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*)

Valentino Kevin Sitanayah Que[1], Ade Iriani[2], Hindriyanto Dwi Purnomo[3]

***Abstract*—Phenomenon of online transportation with some problems like crime and fraud in Indonesia triggers pros and cons to Twitter users. This study aims to find out sentiments of the society on online transportation and compare the accuracy of SVM and SVM-PSO with default parameters value. The proposed solution divided the dataset into training and testing data, because some researches only used one dataset that had already been classified. The research data is tweet data, which is obtained through scraping method using Octoparse. A total of 1,852 tweets from 1/1/2019 to 15/10/2019 were divided into 1,130 tweet testing data and 722 tweet training data. Then, RapidMiner was used for analysis process. Analysis positive sentiment using SVM is 62% and negative sentiment is 38%, while in SVM-PSO, positive opinion is 53% and negative opinion is 47%. The results of research using 10 k-fold CV produce accuracy on SVM is 95.46% and AUC is 0.979 (excellent classification), while in SVM-PSO accuracy is 96.04% and AUC is 0.993 (excellent classification). The results show that use of training and testing data on this study can be done and prove that SVM-PSO is better than ordinary SVM, although the parameters value is default.**

***Intisari*—Terdapat fenomena transportasi *online* dengan masalah seperti kriminalitas dan penipuan di Indonesia yang memicu pro dan kontra pada pengguna Twitter. Makalah ini bertujuan mengetahui sentimen masyarakat terhadap transportasi *online* dan membandingkan akurasi SVM dan SVM-PSO dengan nilai parameter *default*. Solusi yang diusulkan adalah membagi *dataset* ke dalam data *training* dan *testing*, karena beberapa penelitian mengenai optimasi hanya menggunakan satu *dataset* yang sudah diklasifikasikan. Data penelitian adalah data *tweet* dengan metode *scraping* menggunakan Octoparse. Total 1.852 data *tweet* dari 1/1/2019 hingga 15/10/2019 yang dibagi menjadi data *testing* 1.130 *tweet* dan *training* 722 *tweet* serta RapidMiner digunakan untuk proses analisis. Analisis sentimen positif menggunakan SVM adalah sebesar 62% dan sentimen negatif sebesar 38%, sedangkan pada SVM-PSO, opini positif sebesar 53% dan negatif 47%. Hasil penelitian menggunakan 10 *k-fold* CV menghasilkan akurasi pada SVM sebesar 95,46% dan AUC 0,979 (*excellent classification*), sedangkan pada SVM-PSO sebesar 96,04% dan AUC 0,993 (*excellent classification*). Hasil menunjukkan bahwa penggunaan data *training* dan *testing* dapat dilakukan dan terbukti bahwa SVM-PSO lebih baik daripada SVM biasa, meskipun menggunakan nilai parameter *default*.**

***Kata Kunci*—Analisis Sentimen, Twitter, *Support Vector Machine*, *Particle Swarm Optimization*, Transportasi *Online*.**

[1] *Program Studi Magister Sistem Informasi Universitas Kristen Satya Wacana, Jl. Dr. O. Notohamidjodjo Kota Salatiga 50715 INDONESIA (e-mail: vkevinsque11@gmail.com, ade.iriani@uksw.edu[2]; hindriyanto.fti@gmail.com[3])*

I. PENDAHULUAN

Electronic Commerce (*e-commerce*) telah mengalami perkembangan teknologi yang sangat pesat. Kenyamanan dan kemudahan yang ditawarkan berbagai platform *e-commerce* membuat banyak pihak beralih meninggalkan transaksi dengan tradisi lama, yaitu dengan cara tatap muka. *E-commerce* yang sering digunakan masyarakat Indonesia adalah *e-commerce* yang menjual jasa transportasi. Berbeda dari transportasi umum, jasa transportasi *online* memanfaatkan aplikasi sebagai mediator antara konsumen dan *driver*.

Transportasi *online* adalah sebuah transportasi yang dimodifikasi dari ojek konvensional yang umumnya berada di suatu pangkalan. Aplikasi transportasi *online* juga terdiri atas beragam pelayanan, yaitu pemesanan makanan, pembelian tiket, antar jemput barang, belanja, dan sebagainya, yang sangat membantu masyarakat dalam efisiensi waktu. Kemudahan dalam menggunakan aplikasi dan fitur-fitur yang membantu mobilitas dengan menggunakan *smartphone* membuat aplikasi transportasi *online* semakin diminati, terutama di kota-kota besar yang angka kemacetannya tinggi.

Berdasarkan survei YLKI, 45 persen konsumen transportasi *online* pernah dikecewakan. Bahkan kini terbukti bahwa transportasi *online* tidak senyaman dan tidak seaman yang dibayangkan sebelumnya. Berbagai kriminalitas, termasuk pembunuhan, beberapa kali terjadi di transportasi *online* dan korban utamanya adalah konsumen. Di sisi lain, *driver* juga hanya menjadi korban eksploitasi para kapitalis yang menguasai sistem transportasi *online* [1]. Sebuah eksperimen mengenai kesan konsumen tentang ojek *online* di Indonesia dilakukan oleh *Research Institute of Economic Development* (RISED). Hasil eksperimen menyatakan bahwa sejumlah 71,12% konsumen transportasi *online* di Indonesia bisa jadi berpaling menggunakan kendaraan pribadi jika tarif transportasi *online* naik [2]. Tarif transportasi *online* yang mahal dapat membuat perusahaan transportasi rugi besar karena kehilangan konsumen. Kelemahan dari pemakaian transportasi *online* adalah konsumen merasa terusik terhadap *driver* yang menyimpan dan menghubungi nomor konsumen melalui Whatsapp. Hal tersebut mengusik ranah pribadi konsumen, terutama pada lingkup seorang karyawan dan konsumennya [3]. Hal ini berakibat banyak pendapat atau 'cuitan' dari masyarakat pengguna Twitter berupa kritikan, keluhan, atau juga dukungan terhadap perusahaan transportasi *online*. Beberapa masalah yang menjadikan transportasi *online* terkesan negatif pada beberapa masyarakat menjadi menarik untuk diteliti, sehingga dibutuhkan sistem untuk menganalisis sentimen masyarakat pengguna transportasi *online*.

