

## Optimasi Algoritma SVM Dan K-NN Berbasis Particle Swarm Optimization Pada Analisis Sentimen Fenomena Tagar #2019GantiPresiden

Atang Saepudin<sup>1</sup>, Riska Aryanti<sup>2</sup>, Eka Fitriani<sup>3</sup>, Dahlia<sup>4</sup>

<sup>1,2</sup>Program Studi Ilmu Komputer  
Universitas Bina Sarana Informatika  
<sup>1</sup>e-mail: atang.aug@bsi.ac.id  
<sup>2</sup>e-mail: riska.rts@bsi.ac.id

<sup>3</sup>Program Studi Sistem Informasi  
Universitas Bina Sarana Informatika  
<sup>3</sup>e-mail: eka.ean@bsi.ac.id

<sup>4</sup>Program Studi Sistem Informasi Akuntansi  
Universitas Bina Sarana Informatika  
<sup>4</sup>e-mail: dahlia.dlx@bsi.ac.id

Diterima	Direvisi	Disetujui
06-11-2019	06-12-2019	01-02-2020

**Abstrak** - Analisis sentimen adalah proses untuk menentukan konten dataset berbasis teks yang positif atau negatif. Saat ini, opini publik menjadi sumber penting dalam keputusan seseorang dalam menemukan solusi. Algoritma klasifikasi seperti *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (k-NN) diusulkan oleh banyak peneliti untuk digunakan dalam analisis sentimen untuk pendapat ulasan. Namun, klasifikasi sentimen teks memiliki masalah pada banyak atribut yang digunakan dalam dataset. Fitur pemilihan dapat digunakan sebagai proses optimasi untuk mengurangi set fitur asli ke subset yang relatif kecil dari fitur yang secara signifikan meningkatkan akurasi klasifikasi untuk cepat dan efektif. Masalah dalam penelitian ini adalah pemilihan fitur untuk meningkatkan nilai akurasi *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (k-NN) dan membandingkan akurasi tertinggi untuk analisis sentimen tweet / komentar yang menggunakan tagar # 2019GantiPresiden. Algoritma perbandingan, SVM menghasilkan akurasi 88,00% dan AUC 0,964, kemudian dibandingkan dengan SVM berdasarkan PSO dengan akurasi 92,75% dan AUC 0,973. Data hasil pengujian untuk akurasi algoritma k-NN adalah 88,50% dan AUC 0,948, kemudian dibandingkan untuk akurasi dengan PSO berbasis k-NN sebesar 75,25% dan AUC 0,768. Hasil pengujian algoritma PSO dapat meningkatkan akurasi SVM, tetapi tidak mampu meningkatkan akurasi algoritma k-NN. Algoritma SVM berbasis PSO terbukti memberikan solusi untuk masalah klasifikasi tweets/ komentar yang menggunakan tagar # 2019GantiPresiden di Twitter agar lebih akurat dan optimal.

Kata Kunci: optimasi analisa sentimen

**Abstract** - *Sentiment analysis is a process to determine the content of text-based datasets which are positive or negative. At present, public opinion be an important resource in the decision of a person in finding a solution. Klasifikasi algorithms such as Support Vector Machine (SVM) and K-Nearest Neighbor (k-NN) is proposed by many researchers to be used in sentiment analysis for review opinion. However, the text sentiment klasifikasi has problems on many of the attributes used in a dataset. Selection feature can be used as process optimization to reduce the feature set of the original to a relatively small subset of the features that significantly improve the accuracy of klasifikasi for fast and effective. The problem in this research is the selection of feature selection to improve accuracy values Support Vector Machine (SVM) and K-Nearest Neighbor (k-NN) and compare the highest accuracy for sentiment analysis tweet/comments that use hashtag #2019GantiPresiden. The comparison algorithms, SVM produces an accuracy of 88.00% and AUC 0.964, then compared with SVM based on PSO with an accuracy of 92.75% and AUC 0.973. The test result data for k-NN algorithm accuracy was 88.50% and AUC 0.948, then compared for accuracy by k-NN based PSO amounted to 75.25% and AUC 0.768. The results of the testing of the PSO algorithm can improve the accuracy of SVM, but are not able to improve the accuracy of the algorithm k-NN. SVM algorithm based on PSO proven to provide solutions to the problems of klasifikasi tweets / comments that use hashtag #2019GantiPresiden on Twitter to be more accurately and optimally.*

