

# **Particle Swarm Optimization pada Analisa Review Software Antivirus Menggunakan Metode *K-Nearest Neighbors***

**Sucitra Sahara<sup>1\*</sup>, Rizqi Agung Permana<sup>2</sup>, Hariyanto<sup>3</sup>**

<sup>1</sup>Sistem Informasi; Universitas Bina Sarana Informatika

Jl. Kramat Raya No. 98, Senen, Jakarta Pusat 10450 Tlp. (021) 23231170; website

<http://bsi.ac.id>; email: [sucitrasahara@gmail.com](mailto:sucitrasahara@gmail.com)

<sup>2</sup>Teknologi Komputer; STMIK Antar Bangsa

Kawasan Bisnis CBD Ciledug, Jl. HOS Cokroaminoto No.29-35 Tangerang, Banten 15157;

email: [rizqiagung@gmail.com](mailto:rizqiagung@gmail.com)

<sup>3</sup>Sistem Informasi; Universitas Bina Sarana Informatika

Jl. Kramat Raya No. 98, Senen, Jakarta Pusat 10450 Tlp. (021) 23231170; website:

<http://bsi.ac.id>; email: [hariyanto.hro@bsi.ac.id](mailto:hariyanto.hro@bsi.ac.id)

\* Korespondensi: e-mail: [sucitrasahara@gmail.com](mailto:sucitrasahara@gmail.com)

Diterima: 15 April 2020 ; Review: 23 April 2020 ; Disetujui: 13 Mei 2020

Cara sitasi: Sahara S, Permana RA, Hariyanto, 2020. *Particle Swarm Optimization* pada Analisa Review Software Antivirus Menggunakan Metode *K-Nearest Neighbors*. *Informatics for Educators and Professionals*. Vol 4 (2): 123-132.

**Abstrak:** Virus pada komputer menjadi hal yang membahayakan bagi para pengguna komputer perorangan maupun perusahaan yang telah menerapkan sistem terkomputerisasi. Virus program yang didesain untuk tujuan jahat dapat merusak bagian tertentu dari komputer, bahkan yang paling merugikan adalah dapat merusak data penting pada perusahaan. Dalam hal ini maka diciptakanlah sebuah software anti virus, perkembangan anti virus selalu lebih lambat dari virus itu sendiri, sehingga peneliti akan mengadakan penyeleksian software anti virus pada suatu opini atau berdasarkan komentar masyarakat yang telah menggunakan software anti virus produk tertentu dan dituangkan ke media online seperti komentar pada suatu situs penjualan produk tersebut. Berdasarkan ribuan komentar akan diolah dan dikelompokkan pada jenis kata teks positif dan teks negatif, dan peneliti membuat klasifikasi data dengan menggunakan metode algoritma *k-Nearest Neighbor* (k-NN), algoritma k-NN adalah salah satu algoritma yang sesuai dalam penelitian kali ini. Peneliti menemukan bahwa algoritma k-NN mampu mengolah data set yang sudah dikelompokkan pada teks positif dan negatif khususnya dalam pemilihan teks, dan penerapan metode optimasi *Particle Swarm Optimization* (PSO) yang dikombinasikan dengan k-NN diharapkan mampu meningkatkan nilai akurasi sehingga datanya lebih kuat dan valid. Sebelum data set diolah menggunakan optimasi PSO hanya menggunakan metode k-NN akurasi data yang diperoleh 70,50% sedangkan hasil akurasi setelah penggunaan metode k-NN dan optimasi PSO didapatkan nilai akurasi sebesar 83,50%. Dapat disimpulkan bahwa penggunaan optimasi PSO dan metode k-NN sangat sesuai pada konsep text mining dan penyeleksian pada data set berupa text.

**Kata kunci:** Analisis Review, Particle Swarm Optimization, k-Nearest Neighbors.

**Abstract:** Viruses on computers become dangerous for individual computer users and companies that have implemented computerized systems. Virus programs that are designed for malicious purposes can damage certain parts of the computer, even the most detrimental is that it can damage important data on the company. In this case an anti-virus software is created, the development of anti-virus is always slower than the virus itself, so researchers will conduct an anti-virus software selection on an opinion or based on public comments that have used a

particular product's anti-virus software and poured it into online media such as comment on a product sales site. Of the thousands of comments will be processed and grouped on the type of positive and negative text words, and researchers make data classification using the *k*-Nearest Neighbor (*k*-NN) algorithm method, the *k*-NN algorithm is one of the appropriate algorithms in this study. The researcher found that the *k*-NN algorithm is able to process data sets that have been grouped in positive and negative texts, especially in text selection, and the application of the Particle Swarm Optimization (PSO) optimization method combined with *k*-NN is expected to be able to increase the accuracy value so that the data is stronger and valid. Before the data set is processed using PSO optimization only using the *k*-NN method the accuracy of the data obtained is 70.50% while the accuracy results after the use of the *k*-nn method and PSO optimization obtained an accuracy value of 83.50%. It can be concluded that the use of PSO optimization and the *k*-NN method are very compatible with the concept of text mining and correction of text data sets.