Masyarakat kerap memberikan opini dan pendapat menggunakan berbagai media sosial, salah satunya adalah Twitter. Menurut data dari *The Statistic Portal*, jumlah pengguna Twitter aktif di Indonesia pada 2019 mencapai 22,8 juta, naik dari 12 juta pada 2014 [4]. Perusahaan transportasi *online* memiliki akun resmi di Twitter yang menampung *tweet* komentar seperti pelayanan, *driver*, dan aplikasi, maupun memberikan informasi yang terkini. Opini dari masyarakat yang dituangkan ke dalam media sosial Twitter menjadi menarik untuk diketahui sentimennya, yaitu positif atau negatif, terhadap transportasi *online*. Faktor-faktor yang dapat memengaruhi sentimen positif adalah lebih fleksibel, banyak diskon, dan mempermudah aktivitas, sedangkan faktor negatif adalah mahal karena harga ongkos naik, meningkatkan angka kemacetan, tingkat kriminalitas, dan penipuan dari *driver* maupun konsumen. Data dari opini sangat berperan sebagai umpan balik layanan tanpa perlu memperoleh opini secara langsung dari masyarakat guna analisis sentimen sebuah produk atau layanan.

Analisis sentimen merupakan metode untuk memperoleh data dari berbagai platform yang tersedia di internet. Analisis sentimen berpusat terhadap analisis serta pengertian emosi dari *review* teks yang bertujuan untuk prediksi, analisis suasana publik, suasana hati, dan gambaran perasaan secara otomatis para *netizen* pada suatu kasus [5], [6]. Tujuan dari analisis sentimen untuk mengumpulkan polaritas dari teks atau opini pada dokumen bersifat positif atau negatif. Media sosial Twitter dipilih karena Twitter merupakan media sosial yang populer di kalangan pengguna internet saat ini.

Berdasarkan latar belakang tersebut, makalah ini bertujuan untuk memahami sentimen masyarakat tentang transportasi *online* menggunakan metode klasifikasi *Support Vector Machine* (SVM) dan SVM optimasi *Particle Swarm Optimization* (SVM-PSO). Fokus makalah adalah menganalisis sentimen pada transportasi *online* dan membandingkan akurasi SVM dan SVM-PSO. Akurasi dan nilai AUC dibandingkan untuk mengetahui metode klasifikasi SVM biasa atau SVM berbasis PSO yang lebih baik dalam proses klasifikasi. Beberapa penelitian mengenai optimasi hanya menggunakan satu *dataset* yang sudah dibagi positif dan negatifnya. Data yang digunakan juga hanya sedikit. Solusi yang diusulkan dalam makalah ini adalah membagi *dataset* ke dalam data *training* dan data *testing*, tidak hanya menggunakan satu *dataset* yang di dalamnya sudah diklasifikasikan berdasarkan positif atau negatif sebuah opini. Parameter yang digunakan tidak di ubah atau *default*. SVM adalah model *supervised learning* untuk mengidentifikasi pola-pola dan menganalisis data untuk klasifikasi. Kelebihan SVM adalah dapat mengidentifikasi *hyperplane* terpisah yang memaksimalkan *margin* pada dua kelas berbeda. Sementara itu, PSO adalah salah satu teknik optimalisasi yang digunakan untuk meningkatkan tingkat akurasi. Dalam hal ini, PSO digunakan sebagai solusi untuk menentukan bobot fitur-fitur yang sesuai, sehingga terjadi peningkatan akurasi.

II. KONTEN UTAMA

Penelitian terkait yang menggunakan SVM dan PSO menunjukkan bahwa SVM optimasi PSO dapat meningkatkan akurasi. Sebanyak tiga ratus data penelitian yang berasal dari *review* hotel diambil dari situs www.tripadvisor.com, yang terbagi menjadi 150 *review* opini negatif dan 150 opini positif. SVM menghasilkan akurasi 91,33% dan nilai AUC 0,988. Akurasi SVM-PSO meningkat 5,61% menjadi 96,94 dan nilai AUC sebesar 0,992 yang termasuk *excellent classification* [7].

Penelitian mengenai analisis sentimen juga dilakukan menggunakan metode SVM dan *Naive Bayes Classifier* (NBC). Penelitian ini menggunakan data *tweet* pada Twitter @ambonlima, dengan total 1.491 data *tweet* per 1 Januari 2015 sampai 30 April 2018 yang diperoleh dengan metode *snipping*, dibagi menjadi 491 data *training* dan 1.000 data *testing*. Hasil penelitian menunjukkan bahwa NBC cenderung positif dengan akurasi 67,20%, sedangkan SVM cenderung negatif dengan akurasi 81,67%, sehingga dapat diasumsikan SVM lebih baik pada kasus ini [8].

Penelitian lainnya mengenai optimasi dilakukan menggunakan *Genetic Algorithm* (GA), dengan menggunakan *dataset* pada situs www.aemo.au, yaitu terkait permintaan konsumsi energi listrik di Australia bagian wilayah Victoria. Data diambil pada tahun 2013 dalam periode waktu tiga bulan, dimulai 01/10/2013 jam 24.00 hingga 31/12/2013 jam 23.30. Nilai RMSE pada NN sebesar 97.174 dan NN+GA sebesar 94.549, sedangkan SVM sebesar 311.032 dan SVM+GA sebesar 215.158. GA mampu meningkatkan akurasi dengan ditandai adanya perubahan nilai RMSE yang lebih kecil dari sebelumnya. Berdasarkan hasil penelitian, NN menggunakan GA lebih baik dibandingkan SVM menggunakan GA [9].

Artikel-artikel sebelumnya yang telah diteliti menjelaskan analisis sentimen berhasil dilakukan untuk mengetahui sentimen masyarakat. Berdasarkan penelitian terkait, dilakukan penelitian mengenai analisis sentimen transportasi *online*. Penelitian optimasi hanya menggunakan satu dataset yang sudah diklasifikasikan ke positif dan negatif dan menggunakan data yang tergolong sedikit [7]. Oleh karena itu, makalah ini menggunakan data *training* dan data *testing* untuk mengklasifikasi sentimen dan mengambil data *tweet* yang tergolong banyak menggunakan aplikasi Octoparse, agar akurasi klasifikasi lebih tinggi. Metode PSO diterapkan pada SVM untuk melihat akurasi sentimen naik atau tidak dan validasi diuji dengan 10 *k-fold cross validation*. Evaluasi pengukuran akurasi menggunakan *confusion matrix* dan hasil olahan data dalam bentuk kurva *Receiver Operating Characteristics* (ROC) untuk mengukur nilai *Area Under Curve* (AUC). Proses analisis sentimen menggunakan aplikasi RapidMiner dengan menggunakan metode SVM dan SVM-PSO. Perbandingan nilai akurasi dan nilai AUC pada metode klasifikasi SVM dan SVM-PSO dilakukan untuk mengetahui klasifikasi SVM-PSO lebih baik dari klasifikasi SVM biasa.