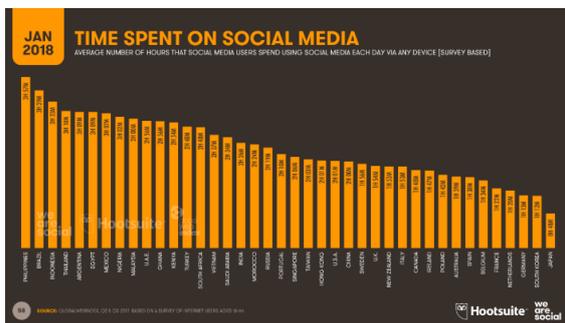
Keywords: Optimization of sentiment analysis

## PENDAHULUAN

Indonesia adalah salah satu negara yang menganut sistem demokrasi. Hal ini dapat dilihat dengan diadakannya suatu Pemilihan Umum (Pemilu) terhadap Presiden dan Wakil Presiden. Pemilu pada suatu negara yang menganut demokrasi biasanya diselenggarakan secara periodik. Di Indonesia sendiri pemilihan Presiden dan Wakil Presiden dilakukan setiap lima tahun sekali. Pada tahun 2019, Indonesia telah menyelenggarakan Pemilu untuk memilih Presiden dan Wakil Presiden untuk periode 2019-2024. Seorang tokoh politik yang ingin maju sebagai Calon Presiden (Capres) tentu akan melihat atau mempertimbangkan popularitas mereka berdasarkan opini dari masyarakat.

Dahulu masyarakat mengungkapkan opini, kritik, dan sarannya melalui media cetak yang tidak semua orang mempunyai kemampuan menulis dan kesempatan menerbitkan tulisannya. Namun, perkembangan teknologi komunikasi saat ini telah merubah kecenderungan kebiasaan masyarakat dalam mengekspresikan opininya pada jejaring sosial. Salah satu jejaring sosial yang populer di kalangan pengguna internet saat ini adalah Twitter (Nurhuda dkk, 2014).

Menurut penelitian yang dilakukan *We Are Social*, perusahaan media asal Inggris yang bekerja sama dengan *Hootsuite*, rata-rata orang Indonesia menghabiskan 3 jam 23 menit sehari untuk mengakses media sosial. Dari laporan berjudul "*Essential Insights Into Internet, Social Media, Mobile, and E-Commerce Use Around The World*" yang diterbitkan tanggal 30 Januari 2018, dari total populasi Indonesia sebanyak 265,4 juta jiwa, pengguna aktif media sosialnya mencapai 130 juta dengan penetrasi 49 persen (*We Are Social*, 2018).



Sumber: *We are social* (2018)

Gambar 1. Persentase Rata-Rata Waktu Pengguna Sosial Media Di Berbagai Negara  
Media sosial khususnya Twitter merupakan salah satu media promosi atau kampanye yang cukup efektif bagi calon presiden dan wakil presiden yang maju pada Pemilu 2019. Menjelang pemilu 2019, masyarakat diramaikan dengan perang tagar antara kubu pendukung Prabowo dan Jokowi di sosial

media. Tagar #2019GantiPresiden yang paling menyita perhatian dan ramai jadi bahan perbincangan publik di sosial media khususnya Twitter. Tagar tersebut sengaja dibuat oleh pendukung Prabowo untuk mendongkrak elektabilitas Prabowo melalui media sosial dalam menghadapi lawan politiknya yaitu Jokowi di Pilpres 2019.

Hashtag sendiri dipopulerkan oleh Twitter, yang mempunyai peran besar dalam meramaikan percakapan dan menjadi trending topik di media sosial. Tagar digunakan untuk mengindeks kata kunci atau topik di Twitter, dan memungkinkan pengguna untuk mengikuti topik yang diminati dengan mudah.

Fenomena tagar-tagar tersebut yang terus ramai diperbincangkan di media sosial khususnya Twitter, tentunya mengundang pro dan kontra dari masyarakat. Antara lain mereka yang setuju dan mendukung hastag #2019GantiPresiden dengan pengguna yang menolak hastag tersebut. Sehingga menimbulkan banyak opini dari pengguna yang tidak hanya bernilai positif, tetapi juga tidak sedikit publik mengungkapkan opini negatif dengan menggunakan kata-kata yang dapat menimbulkan permusuhan, penghinaan, perdebatan dan perselisihan.