**Keywords:** Analysis Review, Particle Swarm Optimization, *k*-Nearest Neighbors.

## 1. Pendahuluan

Beberapa penelitian sebelumnya sudah dilakukan dalam klasifikasi sentimen terhadap review produk yang tersedia secara online seperti Analisa sentiment aplikasi smartphone dengan membandingkan metode *Support Vector Machine* (SVM) dan Naïve Bayes, Karena SVM adalah classifier, kemudian diberikan satu set contoh pelatihan, masing-masing ditandai sebagai milik salah satu dari dua kategori, algoritma pelatihan SVM membangun model yang memberikan contoh baru ke dalam satu kategori atau yang lain [1]. Kategorisasi teks merupakan solusi yang tepat untuk mengelola informasi yang saat ini berkembang dengan sangat cepat dan melimpah. Kategorisasi teks membuat pengelolaan informasi tersebut menjadi efektif dan efisien.

Dengan menggunakan kategorisasi teks, dapat dilakukan penyusunan dokumen menurut kategorinya, penyaringan terhadap email *spam*, melakukan penggalian opini (*opinion mining*) dan analisis sentimen. Algoritma kategorisasi teks saat ini telah banyak berkembang, antara lain: *Support Vector Machines* (SVM), *Naive Bayessian* (NB), pohon keputusan, *K-Nearest Neighbour* (*k*-NN). Metode *k*-Nearest Neighbors (*k*-NN) adalah Metode yang paling umum digunakan untuk menyelesaikan masalah ini, berdasarkan teknik *cross-validasi*, secara otomatis akan menentukan jumlah cluster dalam dataset. [2].

Pada penelitian lain metode Pendekatan berbasis jarak: *k*-nearest neighbors (*k*-NN) diperkenalkan dalam kerangka MIL [3]. Kategorisasi teks merupakan solusi yang tepat pada pengolahan informasi yang saat ini berkembang dengan sangat cepat dan melimpah. Kategorisasi teks membuat pengelolaan informasi tersebut menjadi efektif dan efisien. Algoritma *k*-Nearest Neighbors (*k*-NN) adalah penentu klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, tetapi bergantung pada label kategori yang melekat pada dokumen pelatihan mirip dengan dokumen tes.

*Particle Swarm Optimization* (PSO) Teknik PSO adalah pencarian berbasis populasi berdasarkan simulasi seperti perkumpulan burung, seperti algoritma komputasi evolusioner lainnya, PSO menghasilkan solusi populasi baru untuk setiap masalah iterasi. merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi secara global optimal dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel [4]. PSO banyak digunakan untuk memecahkan masalah optimasi serta pada seleksi fitur [5].

Banyak dari konsumen yang akan membeli produk khususnya antivirus, bingung dalam menentukan produk mana yang paling baik dari sekian produk yang ada, didukung betapa banyaknya produk software yang kurang akurat dalam menghentikan virus pada perangkat komputer mereka, dan banyak pula produk palsu. Kegiatan pemalsuan produk ini semakin marak pada masyarakat karena permintaan yang terus meningkat sehingga pemilik merek merasa kewalahan akan kegiatan pemalsuan barang ini [6]. *k*-NN menghasilkan model dimana peneliti menentukan nilai *k* untuk mencari tingkat keakurasian yang tinggi pada pengujian dataset. Nilai akurasi yang paling tinggi akan digunakan peneliti dalam menentukan apakah nilai tersebut optimal atau tidak, dan ternyata nilai tersebut sudah cukup baik. Maka disimpulkan bahwa struktur algoritma *k*-NN yang dirancang mencapai ideal dalam pemecahan masalah pada penelitian ini.

Pada penelitian kali ini, penggunaan metode pengklasifikasian k-Nearest Neighbors dengan metode optimasi *Particle Swarn Optimization* (PSO) sebagai metode pemilihan fitur akan penulis terapkan untuk klasifikasi *text* pada opini public mengenai *review* produk *software* antivirus yang terdapat pada situs penjualan produk tersebut dalam bentuk komentar, dan data komentar akan di proses dan diklasifikasikan berdasarkan object yang sudah ditentukan, pengolahan data tersebut akan menghasilkan 2 kategori yaitu komentar positif dan negatif, hasil akhir akan dituangkan pada aplikasi *review* produk *software antivirus* yang menunjukkan *software* tersebut baik atau tidak. Berikut ini ringkasan perangkat lunak antivirus terbaik di tahun 2020 dapat dilihat pada gambar 1 berikut ini.