A. Analisis Sentimen

Opinion mining atau analisis sentimen merupakan suatu bidang ilmu dari *data mining* yang berguna untuk menganalisis, mengolah, dan mengekstrak data tekstual pada entitas, seperti layanan, produk, individu, organisasi, peristiwa, atau masalah dan topik tertentu [10]. Analisis ini berfungsi untuk mendapatkan sebuah informasi dari suatu

himpunan data yang ada. Analisis sentimen adalah penelitian yang baru pada *Natural Language Processing* (NLP) dan bertujuan menemukan subjektivitas dalam teks maupun mengekstraksi dan menjalankan klasifikasi sentimen pada opini [11].

Terdapat tiga teknik pada metode klasifikasi sentimen, yakni *hybrid approach*, *lexicon based*, dan *machine learning* [11], [12]. Pada era ini, penelitian analisis sentimen dilakukan dengan *machine learning* karena dapat memprediksi polaritas sentimen (positif, negatif, ataupun netral) berdasarkan data *training* pada data *testing* [11]. Proses analisis sentimen sebagaimana diilustrasikan adalah teks tidak teratur, mencakup teks pada *review*, forum, *tweet*, dan *blog*. *Pre-processing* data mencakup proses *tokenisasi*, *stopword removal*, *stemming*, identifikasi sentimen, dan klasifikasi sentimen [12].

B. Pre-Processing

Pre-processing penting pada data *training* guna mendukung proses melatih algoritme agar dapat mengoreksi data yang tidak terorganisasi menjadi terorganisasi, sehingga mampu menyederhanakan pemrosesan data. Pengumpulan data opini dari media sosial Twitter terkadang tidak sama dengan kata baku, kata-kata yang tidak terdapat di kamus, memakai bahasa daerah, atau disingkat. Untuk mengembalikan sejumlah teks ke teks alami dengan mengeliminasi ekspresi atipikal agar dapat meminimalkan *noise* pada tahap selanjutnya, diperlukan *pre-processing* atau normalisasi untuk mengatasi hal ini [13].

Pre-processing dilakukan dalam enam tahap, yaitu sebagai berikut [12]-[14].

1) Cleansing: Tahap ini adalah tahap eliminasi aksara nonalfabetis untuk menurunkan *noise*. Aksara yang dihapus adalah tanda baca seperti titik (.), koma (,), tanda tanya (?), dan tanda seru (!), serta simbol-simbol seperti tanda '@' untuk *username*, *hashtag* (#), emotikon, dan alamat *website*.

2) Case Folding: Case folding adalah tahap untuk mengonversi karakter alfabet yang telah melalui tahap *cleansing* ke huruf kecil (*lower case*).

3) Tokenizing: Tahap ini berfungsi sebagai pemecah kalimat berdasarkan tiap kata yang menyusunnya, yang disebut *term* atau *token*. *Tokenizing* dipecah berdasarkan spasi.

4) Normalization atau Konversi Slangword: Tahap ini dilakukan agar kata-kata yang disingkat atau diperpanjang menjadi kata-kata yang normal sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Konversi *slangword* adalah proses mengubah kata tidak baku menjadi kata baku. Tahap ini dilakukan dengan bantuan kamus *slangword* dalam kata-kata baku dan memeriksa kata tersebut terdapat dalam kamus *slangword* atau tidak. Jika kata tidak baku terdapat dalam kamus, maka kata tidak baku diubah ke kata baku yang terdapat di dalam kamus.

5) Filtering atau Stopword Removing: Tahap ini memroses agar kata-kata yang tidak penting atau tidak bermakna dihapus untuk analisis sentimen. Contoh kata-kata tersebut adalah atau, yang, dengan, di, ke, dan tetapi.

6) Stemming: Tahap ini berfungsi mengubah kata berimbuhan pada tiap kata yang telah terseleksi menjadi kata dasar.

C. Term Weighting

Term weighting merupakan pembobotan tiap-tiap kata agar dapat menaikkan kemampuan analisis sentimen pada proses *text mining* [15], [16]. Penelitian ini memanfaatkan *Term Frequency-Inverse Document Frequency* (TF-IDF), yang diimplementasi menggunakan *tools* RapidMiner yang dilakukan dengan operator *Process Document from Data* (ekstensi dari *Text Processing*). *Term Frequency* (tf(w,d)) dianggap memiliki proporsi kepentingan sesuai total kemunculannya dalam teks atau dokumen. *Inverse Document Frequency (IDF)* merupakan metode pembobotan token yang berfungsi untuk memonitor kemunculan *token* dalam himpunan teks. TF-IDF adalah statistik untuk memperlihatkan vitalnya sebuah kata pada *dataset* atau dokumen [12].

Data yang telah melalui tahap *pre-processing* harus berbentuk numerik. TF-IDF digunakan untuk mengubah data tersebut menjadi numerik. Dalam perhitungan bobot menggunakan TF-IDF, dihitung lebih dulu nilai TF per kata dengan bobot masing-masing kata adalah 1. IDF(*word*) adalah nilai IDF dari setiap kata yang dicari, *td* adalah jumlah keseluruhan dokumen yang ada, dan *df* adalah jumlah kemuculan kata pada semua dokumen. IDF diformulasikan sebagai (1).

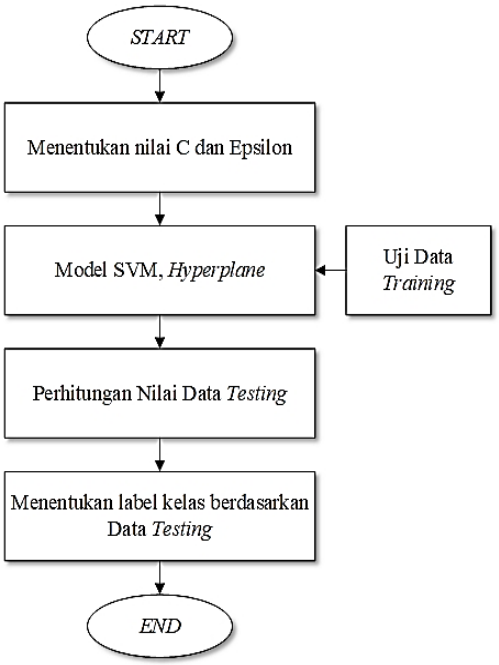
$$IDF(word) = log\frac{td}{df} \quad (1)$$

D. Support Vector Machine

Metode SVM merupakan suatu metode yang relatif baru untuk melakukan prediksi pada kasus regresi atau klasifikasi. SVM adalah model *supervised learning* yang implementasinya membutuhkan tahap pelatihan menggunakan *sequential training* SVM dan diikuti proses pengujian [17]. Klasifikasi SVM mencoba memisah ruang data menggunakan klasifikasi nonlinier atau linier antara kelas yang berbeda [18]. Konsep klasifikasi SVM adalah sebagai *hyperplane* yang berperan menjadi pemisah dua kelas data (hal positif dan hal negatif) [19]. Permulaan tahap SVM yakni mengonversi data teks ke bentuk data vektor (vektor dalam penelitian ini yaitu bobot) lalu digabungkan TF-IDF untuk pembobotan [20]. SVM mempunyai kelebihan yaitu dapat menyelesaikan permasalahan *over-fitting*, solusi *optimal local*, memiliki rasio konvergensi rendah, dan memiliki kemampuan tinggi generalisasi dalam kasus *minor sample*. Gbr. 1 menunjukkan model *flowchart* SVM.