Meluasnya penggunaan internet telah meningkatkan jumlah informasi yang disimpan dan diakses melalui web dalam kecepatan yang sangat cepat, karena banyaknya data yang terdapat di internet tersebut, tanpa diolah untuk dimanfaatkan lebih dalam maka munculah Opinion Mining yang merupakan cabang penelitian dari Text Mining. Fokus dari penelitian Opinion Mining adalah melakukan analisis opini dari suatu dokumen teks (Rozi, dkk, 2012).

Berdasarkan penjelasan di atas, untuk mengatasi permasalahan tersebut perlu dilakukan analisis yang disebut dengan sentiment analysis. Analisis sentimen secara umum dapat didefinisikan sebagai studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik. Topik tersebut kemungkinan besar dapat berupa review. Sentiment analysis digunakan untuk mengotomatisasi proses identifikasi pendapat apakah itu adalah pandangan positif atau negatif (Aryanti dkk., 2019).

K-Nearest Neighbor adalah metode sederhana namun efektif untuk teks kategorisasi tetapi memiliki beberapa kelemahan yaitu kompleksitas pada sample yang komputasi kesamaan besar, performance k-NN mudah dipengaruhi oleh sample tunggal, seperti noisy sample dan k-NN tidak membangun model klasifikasi karena termasuk ke

dalam lazy learning method. k-NN merupakan salah satu metode yang banyak digunakan untuk klasifikasi pada teknik data mining (Mentari dkk, 2018).

Support Vector Machine (SVM) merupakan metode supervised learning yang menganalisa data dan mengenali pola-pola yang digunakan untuk klasifikasi. SVM memiliki kelebihan yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda (Chou dkk, 2014). Namun SVM memiliki kekurangan terhadap masalah pemilihan parameter atau fitur yang sesuai (Basari dkk, 2013). Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi hal tersebut dikemukakan oleh Zhao dkk., 2011 dalam penelitian yang dilakukan sebelumnya (Utami, 2017).

Salah satu algoritma optimasi yang cukup populer adalah Particle Swarm Optimization (PSO) (Ibrahim dkk., 2015). Dalam teknik PSO terdapat beberapa cara untuk melakukan pengoptimasian diantaranya meningkatkan bobot atribut (attribute weight) terhadap semua atribut atau variabel yang dipakai, menseleksi atribut (attribute selection) dan feature selection. PSO adalah suatu teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter (Basari dkk., 2013).

Pada penelitian ini algoritma Particle Swarm Optimization digunakan sebagai seleksi fitur untuk tweet/komentar di sosial media khususnya twitter yang menggunakan tagar #2019GantiPresiden. PSO digunakan untuk meningkatkan akurasi dari kedua algoritma yang digunakan yaitu Support Vector Machine (SVM) dan k-Nearest Neighbor (k-NN).

## METODOLOGI PENELITIAN

Adapun metodologi penelitian yang penulis gunakan melalui beberapa tahapan sebagai berikut:

1. Pengumpulan Data  
Data yang digunakan untuk melakukan eksperimen dikumpulkan melalui tweet dan komentar di media sosial Twitter yang menggunakan tagar #2019GantiPresiden. Kemudian data tweet dan komentar tersebut diseleksi serta disimpan kedalam file *Microsoft Excel* untuk diolah dalam pengujian data selanjutnya.
2. Pengolahan Data Awal  
Tahap pengolahan data awal (*preprocessing*) dilakukan sebelum metode atau algoritma yang dipilih diterapkan. Proses pengolahan data awal yang digunakan yaitu *stopword removal, stemming, tokenize, @anotation removal, filter*

*by length, transform case* dan *n-grams*. Kemudian memilih metode yang akan digunakan pada saat pengujian data. Metode yang dipilih, berdasarkan penelitian yang terdahulu. Penulis menggunakan Metode Algoritma *Support Vector Machine* dan *k-Nearest Neighbor*.

3. Metode yang Diusulkan  
Metode yang diusulkan penulis tambahkan optimasi agar dapat meningkatkan nilai akurasi. Optimasi yang digunakan yaitu *Particle Swarm Optimization*. *Support Vector Machine* berbasis *Particle Swarm Optimization* dengan *k-Nearest Neighbor* berbasis *Particle Swarm Optimization*.
4. Eksperimen dan Pengujian Metode  
Eksperimen yang dilakukan peneliti menggunakan *framework RapidMiner Studio Versi 8.2* untuk mengolah data sehingga menghasilkan nilai akurasi yang akurat. Untuk implementasi atau pengujian metode penulis membuat aplikasi sederhana menggunakan bahasa pemrograman PHP dan HTML.
5. Evaluasi dan Validasi Hasil Evaluasi  
Evaluasi berfungsi untuk mengetahui akurasi dari model algoritma yang diusulkan. Validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah *Cross Validation*. Akurasi algoritma akan diukur menggunakan *Confusion Matrix* dan hasil perhitungan akan ditampilkan dalam bentuk *Curve ROC (Receiver Operating Characteristic)*.