- 1 [Norton 360](#), Pilihan Teratas untuk tahun 2020. Perlindungan malware yang kuat dengan beragam fitur, termasuk pengelola kata sandi, VPN, dan kontrol orang tua yang luar biasa
- 2 [McAfee Total Protection](#), Perlindungan Jaringan Rumah Terbaik. Segala yang Anda butuhkan untuk melindungi rumah Anda.
- 3 [Avira Prime](#), Terbaik untuk Optimasi Sistem. Melindungi komputer Anda dan membantu agar komputer itu berjalan lebih cepat.
- 4 [Bitdefender Total Security](#), Dengan Fitur Tambahan Terbanyak. Semua yang Anda butuhkan untuk melindungi setiap aspek keamanan online Anda.
- 5 [Kaspersky Internet Security](#), Terbaik untuk Belanja dan Perbankan Online. Perlindungan pembayaran yang aman dengan kemampuan antivirus yang kuat.
- 6 [TotalAV](#), Perangkat Lunak Antivirus Paling Mudah Digunakan. Paling tepat bagi pengguna yang baru mengenal perangkat lunak antivirus dan membutuhkan perlindungan yang kuat namun sederhana.
- 7 [BullGuard](#), Terbaik untuk Gamer. Melindungi dan mempercepat sistem Anda saat Anda bermain game..
- 8 [Panda Dome](#), Dengan Paket-paket yang Paling Fleksibel. Aneka pilihan harga untuk semua anggaran.
- 9 [Trend Micro](#), Terbaik untuk Perlindungan Phishing. Mencegah Anda agar tidak menjadi korban penipuan online.
- 10 [Cylance Smart Antivirus](#), Solusi Antivirus Paling Sederhana. Tidak ada fitur tambahan, tidak ada paket harga yang rumit hanya perlindungan malware yang kuat dan ringan.

Sumber: <https://id.safetydetectives.com/>

Gambar 1. Ringkasan perangkat lunak antivirus terbaik di tahun 2020.

Namun, dari data diatas kita belum dapat menyimpulkan bahwa produk tersebut baik untuk digunakan pada perangkat komputer kita. Dikatakan bahwa memiliki terlalu banyak komputer dalam kehidupan kita sehari-hari menyebabkan masalah keamanan, karena orang tidak dapat melindungi komputer mereka dari peretas dan virus internet [7]. Trend analisis ini menunjukkan peningkatan penargetan platform pada komputer sebagai sasaran malware sehingga aplikasi review pada produk software antivirus yang penulis buat sangat perlu untuk diprediksi dengan akurat sebelum konsumen membeli produk tersebut, dan sebagai bahan pertimbangan konsumen dalam memilih produk software antivirus. Analisis Sentimen bisa dikatakan jenis pengolahan Bahasa atau kalimat alami untuk melacak mood atau perasaan pada masyarakat menyangkut produk tertentu atau topik tertentu. Analisis sentimen, yang juga disebut tambang pendapat, melibatkan dalam membangun sistem untuk mengumpulkan dan meneliti pendapat tentang produk yang dibuat dalam posting blog, komentar, ulasan atau tweet. Analisis Sentimen dapat berguna dalam beberapa cara. Berikut adalah persamaan model Confution Matrix, terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix*. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Agar lebih mudah memahaminya, saya menggunakan contoh kasus sederhana untuk klasifikasi text pada review komentar Produk Software Antivirus.

Nilai *Accuracy* adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots(1)$$

*Sensitivity* digunakan untuk membandingkan proporsi TP terhadap tupel yang positif, yang dihitung dengan menggunakan persamaan:

$$Sensitivity = \frac{TP}{TP + FN} \dots\dots\dots(2)$$

*Specificity* digunakan untuk membandingkan proporsi TN terhadap tupel yang negatif, yang dihitung dengan menggunakan persamaan:

$$Specificity = \frac{TN}{TN + FP} \dots\dots\dots(3)$$

PPV (*positive predictive value*) adalah proporsi kasus dengan hasil diagnosa positif, yang dihitung dengan menggunakan persamaan:

$$PPV = \frac{TP}{TP + FP} \dots\dots\dots(4)$$

NPV (*negative predictive value*) adalah proporsi kasus dengan hasil diagnosa negatif, yang dihitung dengan menggunakan persamaan:

$$NPV = \frac{TN}{TN + FN} \dots\dots\dots(5)$$

Analisis Sentimen adalah jenis pengolahan bahasa alami untuk melacak mood masyarakat tentang produk tertentu atau topik. Analisis sentimen, yang juga disebut tambang pendapat, melibatkan dalam membangun sistem untuk mengumpulkan dan meneliti pendapat tentang produk yang dibuat dalam posting blog, komentar, ulasan atau tweet. Analisis Sentimen dapat berguna dalam beberapa cara. Misalnya, dalam pemasaran membantu in judging keberhasilan kampanye iklan atau peluncuran produk baru, menentukan versi produk atau jasa yang populer dan bahkan mengidentifikasi demografi suka atau tidak suka terhadap fitur tertentu [11].