E. Particle Swarm Optimization (PSO)

PSO adalah suatu metode optimasi paling sederhana untuk memodifikasi beberapa parameter. Optimasi pada PSO dapat dilakukan dengan cara menyeleksi atribut (*attribute selection*) dan *feature selection*, serta meningkatkan bobot atribut (*attribute weight*) pada semua atribut atau variabel yang



Gbr. 1 *Flowchart* SVM [21].

digunakan [9]. PSO terinspirasi oleh tabiat sekelompok burung yang terbang bergerombol atau ikan yang berenang bergerombol. Ratusan burung atau ratusan ikan dapat bergerak dengan cepat tanpa saling bertabrakan, padahal jarak mereka sangat dekat. Kelebihan PSO adalah sederhana, mudah diterapkan, dan kecepatan konvergensinya. Data mentah atau dalam kalimat opini dikonversi ke dalam numerik dan TF-IDF digunakan untuk mengubah data opini menjadi numerik.

Pencarian solusi pada algoritme PSO oleh populasi tertentu berdasarkan sejumlah partikel. Populasi dinilai secara acak dan memiliki batasan nilai terkecil dan terbesar. Partikel-partikel melacak solusi dengan melalui ruang pencarian dengan cara beradaptasi terhadap letak terbaiknya (*local best*) dan beradaptasi terhadap letak partikel terbaik pada seluruh kelompok (*global best*) sewaktu melalui *search space* [22]. Gbr. 2 menunjukkan cara PSO diterapkan.

F. Validasi

Validasi yaitu tahap mengevaluasi akurasi prediksi dari suatu model. *Bootstrap, stratified sampling*, *cross-validation, random sub-sampling,* dan *holdout* adalah beberapa metode validasi yang berfungsi untuk memvalidasi sebuah model yang bersumber pada data yang diperoleh [23]. *K-fold cross-validation* adalah metode validasi yang memisahkan data awal secara acak kedalam *k* bagian yang sama-sama terbagi atau "*fold"* [7]. Fungsi *k-fold* adalah supaya tidak ada data *overlapping* terhadap data *testing*. Validasi pada makalah ini menggunakan 10 *k-fold*, yaitu data dibagi menjadi sepuluh bagian dengan total yang sama, lalu dilakukan tahap validasi sebanyak sepuluh kali secara repetitif.

G. Evaluasi

Grafik ROC merupakan metode yang berfungsi mengilustrasikan, menentukan pengklasifikasi, dan

Gbr. 2 Proses penerapan PSO [21].

mengorganisasi menurut kemampuan atau dapat mengevaluasi akurasi pada klasifikasi secara visual. Grafik ROC merupakan alur dua indikator dengan skala *false positive* pada sumbu x (horizontal) dan *true positive* pada sumbu y (vertikal). Kurva ROC digunakan untuk mengukur nilai AUC [7]. Berikut adalah panduan untuk mengklasifikasikan akurasi pengujian menggunakan nilai AUC [23].

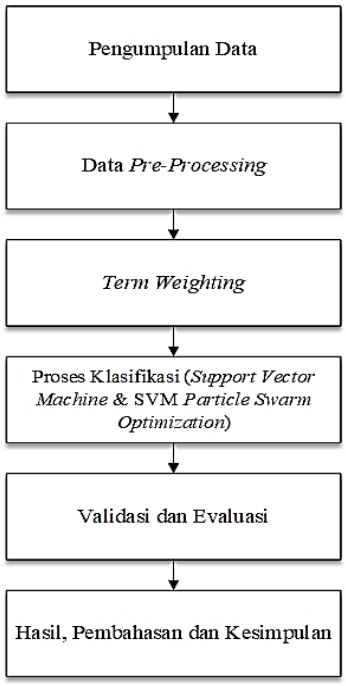
1. *Excellent classification* = 0,90 - 1,00
2. *Good classification* = 0,80 - 0,90
3. *Fair classification* = 0,70 - 0,80
4. *Poor classification* = 0,60 - 0,70
5. *Failure classification* = 0,50 - 0,60.

III. METODOLOGI

Makalah ini bertujuan mengetahui sentimen masyarakat tentang transportasi *online* dengan menggunakan metode klasifikasi SVM dan SVM optimasi PSO. Metode yang digunakan adalah metode kualitatif. Ahli bahasa yang sesuai dengan bahasa yang digunakan berperan untuk proses pelabelan data. Tujuannya agar ahli bahasa lebih paham menginterpretasikan maksud dari setiap teks untuk pelabelan positif atau negatif. Aplikasi RapidMiner melakukan proses *training* dan *testing* untuk menghitung akurasi dengan metode SVM dan SVM-PSO. Gbr. 3 menjelaskan tahapan penelitian dan Gbr. 4 menjelaskan model SVM-PSO.

1) Tahap Pertama: Tahap ini adalah pengumpulan data yang mengambil data *tweet* dari kata kunci "transportasi *online*" menggunakan aplikasi Octoparse dengan metode *scraping.* Data kemudian dibagi dua, menjadi data *training* dan data *testing.*

2) Tahap Kedua: Pada tahap ini dilakukan *pre-processing* data dengan melalui enam tahap, yaitu *cleansing*, *case folding,*