## HASIL DAN PEMBAHASAN

Data training yang digunakan pada saat pengujian data diambil dari media sosial twitter. Pengujian data dilakukan dengan menggunakan tweet atau komentar yang menggunakan tagar #2019GantiPresiden (400 data training, yang terdiri dari 200 review negatif dan 200 review positif) kemudian dilakukan testing dan training dataset sehingga didapatkan *accuracy* dan *AUC (Area Under Curve)*.

Berikut merupakan tahapan-tahapan dalam melakukan pengolahan data yaitu:

### 1. Pengumpulan Data

Pengambilan data dibatasi berdasarkan tweet-tweet yang menggunakan tagar #2019GantiPresiden menggunakan Rapidminer Studio 8.2. Data tweet yang diperoleh kemudian disimpan kedalam *Microsoft Excel* untuk selanjutnya dilakukan proses labeling yaitu mengklasifikasikan tweet-tweet tersebut kedalam kelas negatif atau positif secara manual.

## 2. Pengolahan Data Awal (*Preprocessing*)

### a. *Stopwords Removal*

Dalam proses ini, kata-kata sambung atau kata-kata yang tidak relevan akan dihapus, seperti kata tetapi, untuk, dengan, yang, pada dan kata sambung lainnya.

Tabel 1. Proses *Stopwords Removal*

<i>Sebelum Stopwords Removal</i>	<i>Setelah Stopwords Removal</i>
Bener apa yang dibilang prof rocky gerung, sebenarnya yang potensial menyebar hoax adalah cebong.	Bener dibilang prof rocky gerung, sebenarnya potensial menyebar hoax cebong.

Sumber: Hasil Penelitian (2019)

### b. *@anotation Removal*

Proses ini bertujuan untuk menghapus atau menghilangkan link nama pengguna twitter (*username*) yang biasanya ditandai dengan simbol “@” yang terdapat pada dataset. Pada proses ini juga penulis masih menggunakan fasilitas yang disediakan oleh *gataframework.com*.

Tabel 2. Proses *@anotation Removal*

<i>Sebelum @anotation Removal</i>	<i>Setelah @anotation Removal</i>
@asae_04 alhamdulillah proses pencoblosan di wilayah karawang berjalan dengan aman @gerindra	alhamdulillah proses pencoblosan di wilayah karawang berjalan dengan aman

Sumber: Hasil Penelitian (2019)

### c. *Stemming*

Kata yang memiliki imbuhan awal dan akhiran seperti me, mem, meny, meng, di, per, ber, an, kan, i, nya dll., akan diubah menjadi kata dasar dengan menghilangkan kata imbuhan tersebut. Melalui *gataframework.com*, karena dataset yang digunakan adalah Bahasa Indonesia.

Tabel 3. Proses *Stemming*

<i>Sebelum Stemming</i>	<i>Setelah Stemming</i>
Bener apa yang dibilang prof rocky gerung, sebenarnya yang potensial menyebar hoax tu cebong.	Bener apa yang bilang prof rocky gerung, benar yang potensial sebar hoax adalah cebong.

Sumber: Hasil Penelitian (2019)

### d. *Tokenize*

Dalam proses tokenize ini, semua kata yang ada di dalam tiap dokumen dikumpulkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol, karakter khusus atau apapun yang bukan huruf.

Tabel 4. Proses *Tokenize*

<i>Sebelum Tokenize</i>	<i>Setelah Tokenize</i>
Bener apa yang dibilang prof rocky gerung, sebenarnya yang potensial	Bener apa yang bilang prof rocky gerung sebenarnya yang

menyebarkan hoax adalah cebond...!!!	potensial menyebarkan hoax adalah cebong
---	---

Sumber: Hasil Penelitian (2019)

### e. *Filter Tokens (By Length)*

Dalam proses ini, kata-kata yang memiliki panjang kurang dari 4 dan lebih dari 25 akan dihapus, seperti kata yg, tdk, jd, ga, ane, gan yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentiment.