## 2. Metode Penelitian

Teks yang akan dilakukan proses text mining, pada umumnya memiliki beberapa karakteristik diantaranya adalah memiliki dimensi yang tinggi, terdapat noise pada data, dan terdapat struktur teks yang tidak baik. Cara yang digunakan dalam mempelajari suatu data teks, adalah dengan terlebih dahulu menentukan fitur-fitur yang mewakili setiap kata untuk setiap fitur yang ada pada dokumen [8]. Data yang olah peneliti pada beberapa studi klasifikasi sentimen dikumpulkan dari situs e-commerce seperti [www.amazon.com](http://www.amazon.com) (review produk), [www.yelp.com](http://www.yelp.com) (ulasan restoran), [www.CNETdownload.com](http://www.CNETdownload.com) (review produk) dan [www.reviewcentre.com](http://www.reviewcentre.com), yang menjadi tuan rumah jutaan ulasan produk oleh konsumen. Selain itu, situs yang tersedia adalah situs review profesional seperti [www.dpreview.com](http://www.dpreview.com), [www.zdnet.com](http://www.zdnet.com) dan situs pendapat konsumen tentang topik yang luas, dan kali ini peneliti mengambil review Software anti virus pada [www.amazon.com](http://www.amazon.com) (review produk).

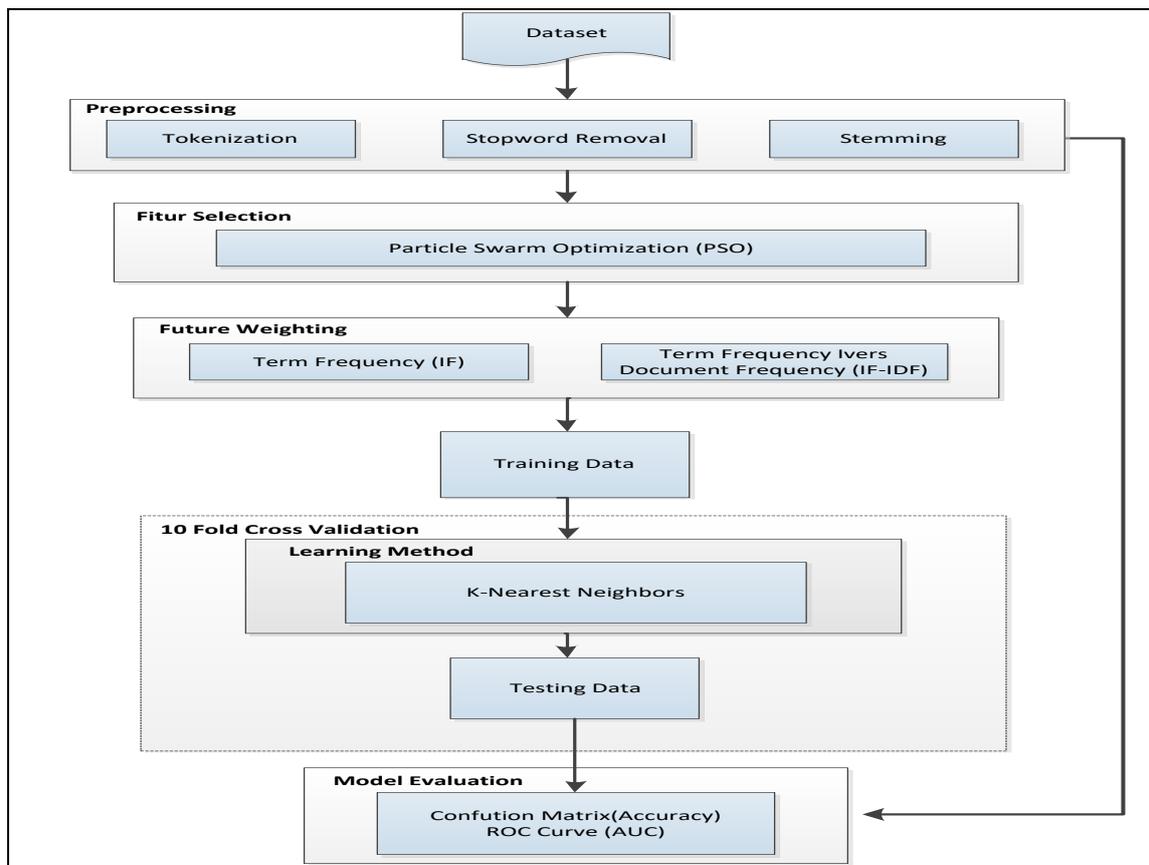
Setelah data dasar (komentar review) menjadi data terstruktur dan berupa nilai numerik maka data dapat disajikan sebagai sumber data yang dapat diolah lebih lanjut. Proses pengubahan bentuk bentuk menjadi data yang terstruktur sesuai kebutuhannya untuk proses dalam data mining, yang biasanya akan menjadi nilai-nilai numerik, proses ini sering disebut dengan *text processing* [9].

Pendekatan analisis mencakup tiga fase: pra-pemrosesan, di mana data mentah ditransformasikan menjadi format yang dapat digunakan, terutama dengan membersihkan, menetapkan atribut, dan mengintegrasikan data [9].

