebanyak 50. Pada pengujian 15 dan 30 partikel yang dibangkitkan mendapat nilai akurasi tertinggi sebesar 68% untuk 15 partikel yang dibangkitkan dengan  $K=6$ , dan 76% untuk 30 partikel dibangkitkan dengan menggunakan  $K=10$  dan  $K=8$ . Pada pengujian klasifikasi tanpa

seleksi fitur dengan metode *particle swarm optimization* didapatkan akurasi tertinggi sebesar 60%.

#### 3.2. Saran

Dalam penelitian yang dikerjakan ini masih terdapat kekurangan yang terjadi, dan penelitian masih dapat dikembangkan lagi agar dapat mendapatkan hasil yang lebih baik. Adapun saran yang dapat diberikan untuk penelitian selanjutnya ialah menggunakan metode pengklasifikasian lainnya sebagai algoritma pembelajaran yang memiliki tingkat kinerja lebih baik dibandingkan dengan metode *K-Nearest Neighbor*.

### DAFTAR PUSTAKA

- [1] A. A. Irfa, Adiwijaya And M. S. Mubarak, "Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan K-Nearest Neighbor," *E - Proceeding Of Engineering*, Vol. 5, Pp. 3631-3640, 2018.
- [2] C. F. Suharno, M. Fauzi And R. S. Perdana, "Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-Square," *Systemic*, Vol. 3, Pp. 25-32, 2017.
- [3] V. Chandani And P. Romi Satria Wahono, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection Pada Analisis Sentimen Review Film," *Journal Of Intelligent Systems*, Vol. 1, Pp. 56 - 60, 2015.
- [4] M. H. Aghdam And S. Heidari, "Feature Selection Using Particle Swarm Optimization In Text Categorization," *Jaiscr*, Vol. 5, Pp. 231-238, 2015.
- [5] H. Februariyanti And E. Zuliarso, "Klasifikasi Dokumen Berita Teks Bahasa Indonesia," *Jurnal Teknologi Informasi Dinamik*, Vol. 17, Pp. 14-23, 2012.
- [6] A. H. Ferdinan, A. B. Osmond And C. Setianingsih, "Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode K-Nearest Neighbor," *E-Proceeding Of Engineering*, Vol. 5, Pp. 6187-6194, 2018.
- [7] G. Hermawan, "Implementasi Algoritma Particle Swarm Optimization Untuk Penentuan Posisi Strategis Agent Pada Simulasi Robot Sepakbola Dua Dimensi," *Jurnal Ilmiah Komputer Dan Informatika (Komputa)*, Vol. 1, Pp. 63-70, 2012.
- [8] K. W. Mahardika, Y. A. Sari And A. Arwan, "Optimasi K-Nearest Neighbor Menggunakan Particle Swarm Optimization Pada Sistem Pakar Untuk Monitoring Pengendalian Hama Pada Tanaman Jeruk," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, Vol. 2, Pp. 3333-3344, 2018.

- [9] B. Xue, *Particle Swarm Optimisation For Feature Selection*, Victoria University, 2014.
- [10] D. W. Suliantoro, I. Wisnubhadra And Ernawati, "Integrasi Pembobotan Tf-Idf Pada Metode K-Means Untuk Clustering Dokumen Teks," *Prosiding Seminar Nasional Manajemen Teknologi*, 2012.
- [11] D. Ariani, A. Fahriza And I. Prasetyaningrum, "Optimasi Penjadwalan Mata Kuliah Di Jurusan Teknik Informatika Pens Dengan Menggunakan Algoritma Particle Swarm Optimization".