Tabel 5. Proses *Filter Tokens (By Length)*

<i>Sebelum Filter Tokens</i>	<i>Setelah Filter Tokens</i>
Saya kok senang ya melihat lembaga survei pd ngaco, setidaknya kita bs lebih hati-hati dlm melihat hasil survei	Saya senang melihat lembaga survei ngaco, setidaknya kita lebih hati-hati melihat hasil survei

Sumber: Hasil Penelitian (2019)

### f. *Transform Cases*

Dalam proses ini, kata-kata yang tidak relevan akan diubah, seperti kata yang mengandung huruf besar yang diubah menjadi huruf kecil sehingga dapat saling berhubungan dengan sentiment.

Tabel 6. Proses *Transform Cases*

<i>Sebelum Transform Case</i>	<i>Setelah Transform Case</i>
Utk mengganti presiden kita harus TENGSELAMKAN BANTENG di setiap Daerah, GANTIPRESIDEN	utk mengganti presiden kita harus tenggelamkan banteng di setiap daerah, gantipresiden

Sumber: Hasil Penelitian (2019)

### g. *Generate n-Grams (bigram)*

Proses n-Grams digunakan untuk menyelesaikan permasalahan klasifikasi dokumen ke dalam sentiment positif atau negatif. Kesalahan klasifikasi sentiment biasanya disebabkan oleh fitur term tunggal. Misalnya kata “buruk” merupakan kata yang termasuk sentiment negatif, akan tetapi kata “tidak buruk” masuk ke dalam sentiment positif apabila berdampingan dengan kata negasi begitupun sebaliknya.

## 3. Klasifikasi

*Text mining* dapat menganalisis dokumen, mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya, serta menentukan kesamaan di antara dokumen untuk mengetahui bagaimana mereka berhubungan dengan variabel lainnya. *Text mining* merupakan teknik yang digunakan untuk menangani masalah *classification*, *clustering*, *information extraction* dan *information retrieval* (Berry & Kogan, 2013).

*Validation* adalah proses untuk mengevaluasi keakuratan prediksi dari model. Validasi digunakan untuk memperoleh prediksi menggunakan model

yang ada dan kemudian membandingkan hasil tersebut dengan hasil yang sudah diketahui, ini mewakili langkah paling penting dalam proses membangun sebuah model (Utami, 2017).

Pada penelitian ini proses menentukan bagian kalimat sebagai anggota kelas positif atau kelas negatif berdasarkan nilai perhitungan *Support Vector Machine* dan *K-Nearest Neighbor*. Jika hasilnya kalimat tersebut untuk kelas positif lebih besar dari pada kelas negatif, maka kalimat tersebut termasuk ke dalam kelas positif begitu sebaliknya.

### 1. Algoritma Support Vector Machine (SVM)

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C, epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycles*.

Tabel 7. Training Cycles SVM

C	Epsilon	Accuracy	AUC
<b>0.0</b>	<b>0.0</b>	<b>88.00 %</b>	<b>0.964</b>
0.1	0.1	86.75 %	0.966
0.2	0.2	87.00 %	0.967
0.3	0.3	86.50 %	0.967
0.4	0.4	86.75 %	0.967
0.5	0.5	86.75 %	0.967
0.6	0.6	86.75 %	0.967
0.7	0.7	86.75 %	0.966
0.8	0.8	86.75 %	0.964
0.9	0.9	86.75 %	0.967
0.0	1.0	50.00 %	0.500
1.0	1.0	50.00 %	0.500
1.0	0.0	86.75 %	0.968

Sumber: Hasil Penelitian (2019)

Hasil pengujian menunjukkan bahwa penerapan metode *Support Vector Machine* pada Tabel 7 dengan C = 0.0 dan Epsilon E = 0.0 dihasilkan Accuracy= 88.00% dan AUC= 0.964.

Tabel 8. Training Cycles SVM

accuracy: 88.00 % +/- 4.72 %			
	true negative	true positive	class precision
pred.negative	170	18	90.43 %
pred.positive	30	182	85.85 %
class recall	85.00 %	91.00 %	

Sumber: Hasil Penelitian (2019)



Sumber: Hasil Penelitian (2019)

Gambar 2. Kurva ROC SVM

Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan tweet/ komentar yang menggunakan tagar #2019GantiPresiden menggunakan metode SVM pada *framework RapidMiner*.

### 2. Algoritma SVM Berbasis PSO

Hasil *training cycles* menggunakan metode SVM berbasis PSO dapat dilihat pada tabel 9 dibawah ini.