Beberapa proses yang dilakukan adalah sebagai berikut: a. *Tokenization* adalah peneliti menggunakan Tokenize untuk memisahkan kata atau huruf dari tanda baca dan simbol. b. *Stopwords Removal* adalah kata yang dianggap tidak perlu dalam pengolahan data sentimen review, sebagai contoh *if, the, of, or, etc.* c. *Steaming* adalah Proses perubahan bentuk kata menjadi kata dasar. Metode perubahan bentuk kata menjadi kata dasar ini menyesuaikan struktur bahasa yang digunakan dalam proses stemming [10].

Berikut langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen adalah: a). Definiskan domain dataset: Pengumpulan dataset yang melingkupi suatu domain, misalnya dataset review film, dataset review produk, dan lain sebagainya. b). *Pre-processing* Tahap pemrosesan awal yang umumnya dilakukan dengan proses *Tokenization, Stopwords removal, dan Stemming*. c). *Transformation*, Proses representasi angka yang dihitung dari data tekstual. *Binary representation* yang umumnya digunakan dan hanya menghitung kehadiran atau ketidakhadiran sebuah kata di dalam dokumen. Berapa kali sebuah kata muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual. Proses yang umumnya digunakan yaitu TF-IDF, *Binary transformation, dan Frequency transformation*. d). *Feature Selection*, Pemilihan fitur (*feature selection*) bisa membuat pengklasifikasi lebih efisien/efektif dengan mengurangi jumlah data untuk dianalisa dengan mengidentifikasi fitur yang relevan yang selanjutnya akan diproses. Metode pemilihan fitur yang biasanya digunakan adalah Expert. Knowledge, Minimum Frequency, Information gain, Chi-Square, dan lain sebagainya. e). *Classification*, Proses klasifikasi umumnya menggunakan pengklasifikasi seperti Naïve Bayes, Support Vector Machine, dan lain sebagainya.

Gambar metode usulan peneliti yang dirancang dalam review product dijelaskan dalam gambar 2 berikut ini:



Sumber: Hasil Penelitian (2020)

Gambar 2. Metode Usulan

Pada evaluasi kali ini penulis mengusulkan penggunaan model dalam kegiatan review aplikasi pada android yaitu model k-Nearest Neighbors (k-NN), yang dilakukan dalam dua tahap penerapan. Algoritma k-NN menghasilkan model dimana peneliti menentukan nilai k untuk mencari tingkat keakurasian yang tinggi pada pengujian dataset. Nilai akurasi yang paling tinggi akan digunakan peneliti dalam menentukan apakah nilai tersebut optimal atau tidak, dan ternyata nilai tersebut sudah cukup baik. Maka disimpulkan bahwa struktur algoritma yang dirancang mencapai ideal dalam pemecahan masalah. Data training yang digunakan dalam pengklasifikasian *text* terdiri atas 100 review positif pada review software antivirus dan 100 review negatif pada review software antivirus. Data review yang akan diolah masih berupa sekumpulan *text* yang dipisah dalam bentuk dokumen. Sebelum diklasifikasikan, data tersebut harus melewati proses tahapan agar data dapat diolah dengan baik.

### 3. Hasil dan Pembahasan

Pada hasil penelitian kali ini, peneliti menjabarkan hasil dari pengujian algoritma dan optimasinya yang menghasilkan nilai akurasi dan AUC.

#### 3.1. Hasil

Data training yang digunakan dalam pengklasifikasian *text* terdiri atas 100 review positif pada produk antivirus software dan 100 review negatif pada produk antivirus software. Data review yang akan diolah masih berupa sekumpulan *text* yang dipisah dalam bentuk dokumen. Sebelum diklasifikasikan, data tersebut harus melewati proses tahapan agar data dapat diolah dengan baik.

Dari beberapa metode yang digunakan untuk memvalidasi suatu model berdasarkan data yang didapat, seperti holdout, cross validation, random sub-sampling, dan lain-lain. *Confusion matrix* memberikan keputusan yang diperoleh dalam training dan testing, *Confusion matrix* memberikan penilaian performance klasifikasi berdasarkan objek dengan benar atau salah. *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi. Hasil Eksperimen Pengujian Metode k-NN, Hasil Nilai query instance dalam penelitian disini ditentukan dengan cara melakukan uji coba memasukkan nilai k (jumlah tetangga terdekat).

Tabel 1. Eksperimen k-NN

k	k-NN		Populations Size	k	k-NN+PSO	
	Accuracy	AUC			Accuracy	AUC
1	64.50%	0.500	5	1	70.10%	0.600
2	66.50%	0.652	5	2	70.30%	0.652
3	65.00%	0.402	5	3	65.00%	0.602
4	70.30%	0.645	5	4	70.50%	0.745
5	64.50%	0.504	5	5	64.50%	0.704
6	70.00%	0.790	5	6	70.00%	0.790
7	64.50%	0.700	5	7	64.50%	0.800
8	70.00%	0.607	5	8	70.00%	0.607
9	64.50%	0.708	5	9	64.50%	0.808
10	60.50%	0.795	5	10	70.50%	0.895

Sumber: Hasil Penelitian (2020)