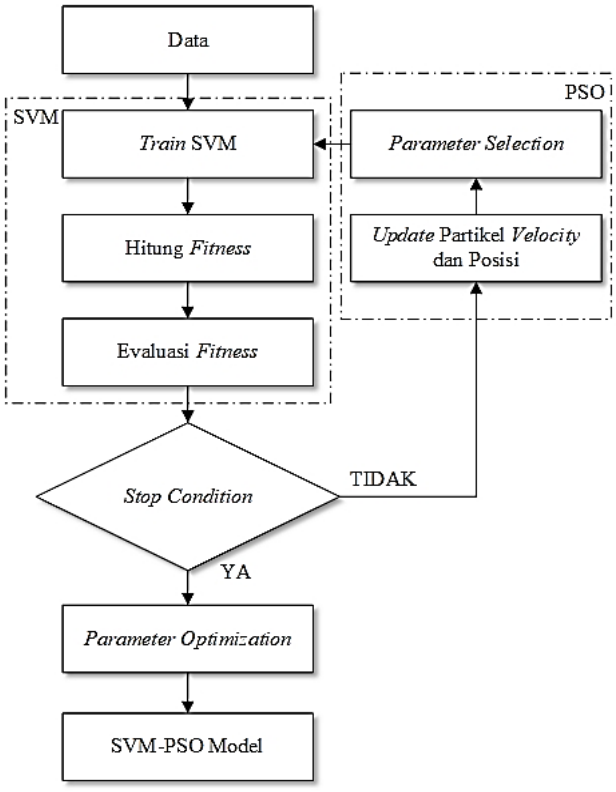
Gbr. 3 Tahapan penelitian.

tokenizing, *normalization*, *filtering*, dan *stemming*. *Pre-processing* dilakukan agar data dapat diklasifikasi dan untuk mempermudah dalam pemrosesan data.

3) Tahap Ketiga: Tahap ini melakukan pembobotan setiap kata atau *term weighting* yang digunakan untuk mengubah data opini menjadi data numerik. Makalah ini menggunakan TF-IDF, diimplementasikan menggunakan *tools* RapidMiner.

4) Tahap Keempat: Tahap ini adalah proses klasifikasi dengan menggunakan metode SVM dan SVM-PSO. Tujuannya adalah untuk membandingkan hasil sentimen positif dan negatif, akurasi, serta AUC dari metode klasifikasi SVM dan SVM-PSO. Usulan metode pada makalah ini adalah menggunakan data *training* dan *testing* pada proses klasifikasi karena banyaknya penelitian yang hanya menggunakan satu *dataset* yang langsung dibagi positif dan negatif secara manual. Data *training* digunakan untuk melatih algoritme sedangkan data *testing* digunakan untuk mengetahui kinerja algoritme yang sudah dilatih sebelumnya. Proses analisis menggunakan SVM dimulai dengan mengubah teks menjadi data vektor. Karena sulitnya menentukan parameter yang cocok dengan SVM, pada makalah ini PSO digunakan untuk mengatur parameter SVM untuk dioptimalkan. Namun, pada makalah ini setiap parameter pada SVM dan PSO diatur *default* dengan nilai 0. *Output* yang didapat adalah kelas positif dan negatif. *Stopping criteria* pada SVM adalah *max iterations* 100.000 (*default*) dan akan berhenti jika sampai pada angka tersebut, sedangkan pada PSO adalah *max evaluations* dengan nilai 500 (*default*) dan akan berhenti jika sampai pada angka tersebut.

5) Tahap Kelima: Tahap ini adalah proses validasi menggunakan 10 *k-fold cross validation* yang merupakan metode untuk menghapus *noise* atau *bias* kata, sehingga dapat menaikkan akurasi. Evaluasi pengukuran akurasi diukur menggunakan *confusion matrix* dengan hasil data dalam bentuk kurva ROC untuk mengukur nilai AUC.



Gbr. 4 Model SVM-PSO.

TABEL I
SENTIMEN DATA *TRAINING*

Positif	Negatif	Jumlah
374	348	722
51,80%	48,20%	100%

6) Tahap Keenam: Tahap terakhir adalah penulisan laporan penelitian dan memaparkan hasil penelitian, pembahasan, dan kesimpulan.

IV. HASIL DAN PEMBAHASAN

Proses analisis sentimen pada makalah ini dilakukan menggunakan *tools* RapidMiner versi 9.5.001. Data yang digunakan adalah data *tweet* dari hasil kata kunci "transportasi *online*" menggunakan aplikasi Octoparse versi 7.1 dengan metode *scraping* (*file .xlsx*) dengan total data 1.852 *tweet* dari 1 Januari 2019 hingga 15 Oktober 2019, yang dibagi menjadi data *testing* 1.130 *tweet* dan data *training* 722 *tweet*. Data *tweet* yang digunakan tidak menggunakan *tweet* dari akun berita dan telah dipilah secara manual. Data *training* dibagi secara manual sesuai kelasnya, yaitu sentimen positif atau sentimen negatif, yang diperlihatkan pada Tabel I.

Data *training* lalu melalui tahap *pre-processing*. Diambil contoh dua *tweet* yang dianggap dapat merepresentasikan *pre-processing*, terlihat pada Gbr. 5, dan diuraikan prosesnya pada Tabel II sampai Tabel V. Contoh implementasi tahap *pre-processing* pada *tweet* ditunjukkan pada Tabel II.

TABEL II
CONTOH HASIL *PRE-PROCESSING*

Tahap *Pre-Processing*	**Hasil**
Cleansing	T1: Dan juga pengguna seharusnya diberi pemilihan rute yang seharusnya bisa ditempuh dengan jarak dekat namun jika menggunakan transportasi online dipilih otomatis dengan rute jarak terjauh mengakibatkan tarif yang dibayarkan lebih mahal T2: gua pribadi merasa sangat terbantu dengan aplikasi transportasi online karena jadi tau ongkosnya berapa dan nyiapin uang dulu sebelum naik tanpa perlu takut kekurangan uang pas mau bayar ojek angkot
Case Folding	T1: dan juga pengguna seharusnya diberi pemilihan rute yang seharusnya bisa ditempuh dengan jarak dekat namun jika menggunakan transportasi online dipilih otomatis dengan rute jarak terjauh mengakibatkan tarif yang dibayarkan lebih mahal T2: gua pribadi merasa sangat terbantu dengan aplikasi transportasi online karena jadi tau ongkosnya berapa dan nyiapin uang dulu sebelum naik tanpa perlu takut kekurangan uang pas mau bayar ojek angkot
Tokenizing	T1: dan; juga; pengguna; seharusnya; diberi; pemilihan; rute; yang; seharusnya; bisa; ditempuh; dengan; jarak; dekat; namun; jika; menggunakan; transportasi; online; dipilih; otomatis; dengan; rute; jarak; terjauh; mengakibatkan; tarif; yang; dibayarkan; lebih; mahal; T2: gua; pribadi; merasa; sangat; terbantu; dengan; aplikasi; transportasi; online; karena; jadi; tau; ongkosnya; berapa; dan; nyiapin; uang; dulu; sebelum; naik; tanpa; perlu; takut; kekurangan; uang; pas; mau; bayar; ojek; angkot;
Normalization	T1: dan; juga; pengguna; seharusnya; diberi; pemilihan; rute; yang; seharusnya; bisa; ditempuh; dengan; jarak; dekat; namun; jika; menggunakan; transportasi; online; dipilih; otomatis; dengan; rute; jarak; terjauh; mengakibatkan; tarif; yang; dibayarkan; lebih; mahal; T2: saya; pribadi; merasa; sangat; terbantu; dengan; aplikasi; transportasi; online; karena; jadi; tau; biayanya; berapa; dan; siapkan; uang; dulu; sebelum; naik; tanpa; perlu; takut; kekurangan; uang; pas; mau; bayar; ojek; angkot;
Filtering	T1: pengguna; pemilihan; rute; tempuh; jarak; transportasi; online; pilih; otomatis; rute; jarak; mengakibat; tarif; bayar; mahal; T2: saya; pribadi; terbantu; aplikasi; transportasi; online; biayanya; uang; takut; uang; pas; bayar; ojek; angkot;
Stemming	T1: guna; pilih; rute; tempuh; jarak; transportasi; online; pilih; otomatis; rute; jarak; akibat; tarif; bayar; mahal; T2: saya; pribadi; bantu; aplikasi; transportasi; online; biaya; uang; takut; uang; pas; bayar; ojek; angkot;