Tabel 9. Training Cycles SVM Berbasis PSO

Size	C	Epsilon	Accuracy	AUC
5	0.0	0.0	90.75 %	0.971
5	0.1	0.1	90.50 %	0.957
5	0.2	0.2	91.00 %	0.965
5	0.3	0.3	89.00 %	0.962
5	0.4	0.4	90.50 %	0.966
5	0.5	0.5	89.00 %	0.970
5	0.6	0.6	91.00 %	0.973
5	0.7	0.7	89.50 %	0.971
5	0.8	0.8	91.75 %	0.964
<b>5</b>	<b>0.9</b>	<b>0.9</b>	<b>92.75 %</b>	<b>0.973</b>
5	0.0	1.0	50.00 %	0.500
5	1.0	1.0	50.00 %	0.500
5	1.0	0.0	91.50 %	0.966

Sumber: Hasil Penelitian (2019)

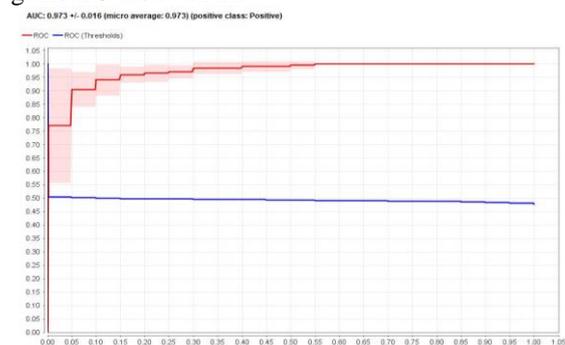
Hasil terbaik pada eksperimen SVM berbasis PSO diatas adalah C=0.9 dan Epsilon E=0.9 serta *population size*=5 menghasilkan accuracy=92.75% dan AUC=0.973. Hal ini menunjukkan bahwa dengan menggunakan optimasi *Particle Swarm Optimization* dapat meningkatkan akurasi yang lebih baik.

Tabel 10. Confusion Matrix SVM + PSO

accuracy: 92.75 % +/- 4.39 %			
	true negative	true positive	class precision
pred.negative	188	17	91.71 %
pred.positive	12	183	93.85 %
class recall	94.00 %	91.50 %	

Sumber: Hasil Penelitian (2019)

Nilai AUC dan akurasi yang dihasilkan setelah dilakukan optimasi menggunakan algoritma PSO meningkat, seperti yang terlihat pada tabel.9 dan gambar.3 dibawah ini.



Sumber: Hasil Penelitian (2019)

Gambar 3. Kurva ROC SVM+PSO

### 3. Algoritma *K-Nearest Neighbor* (k-NN)

Nilai k yang digunakan menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas pada data uji. Untuk memperkirakan nilai k yang terbaik, bisa dilakukan dengan menggunakan teknik validasi silang (*Cross Validation*).

Tabel 11. Eksperimen penentuan nilai *training* k-NN

k	Accuracy	AUC
1	79.00 %	0.500
2	79.00 %	0.816
3	84.25 %	0.888
4	83.25 %	0.907
5	84.25 %	0.919
6	85.75 %	0.931
7	86.25 %	0.936
8	88.00 %	0.940
9	87.25 %	0.946
10	88.50 %	0.948

Sumber: Hasil Penelitian (2019)

Hasil pengujian menunjukkan bahwa penerapan metode *k-Nearest Neighbor* pada tabel 11 dengan penentuan nilai k=10 menghasilkan *Accuracy*=88.50% dan *AUC*= 0.948 adalah nilai yang paling tertinggi.

Tabel 12. *Confusion Matrix* k-NN

accuracy: 88.50 % +/- 4.64 %			
	<i>true negative</i>	<i>true positive</i>	<i>class precision</i>
<i>pred.negative</i>	171	17	90.96 %
<i>pred.positive</i>	29	183	86.32 %
<i>class recall</i>	85.50 %	91.50 %	

Sumber: Hasil Penelitian (2019)



Sumber: Hasil Penelitian (2019)

Gambar 4. Kurva ROC k-NN

Kurva ROC tersebut diartikan dengan *False Positive* untuk garis *horizontal* dan *True Positive* untuk garis *vertikal* dengan nilai *AUC*= 0.948.

### 4. Algoritma k-NN Berbasis PSO

Hasil pengujian dengan menggunakan metode k-NN berbasis PSO dapat dilihat pada tabel 13 di bawah ini.