Hasil eksperimen menunjukkan pada penerapan metode k-Nearest Neighbors pada tabel IV.5 dengan penentuan nilai k=10 yang nilai akurasinya mencapai 70,50% dan AUC 0.895 menunjukkan hasil yang paling tertinggi diantara penentuan nilai k yang lain. Metode k-Nearest Neighbors berbasis *Particle Swarm Optimization* Nilai training dalam penelitian ditentukan dengan cara melakukan uji coba terhadap nilai k sebagai nilai tetangga terdekat, berikut tahap uji cobanya:

Tabel 2. Eksperimen k-NN dengan PSO

Nilai k	Accuracy	AUC
1	70.10%	0.600
2	70.30%	0.652
3	65.00%	0.602
4	70.50%	0.745
5	64.50%	0.704
6	70.00%	0.790
7	64.50%	0.800
8	70.00%	0.607
9	64.50%	0.808
10	70.50%	0.895

Sumber: Hasil Penelitian (2020)

Hasil yang didapatkan pada eksperimen diatas, bahwa *k-Nearest Neighbors* berbasis PSO dengan menentukan nilai  $k=10$  dan population Size= $5$  menghasilkan accuracy= $83.50\%$ , nilai AUC= $0.896$  adalah nilai yang paling tertinggi.

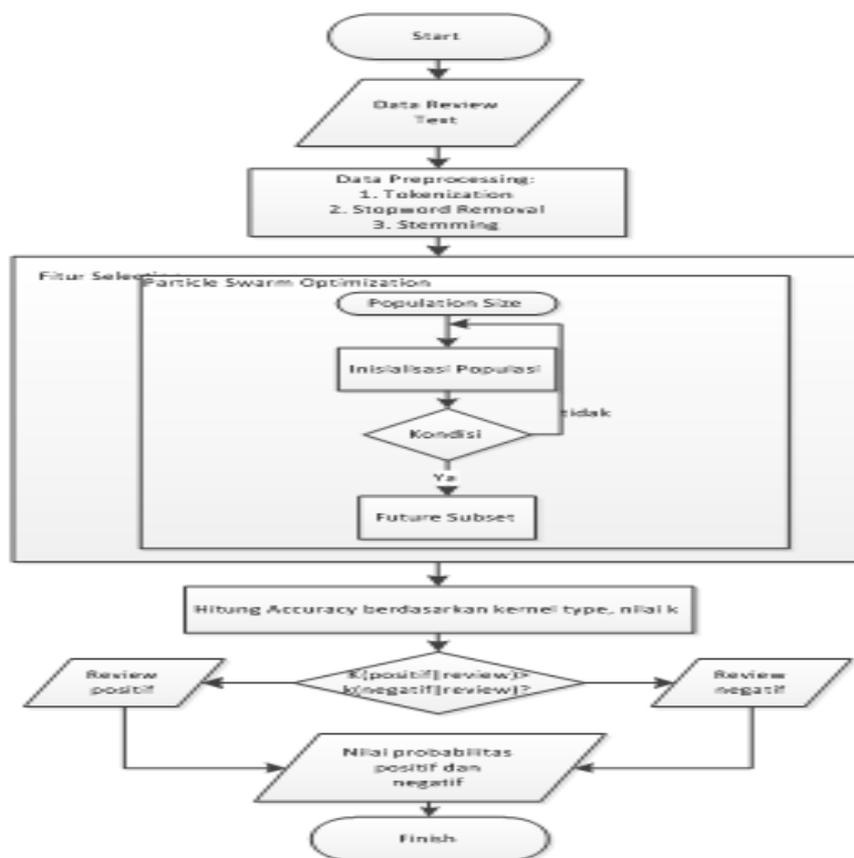
**3.2. Analisis Evaluasi Hasil dan Validasi Model**

Dari hasil pengujian yang peneliti lakukan dari awal pembahasan, pengukuran akurasi menggunakan confusion matrix dan kurva ROC membuktikan bahwa hasil pengujian algoritma *k-Nearest Neighbors* (*k-NN*) berbasis Particle Swarm Optimization (PSO) menghasilkan akurasi yang lebih tinggi dibandingkan dengan logika algoritma *k-Nearest Neighbors* (*k-NN*) tidak menggunakan model optimasi. Nilai akurasi untuk model algoritma *k-NN* sebesar  $70,50\%$  pada  $k=10$  dan nilai akurasi untuk model *k-NN* berbasis PSO sebesar  $83.50\%$  pada  $k=8$ , mempunyai peningkatan akurasi sebesar  $13.00\%$ . Kesimpulan pengujian ini adalah penerapan PSO pada algoritma *k-NN* dapat meningkatkan nilai akurasi yang merupakan solusi yang baik dalam permasalahan pada klasifikasi sentimen review produk antivirus.