Gbr. 5 Contoh *tweet*.

Pembobotan atau *term weighting* menurut TF-IDF dilakukan setelah proses *pre-processing*. Diambil sejumlah kata untuk contoh atribut pemilihan data: kelas positif seperti terima kasih, aman, murah, baik, dan normal; kelas negatif seperti mahal, susah, jahat, kasar, dan parah. Bobot nilai pada masing-masing *tweet* yang dihasilkan dari hasil pengolahan data teks pada atribut lalu dibandingkan setiap probabilitasnya menurut pembobotan TF-IDF, yang ditunjukkan pada Tabel III dan Tabel IV.

Hasil probabilitas setiap dokumen berdasarkan atribut dapat dilihat pada Tabel III dan Tabel IV. Hasil dari probabilitas atribut negatif dan atribut positif dibandingkan untuk mengetahui atribut yang memiliki probabilitas lebih besar, seperti yang tertera pada Tabel V. Apabila probabilitas opini positif melebihi opini negatif, maka dokumen tersebut merupakan opini positif, dan sebaliknya.

TABEL III
CONTOH ATRIBUT KELAS POSITIF PADA DATA *TRAINING*

Dokumen	**"terima kasih"**	**"aman"**	**"murah"**	**"baik"**	**"normal"**	**Prob.**
Negatif 256	0,000	0,000	0,000	0,000	0,000	0,000
Negatif 352	0,000	0,093	0,000	0,000	0,000	0,093
Negatif 681	0,000	0,000	0,000	0,000	0,000	0,000
Positif 209	0,000	0,106	0,000	0,129	0,000	0,235
Positif 285	0,157	0,120	0,000	0,000	0,000	0,277
Positif 566	0,000	0,163	0,163	0,000	0,000	0,326

A. Analisis Data Menggunakan SVM dan SVM-PSO

Pengujian metode klasifikasi SVM dan SVM-PSO menggunakan 10 *k-fold* atau nilai *fold default* pada *cross validation*, yaitu total data dibagi menjadi sepuluh bagian data secara acak dan prosesnya berulang sebanyak jumlah kelompok yang telah ditentukan. Pada Tabel VI ditunjukkan nilai akurasi SVM sebesar 95,46% dan SVM-PSO sebesar 96,04%. Perbedaan nilai akurasi sebesar 0,58% membuat metode SVM-PSO memperoleh akurasi yang lebih baik dari metode SVM. Pada nilai AUC, SVM dan SVM-PSO

TABEL IV
CONTOH ATRIBUT KELAS NEGATIF PADA DATA *TRAINING*

Dokumen	"mahal"	"susah"	"jahat"	"kasar"	"parah"	Prob.
Negatif 256	0,102	0,000	0,000	0,000	0,000	0,102
Negatif 352	0,000	0,000	0,125	0,000	0,000	0,125
Negatif 681	0,136	0,171	0,000	0,000	0,000	0,307
Positif 209	0,000	0,000	0,000	0,000	0,000	0,000
Positif 285	0,108	0,000	0,000	0,000	0,000	0,108
Positif 566	0,000	0,000	0,000	0,000	0,000	0,000

TABEL V
CONTOH PENENTUAN KELAS SENTIMEN

Dokumen	Prob. Positif	Prob. Negatif	*Class*
Negatif 256	0,000	0,102	Negatif
Negatif 352	0,093	0,125	Negatif
Negatif 681	0,000	0,307	Negatif
Positif 209	0,235	0,000	Positif
Positif 285	0,277	0,108	Positif
Positif 566	0,326	0,000	Positif

TABEL VI
PERSENTASE HASIL PENGUJIAN DATA MENGGUNAKAN SVM DAN SVM-PSO

Algoritme Klasifikasi	*K-Fold*	Akurasi	AUC
SVM	10	95,46 %	0,979
SVM-PSO	10	96,04 %	0,993

TABEL VII
MODEL *CONFUSION MATRIX* UNTUK SVM DENGAN 10 *FOLD*

Akurasi: 95,46 % +/- 14,02 % (*micro average*: 99,36 %)			
	True Negative	***True Positive***	***Class Precision***
Pred. Negative	2.833	3	99,89 %
Pred. Positive	35	3.049	98,87 %
Class Recall	98,78 %	99,90 %	

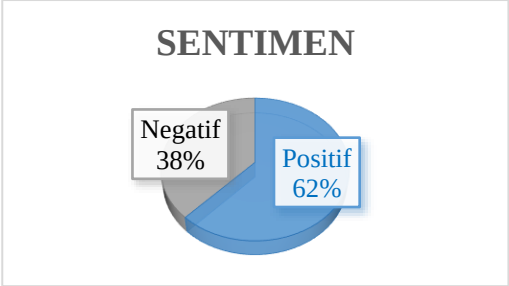
menghasilkan nilai 0,979 dan 0,993, yang keduanya temasuk dalam *excellent classification*. Perbedaan nilai 0,014 membuat SVM-PSO lebih baik dari segi nilai AUC daripada SVM. Akurasi pada Tabel VI membuktikan bahwa meskipun parameter di atur *default*, PSO dapat menaikkan akurasi SVM. Hal ini dikarenakan PSO bekerja mencari nilai parameter terbaik dengan cara beradaptasi terhadap *local best* dan beradaptasi terhadap letak partikel terbaik pada seluruh kelompok (*global best*).