Tabel 13. Eksperimen penentuan nilai *training* k-NN berbasis PSO

Size	k	Accuracy	AUC
5	1	73.50 %	0.500
5	2	70.75 %	0.638
5	3	67.50 %	0.682
5	4	75.25 %	0.768
5	5	69.00 %	0.711
5	6	69.00 %	0.727
5	7	74.25 %	0.772
5	8	71.00 %	0.746
5	9	72.00 %	0.771
5	10	72.00 %	0.791

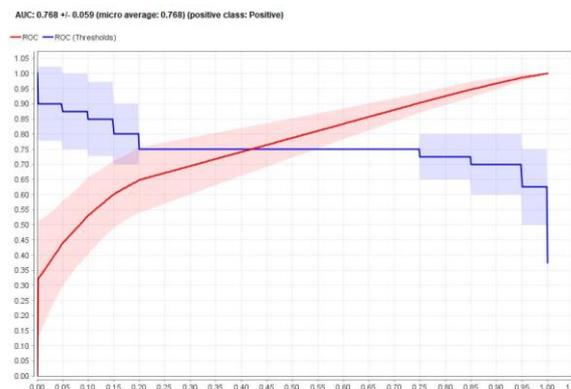
Sumber: Hasil Penelitian (2019)

Hasil perhitungan dari Tabel 13 di atas menunjukkan dengan memasukkan nilai k=4 serta *Population Size* = 5 dihasilkan *Accuracy*=75.25% dan *AUC*=0.768 adalah nilai yang tertinggi diantara nilai k yang lainnya, namun ternyata terjadi penurunan hasil akurasi pada k-NN sekitar 13% dari hasil akurasi metode k-NN tanpa optimasi berbasis PSO.

Tabel 14. *Confusion Matrix* k-NN berbasis PSO

accuracy: 75.25 % +/- 6.07 %			
	<i>true negative</i>	<i>true positive</i>	<i>class precision</i>
<i>pred.negative</i>	183	82	69.06 %
<i>pred.positive</i>	17	118	87.41 %
<i>class recall</i>	91.50 %	59.00 %	

Sumber: Hasil Penelitian (2019)



Sumber: Hasil Penelitian (2019)

Gambar 5. Kurva ROC K-NN Berbasis PSO

### KESIMPULAN

Berdasarkan pengujian yang telah dilakukan terhadap tweet fenomena tagar #2019GantiPresiden yang dikumpulkan menggunakan *Rapidminer Studio* sebanyak 400 data (200 positif dan 200 negatif). Dengan menggunakan metode *Support Vector Machine* (SVM), *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM+PSO), *k-Nearest Neighbor* (k-NN), dan *k-Nearest Neighbor* berbasis *Particle Swarm Optimization* (k-NN+PSO) maka hasilnya adalah hipotesa awal tidak sesuai

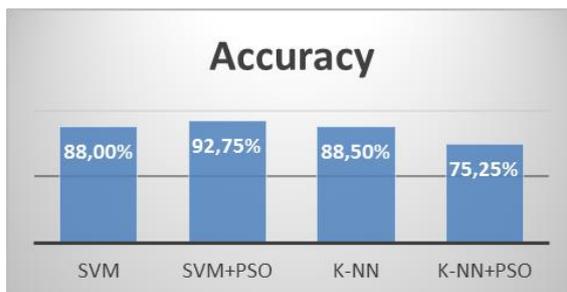
dengan hasil akhir. Dalam penelitian ini, hasil perhitungan metode SVM memiliki *Accuracy* sebesar 88.00% dan AUC sebesar 0.964 sedangkan Metode SVM+PSO menghasilkan *Accuracy* sebesar 92.75% dan AUC sebesar 0.973.

Pengujian juga telah dibandingkan dengan menggunakan metode k-NN dan k-NN berbasis PSO. Hasil perhitungan yang diperoleh dari pengujian data dengan metode k-NN yaitu menghasilkan *Accuracy* sebesar 88.50% dan AUC sebesar 0.948. Sementara itu metode k-NN berbasis PSO menghasilkan nilai *Accuracy* yang justru menurun yaitu sebesar 75.25% dan AUC sebesar 0.768.