**3.3. Pembahasan**

Penelitian ini merupakan Penerapan PSO pada algoritma *k-NN* dapat meningkatkan nilai akurasi pada klasifikasi review produk antivirus software untuk mengidentifikasi antara review positif dan review negatif, dengan memiliki model klasifikasi teks pada review, pembaca dapat dengan mudah mengidentifikasi mana review yang positif maupun yang negatif. Dari data review yang sudah ada, dipisahkan menjadi kata-kata, lalu diberikan bobot pada masing-masing kata tersebut. Dapat dilihat kata mana saja yang berhubungan dengan sentimen yang sering muncul dan mempunyai bobot paling tinggi. Dengan demikian dapat diketahui review tersebut termasuk review positif atau review negatif.

Dapat dilihat dari diagram alir atas proses klasifikasi pada aplikasi yang dibentuk.



Sumber: Hasil Penelitian (2020)

Gambar 3. Diagram Alir Tahapan Proses Klasifikasi Algoritma Support Vector Machine Berbasis PSO

Dalam penelitian ini, hasil pengujian model akan dibahas melalui *confusion matrix* untuk menunjukkan model yang terbaik. Tanpa menggunakan metode pemilihan fitur, k-Nearest Neighbors sendiri sudah menghasilkan akurasi yang cukup tinggi sebesar 70,50% dan nilai AUC 0.825. Akurasi tersebut masih bisa ditingkatkan lagi, sehingga perlu ditingkatkan menggunakan metode pemilihan fitur PSO. Setelah menggunakan metode pemilihan fitur dari filter 83.50% dan nilai AUC 0.928.

Peneliti menyediakan aplikasi berbasis web untuk menguji model menggunakan dataset yang berbeda dan belum diklasifikasikan sesuai dengan kelasnya. Diaplikasikan dengan menggunakan bahasa pemrograman PHP berbasis Web. Peneliti juga menyediakan aplikasi untuk menguji model menggunakan dataset yang berbeda dan belum diklasifikasikan sesuai dengan kelasnya. Diaplikasikan dengan menggunakan bahasa pemrograman PHP berbasis Web.

### 3.4. Implikasi Penelitian

Implikasi penelitian ini mencakup beberapa aspek, di antaranya: a. Implikasi terhadap aspek sistem Hasil evaluasi menunjukkan penerapan Particle Support Optimization (PSO) untuk fitur selection yang dapat meningkatkan akurasi k-Nearest Neighbors (k-NN) dan merupakan metode yang cukup baik dalam mengklasifikasi teks review aplikasi pada produk antivirus. Metode ini dapat membantu para calon pengguna komputer dalam menentukan aplikasi antivirus apa saja yang layak mereka pasang pada perangkat komputer mereka, supaya tidak lagi asal menginstall aplikasi antivirus yang banyak tersedia diberberapa media online, yang dampak buruknya bisa mengakibatkan memory dan kualitas mobile pengguna melemah kualitasnya, bahkan ada beberapa produk antivirus yang palsu bahkan malah di tect sebagai virus, dan mengurangi waktu dalam membaca review dan komentar mengenai produk antivirus software. b. Implikasi Aspek Manajerial Membantu para pengembang dan vendor sistem yang berkaitan dengan dunia komputer khususnya antivirus, baik dari sumber sosial media atau dari situs resmi para pengusaha dibidang antivirus software, agar menggunakan aplikasi RapidMiner dalam membangun suatu sistem. c. Implikasi terhadap aspek penelitian lanjutan Penelitian selanjutnya bisa menggunakan metode pemilihan fitur ataupun dataset dari domain yang berbeda, seperti review hotel, review restoran, dan banyak lainnya yang bisa dicari dalam bidang pengembangannya.

Penelitian ini mencakup beberapa aspek, di antaranya: 1). Implikasi terhadap aspek sistem Hasil evaluasi menunjukkan penerapan Algoritma k-Nearest Neighbors (k-NN) merupakan metode yang cukup baik dalam mengklasifikasi teks review Software Anti Virus. Metode ini dapat membantu para calon pengguna Komputer dalam menentukan software yang tepat dan layak mereka install, supaya tidak lagi asal menginstall software yang banyak tersedia diberberapa media online yang belum terjamin kualitasnya, yang dapat dampak buruknya bisa mengakibatkan memory dan kualitas mobile pengguna melemah kualitasnya karena terganggu oleh virus, dan mengurangi waktu dalam membaca review dan komentar mengenai produk antivirus. 2). Implikasi Aspek Manajerial Membantu para pengembang dan vendor sistem yang berkaitan dengan dunia software, baik dari sumber sosial media atau dari situs resmi para pengusaha dibidang produk ini, agar menggunakan aplikasi penyeleksian review dalam membangun suatu sistem. 3). Implikasi terhadap aspek penelitian lanjutan Penelitian selanjutnya bisa menggunakan metode pemilihan fitur ataupun dataset dari domain yang berbeda, seperti review maskapai, review handphone, dan banyak lainnya yang bisa dicari dalam bidang pengembangannya.