B. Analisis Confusion Matrix dan Kurva ROC Menggunakan SVM dan SVM-PSO

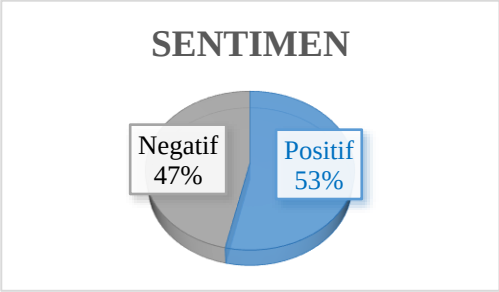
Tabel VII dan Tabel VIII menunjukkan evaluasi metode menggunakan *confusion matrix* dengan 10 *fold*. *True Positive* (TP) *rate* adalah persentase dari *class positive* yang berhasil

TABEL VIII
MODEL *CONFUSION MATRIX* UNTUK SVM-PSO DENGAN 10 *FOLD*

Akurasi: 96.04 % +/- 0.67 % (*micro average*: 96.04 %)			
	True Negative	***True Positive***	***Class Precision***
Pred. Negative	3.033	158	95,05 %
Pred. Positive	99	3.208	97,01 %
Class Recall	96,84 %	95,31 %	



Gbr. 8 Hasil analisis sentimen dengan SVM.



Gbr. 9 Hasil analisis sentimen dengan SVM-PSO.

diklasifikasi sebagai *class positive*, sedangkan *True Negative* (TN) *rate* adalah persentase dari *class negative* yang berhasil diklasifikasi sebagai *class negative*. *False Positive* (FP) *rate* adalah *class negative* yang diklasifikasi sebagai *class positive*. *False Negative* (FN) *rate* adalah *class positive* yang diklasifikasi sebagai *class negative*. TP dari metode SVM adalah 3.049, TN sebesar 2.833, FP sebesar 3, dan FN adalah 35. TP dari metode SVM-PSO adalah 3.208, TN sebesar 3.033, FP sebesar 158, dan FN adalah 99.

Gbr. 6 dan Gbr. 7 pada bagian akhir makalah merupakan kurva ROC untuk menilai hasil prediksi dari algoritme SVM dan SVM-PSO dengan nilai 0,979 dan 0,993, yang termasuk dalam *excellent classification*. Kurva ROC mengekspresikan *confusion matrix*, yaitu sumbu x merupakan *false positive* dan sumbu y adalah *true positive*.

C. Analisis Sentimen Menggunakan SVM dan SVM-PSO

Gbr. 8 menunjukkan hasil analisis sentimen menggunakan SVM. Sentimen positif mendominasi pada data *testing* dengan total data 702 atau 62%, sedangkan sentimen negatif dengan total data 428 atau 38%.

Gbr. 9 menunjukkan hasil analisis sentimen menggunakan SVM-PSO. Sentimen positif dan negatif hanya berbeda 6%. Sentimen positif pada data *testing* dengan jumlah data 604 atau 53%, sedangkan sentimen negatif dengan jumlah data 526 atau 47%.

Penggunaan data *training* dan data *testing* berhasil dilakukan dengan melihat data yang ada. Hasil menunjukkan bahwa SVM-PSO lebih baik dari segi akurasi maupun nilai AUC, juga dalam hal analisis sentimen dengan nilai positif sebesar 53%. Peningkatan nilai akurasi SVM-PSO juga membuktikan bahwa nilai parameter yang diatur *default* dapat meningkatkan akurasi. Hasil ini menunjukkan SVM yang telah dioptimasi menggunakan PSO lebih baik dari SVM biasa.

V. KESIMPULAN

Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa analisis sentimen transportasi *online* yang bertujuan untuk menganalisis sentimen dan meningkatkan akurasi metode dapat dilakukan dengan metode klasifikasi SVM dan SVM-PSO. Hasil menunjukkan bahwa sentimen masyarakat terhadap transportasi *online* di Indonesia adalah positif dengan hasil opini positif pada metode klasifikasi SVM sebesar 62% dan negatif sebesar 38%, sedangkan pada metode klasifikasi SVM-PSO opini positif sebesar 53% dan negatif sebesar 47%. Faktor-faktor yang menyebabkan sentimen positif antara lain adalah fleksibel, banyaknya diskon yang membuat ongkos lebih murah, dan mempermudah aktivitas. Hasil sentimen negatif yang tidak lebih dari 47% membuktikan bahwa transportasi *online* masih dapat diterima masyarakat dengan adanya masalah-masalah yang ada di sekitar. Hal ini juga dipengaruhi dengan adanya akun Twitter dari beberapa layanan transportasi *online* yang melayani keluhan pengguna dengan cara *mention* ke akun tersebut atau mengirim *direct message*.

Hasil pengujian menggunakan 10 *k-fold cross validation* mendapatkan akurasi dan nilai AUC pada metode klasifikasi SVM sebesar 95,46% dan AUC 0,979 (*excellent classification)*, sedangkan pada metode SVM-PSO akurasi sebesar 96,04% dan AUC sebesar 0,993 (*excellent classification)*. Dengan hasil dan data penelitian yang didapat, terlihat bahwa penggunaan data *training* dan data *testing* pada pengujian ini dapat dilakukan. Hasil komparasi metode klasifikasi analisis sentimen pada masalah ini membuktikan bahwa metode SVM yang sudah dioptimasi menggunakan PSO lebih baik daripada SVM biasa, meskipun menggunakan nilai parameter yang diatur *default*.

REFERENSI

[1] T. Abadi (2018) Catatan Perlindungan Konsumen 2018 (edisi 1): Ekonomi Digital dan Marginalisasi Hak Konsumen, [Online], https://ylki.or.id/2018/12/catatan-perlindungan-konsumen-2018-edisi-1-ekonomi-digital-dan-marginalisasi-hak-konsumen/, tanggal akses: 20-Okt-2019.

[2] P. Arhando (2019) Jika Tarif Ojek Online Naik, 71 Persen Konsumen Kembali Gunakan Kendaraan Pribadi, [Online], https://www.moneysmart.id/kenaikan-tarif-ojek-online-membuat-konsumer-terancam-beralih/, tanggal akses: 20-Okt-2019.

[3] A. Pradyant, (2017) Melihat Transportasi Umum Online dan Konvensional dari Kedua Sisi, [Online], https://www.kompasiana.com/tetikusliterasi/5a082942a4b06842cf3fe392/melihat-transportasi-umum-online-dan-konvensional-dari-kedua-sisi?page=all, tanggal akses: 20-Okt-2019.

[4] (2015) Indonesia: number of Twitter users 2014-2019. [Online], https://www.statista.com/statistics/490548/twitter-users-indonesia/, tanggal akses: 20-Okt-2019.