Tabel 15. Perbandingan *Accuracy* dan AUC Algoritma Klasifikasi

Algoritma	Accuracy	AUC
SVM	88.00 %	0.964
SVM+PSO	92.75 %	0.973
k-NN	88.50 %	0.948
k-NN+PSO	75.25 %	0.768

Sumber: Hasil Penelitian (2019)



Sumber: Hasil Penelitian (2019)

Gambar 6. Perbandingan Akurasi

Penerapan *Particle Swarm Optimization* (PSO) terbukti dapat meningkatkan akurasi algoritma SVM pada klasifikasi tweet fenomena tagar #2019GantiPresiden untuk mengidentifikasi antara tweet positif dan tweet negatif. Sedangkan untuk algoritma k-NN justru menghasilkan akurasi yang sebaliknya yaitu lebih rendah dibandingkan algoritma k-NN tanpa PSO. Penerapan optimasi PSO pada algoritma k-NN mengalami penurunan sebesar 13%. Hal ini dapat disimpulkan bahwa dalam penelitian *text mining* penerapan optimasi menggunakan PSO belum tentu dapat mengoptimalkan nilai akurasi terutama pada algoritma k-NN.

Mengingat banyaknya penelitian terdahulu yang telah menerapkan *text mining* berbahasa Inggris dengan sukses meningkatkan nilai akurasi k-NN menggunakan PSO, maka dapat dianalisa bahwa optimasi PSO pada algoritma k-NN dengan menggunakan *dataset* Bahasa Indonesia tidak dapat meningkatkan nilai akurasi. Metode SVM terbukti lebih unggul dalam klasifikasi teks *tweet* fenomena

#2019GantiPresiden ini karena SVM bekerja dengan mencari parameter *hyperplane* yang terbaik yaitu nilai *C* dan *Epsilon* sehingga ada banyak kemungkinan akurasi dapat lebih optimal. Namun waktu pengujian data lebih lama dilakukan oleh SVM+PSO dibandingkan metode KNN+PSO.

Saran kedepan untuk penelitian selanjutnya dapat dilakukan dengan menggunakan dataset dari domain yang berbeda, misalnya review produk, review film, review saham dan lain sebagainya dengan teks berbahasa Indonesia dan meneliti optimasi perbaikan dari KNN+PSO untuk meningkatkan akurasi serta menggunakan teknik n-grams dengan tiga term (three-grams) pada tahap preprocessing untuk meningkatkan performa klasifikasi sebagai kontribusi penelitian lanjutan.

Menggunakan teknik pengklasifikasian teks yang lain dan metode pemilihan fitur yang lainnya, seperti Chi Square, Gini Index, Mutual Information, dan lain-lain agar hasil akurasinya dapat dibandingkan.

## REFERENSI

- Aryanti, R., Saepudin, A., Fitriani, E., Permana, R., & Saefudin, D. F. (2019). Komparasi Algoritma Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Pengguna Busway. *Jurnal Teknik Komputer*, 5(2), 89–96. <https://doi.org/10.31294/jtk.v5i2.5406>
- Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453–462. <https://doi.org/10.1016/j.proeng.2013.02.059>
- Berry, W., & Kogan, J. (2013). Michael W. Berry and Jacob Kogan (eds.): Text mining: applications and theory. *Information Retrieval*, 14(2), 208–211. <https://doi.org/10.1007/s10791-010-9153-5>
- Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Pham, A.-D. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Systems with Applications*, 41(8), 3955–3964. <https://doi.org/10.1016/j.eswa.2013.12.035>
- Ibrahim, I., Ibrahim, Z., Ahmad, H., & Yusof, Z. M. (2015). An Improved Multi-state Particle Swarm Optimization for Discrete Optimization Problems. *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*, 1, 3–8. <https://doi.org/10.1109/CICSyN.2015.11>
- Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan

- Metode K-Nearest Neighbor dan FeFeature Selection Query Expansion Ranking. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2739–2743.
- Nurhuda, F., Sihwi, S. W., & Doewas, A. (2014). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *ITSmart: Jurnal Ilmiah Teknologi Dan Informasi*, 2(2), 35–42.
- Rozi, I. F., Hadi, S., & Achmad, E. (2012). Implementasi Opinion Mining ( Analisis Sentimen ) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Universitas Stuttgart*, 6(1), 37–43.
- Utami, L. A. (2017). Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization. *Jurnal Pilar Nusa Mandiri*, 13(1), 103–112. Retrieved from <http://ejournal.nusamandiri.ac.id/ejournal/index.php/pilar/article/view/344/276>
- Wahyunanda Kusuma Pertiwi. (n.d.). Riset Ungkap Pola Pemakaian Medsos Orang Indonesia - Kompas.com. Retrieved July 5, 2018, from <https://tekno.kompas.com/read/2018/03/01/10340027/riset-ungkap-pola-pemakaian-medsos-orang-indonesia>