### 4. Kesimpulan

Klasifikasi text dengan data berupa review produk antivirus software yang peneliti buat, ternyata akurasi data lebih optimal dan nilai nya cukup tinggi dengan metode pengklasifikasian k-Nearest Neighbors (k-NN). Hal ini dikarenakan k-NN metode yang dapat sesuai dengan klasifikasi data dan mudah dipahami. Dapat dilihat kata mana saja yang berhubungan dengan sentimen yang sering muncul dan mempunyai bobot paling tinggi. Dengan demikian dapat diketahui review tersebut termasuk review positif atau review negatif. k-NN juga sering digunakan pada beberapa peneliti dalam klasifikasi teks dan memiliki performa yang baik. k-NN

juga sering digunakan pada beberapa peneliti dalam klasifikasi teks dan memiliki performa yang baik.

Berdasarkan pengolahan data yang sudah dilakukan. Data review yang peneliti olah dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi k-NN sebelum menggunakan penggabungan metode pemilihan fitur mencapai 70,50% Sedangkan setelah menggunakan penggabungan metode selection fitur *Particle Swarm Optimization* (PSO) akurasi meningkat hingga mencapai 83.50%. Peningkatan akurasi mencapai 14.5%. Untuk memudahkan penelitian, dibuatlah aplikasi review produk antivirus software untuk mengklasifikasikan review positif dan negatif yang ditampilkan dalam bentuk chart menggunakan bahasa pemrograman PHP.

Model yang terbentuk dapat diterapkan pada seluruh data review produk antivirus software dari berbagai sumber, sehingga dapat dilihat secara langsung hasilnya dalam bentuk positif dan negatif (chart). Hal ini dapat membantu seseorang untuk menghemat waktu saat mencari produk antivirus yang akan digunakan baik atau tidak.

Kedepannya peneliti akan mencari model fitur selection lain yang dapat digunakan untuk klasifikasi data berupa text supaya perbandingan hasilnya dapat lebih optimal, mengimplementasikan model k-NN berbasis PSO pada data review produk atau jasa lainnya, dalam pengembangan pengujian selanjutnya adalah memperbaiki proses data preparation dimana kualitas data yang akan diolah menjadi lebih baik sehingga pengolahan pada proses text mining menjadi lebih optimal.

### Ucapan Terima Kasih

Alhamdulillah puji syukur kepada Allah SWT, karena atas kehendak dan ridhaNya peneliti dapat menyelesaikan karya ilmiah ini. Kami sadari penelitian ini tidak akan selesai tanpa doa, dukungan dan dorongan dari berbagai pihak. Adapun dalam kesempatan ini peneliti ingin mengucapkan banyak terima kasih kepada penyelenggara jurnal *Informatics For Educators And Professionals* yang telah memberi kesempatan kepada kami dan membantu kami para penulis untuk penerbitan karya ilmiah ini.

### Referensi

- [1] Y. Yin, D. Han, and Z. Cai, "Explore Data Classification Algorithm Based on SVM and PSO for Education Decision," *J. Conver. Inf. Technol.*, vol. 6, no. 10, pp. 122–128, 2011.
- [2] F. Gorunescu, *Data mining: concepts and techniques*. 2011.
- [3] A. Zafra and S. Ventura, "Multi-instance genetic programming for predicting student performance in web based educational environments," *Appl. Soft Comput. J.*, vol. 12, no. 8, pp. 2693–2706, 2012.
- [4] P. J. García Nieto, E. García-Gonzalo, F. Sánchez Lasheras, and F. J. de Cos Juez, "Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability," *Reliab. Eng. Syst. Saf.*, vol. 138, pp. 219–231, 2015.
- [5] F. Zhao, Y. Liu, C. Zhang, and J. Wang, "A self-adaptive harmony PSO search algorithm and its performance analysis," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7436–7455, 2015.
- [6] P. Kasih, "Perlindungan hukum bagi masyarakat terhadap peredaran barang palsu," *J. Univ. Atma Jaya Yogyakarta*, vol. 5, no. 6, pp. 12–21, 2016.
- [7] M. Berndtsson, J. Hansson, B. Olsson, and B. Lundell, *Thesis Guide - A Guide for Students in Computer Science and Information Systems*. 2010.
- [8] D. P. Langgeni, Z. K. A. Baizal, and Y. F. a W, "Clustering Artikel Berita Berbahasa Indonesia Menggunakan Unsupervised Feature Selection," *Semin. Nas. Inform. 2010*, vol. 2010, no. semnasIF, pp. 1–10, 2010.
- [9] M. Abdous, W. He, and C. J. Yen, "Using data mining for predicting relationships

- between online question theme and final grade,” *Educ. Technol. Soc.*, vol. 15, no. 3, pp. 77–88, 2012.
- [10] M. Sholehudin, M. Fauzi Ali, and S. Adinugroho, “Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi ( Studi Kasus: Universitas Brawijaya ),” vol. 2, no. 11, pp. 5518–5524, 2018.
- [11] G, Vinodhini & Chandrasekaran, Dr. (2012). Sentiment Analysis and Opinion Mining: A Survey. *Int J Adv Res Comput Sci Technol.* 2.