[5] M. Kanakaraj dan R.M.R. Guddeti, "Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis Using NLP Techniques," *Proc. 2015 IEEE 9th Int. Conf. Semantic Computing*, 2015, hal. 169-170.

[6] P.K. Gajakosh, T. Ghorpade, dan R. Shedge, "Opinion Mining for Multi-Mix Languages Hotel Review by using Fuzzy Sets," *Proc. of Int. Conf. on Advances in Science and Technology (ICAST)*, 2015, hal. 1-4.

[7] E. Indrayuni, "Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization," *Jurnal Evolusi*, Vol. 4, No.2, hal. 20-27, 2016.

[8] H. Tuhuteru dan A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *Jurnal Pengembangan IT*, Vol. 3, No. 3, hal. 394-401, 2018.

[9] O. Somantri dan C. Supriyanto, "Algoritme Genetika untuk Peningkatan Prediksi Kebutuhan Permintaan Energi Listrik," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, Vol. 5, No. 2, hal. 108-114, 2016.

[10] B. Liu, *Sentiment Analysis and Opinon Mining*, California, USA: Morgan & Claypool Publishers, 2012.

[11] A. D'Andrea, F. Ferri, P. Grifoni, dan T. Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation," *International Journal of Computer Applications*, Vol. 125, No. 3, hal. 26-33, 2015.

[12] S.B. Bhonde dan J.R. Prasad, "Sentiment Analysis - Methods, Applications and Challenges," *International Journal of Electronics Communication and Computer Engineering*, Vol. 6, No. 6, hal. 634-640, 2015.

[13] M. Fernández-Gavilanes, T. Álvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, dan F.J. González-Castaño, "Unsupervised Method for Sentiment Analysis in Online Texts," *Expert Systems with Applications*, Vol. 58, hal. 57-75, 2016.

[14] I. Zulfa dan E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network," *Indonesian Journal of Computing and Cybernetics Systems*, Vol. 11, No. 2, hal. 187-198, 2017.

[15] Z.H. Deng, K.H. Luo, dan H.L. Yu, "A Study of Supervised Term Weighting Scheme for Sentiment Analysis," *Expert Systems with Applications*, Vol. 42, No. 7, hal. 3506-3513, 2014.

[16] M. Lan, C.L. Tan, J. Su, dan Y. Lu, "Supervised and Traditional Term Weighting Methods for Automatic Text Categorization," *Pattern Analysis and Machine Intelligence*, Vol. 31, No. 4, hal. 721-735, 2009.

[17] B. Santosa, "Tutorial Support Vector Machine," Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, Tutorial, hal. 1-19, 2015.

[18] R. Feldman dan J. Sanger, *The Text Mining Handbook*, New York, USA: Cambridge University Press, 2007.

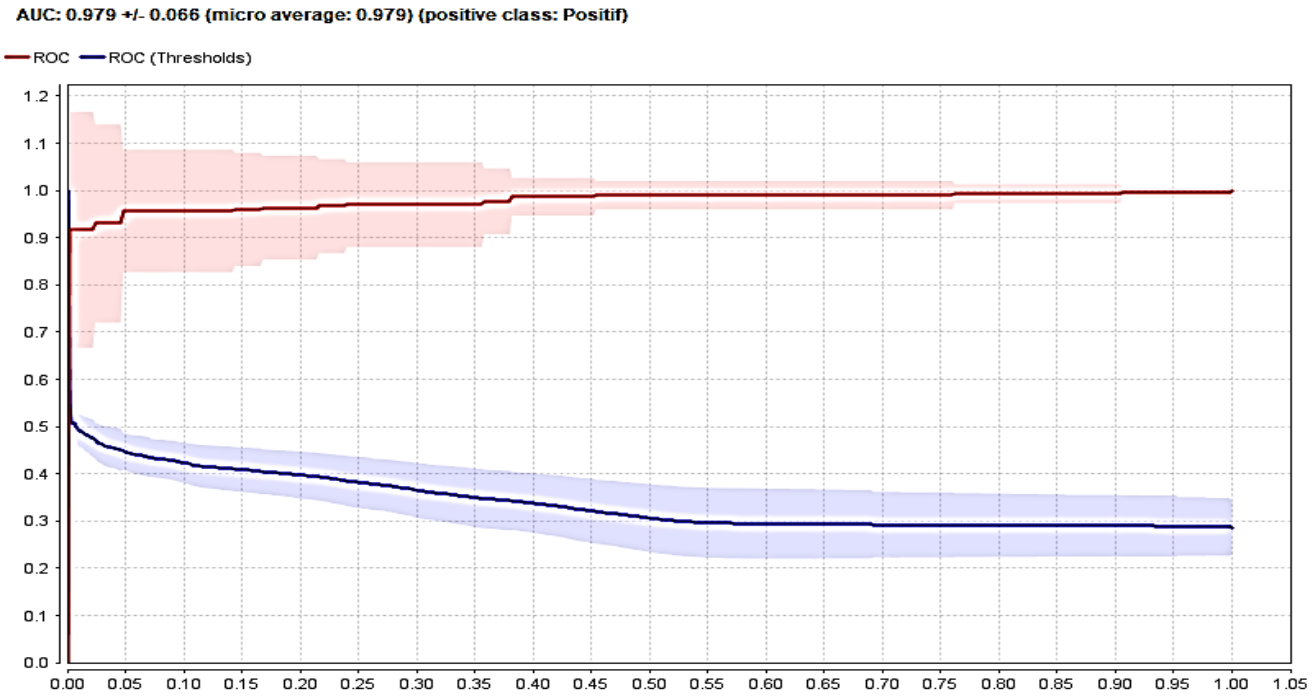
[19] W. He, S. Zha, dan L. Li, "Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry," *International Journal of Information Management*, Vol. 33, No.3, hal. 464-472, 2013.

[20] N. Yunita, "Analisis Sentimen Berita Artis dengan Menggunakan Algoritma Support Vector Machine dan Particle Swarm Optimization," *Jurnal Sistem Informasi*, Vol. 5, No. 2, hal. 104-112, 2017.

[21] S.S. Istia dan H.D. Purnomo, "Sentiment Analysis of Law Enforcement Performance Using Support Vector Machine dan K-Nearest Neighbor," *3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, 2018, hal. 84-89.

[22] T.B. Sasongko, "Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA)," *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 2, No. 2, hal. 244-253, 2016.

[23] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, Berlin: Germany: Springer-Verlag Berlin Heidelberg, 2011.



Gbr. 6 Kurva ROC SVM dengan 10 *fold.*

Gbr. 7 Kurva ROC SVM-PSO dengan 10 *fold*.