

# ANALISIS SENTIMEN BERITA ARTIS DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN PARTICLE SWARM OPTIMIZATION

Norma Yunita

**Abstract** - Media to get information about the artist is very much like the media tv, radio, newspapers, magazines, websites and more. But from most of the media website is a medium that is very flexible because it can be accessed in a variety of places that are connected to the internet, the information provided is very up to date and also anybody can comment on related articles. Classification techniques from some of the most frequently used is the Support Vector Machine (SVM). SVM has the advantage that is able to identify separate hyperplane that maximizes the margin between two different classes. But the SVM have disadvantages against the problem of election parameters or features accordingly. So that it can influence the accuracy results. Therefore, in this study using the Merge method of the selection of features, i.e. Particle Swarm Optimization to improve accuracy at the Support Vector Machine. As for the accuracy of the resulting algorithm Support Vector Machine-based Particle Swarm Optimmmization with 76.00% accuracy.

**Intisari** — Media untuk mendapatkan informasi tentang artis sangat banyak seperti media tv, radio, koran, majalah, website dan lain-lain. Tetapi dari sebagian besar media tersebut website merupakan media yang sangat fleksibel karena dapat di akses di berbagai macam tempat yang terkoneksi jaringan internet, informasi yang disediakan sangatlah up to date dan juga setiap orang bisa mengomentari artikel yang terkait. Dari beberapa teknik klasifikasi yang paling sering digunakan adalah Support Vector Machine (SVM). SVM memiliki kelebihan yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda. Tetapi SVM memiliki kekurangan terhadap masalah pemilihan parameter atau fitur yang sesuai. Sehingga dapat mempengaruhi hasil akurasi. Oleh karena itu, dalam penelitian ini menggunakan penggabungan metode pemilihan fitur, yaitu Particle Swarm Optimization (PSO) agar bisa meningkatkan akurasi pada Support Vector Machine. Adapun akurasi yang dihasilkan pada algoritma Support Vector Machine berbasis Particle Swarm Optimmmization dengan akurasi 76.00%.

**Kata Kunci:** Website, Classification, Support Vector Machine, Particle Swarm Optimization

## I. PENDAHULUAN

Di masa saat ini informasi sangat mudah diperoleh terutama informasi atau kabar tentang artis tertentu, banyak pembaca yang rela berjam-jam membaca untuk mendapatkan informasi terkini dari artis tertentu. Media untuk mendapatkan informasi

tentang artis sangat banyak seperti melalui media tv, radio, koran, majalah, website dan lain-lain. Tetapi dari sebagian besar media tersebut website merupakan media yang sangat fleksible karena dapat di akses di berbagai macam tempat yang terkoneksi jaringan internet, informasi yang disediakan sangatlah up to date dan juga setiap orang bisa mengomentari artikel yang terkait. Informasi yang menyebar sangat cepat dan diiringi dengan kebebasan mengeluarkan pendapat dapat menimbulkan berbagai jenis opini, baik opini negatif atau positif. Opini negatif disini mengandung arti bahwa kata atau opini yang dapat menimbulkan permusuhan, penghinaan, perdebatan dan perselisihan di dunia maya. Sedangkan opini positif yaitu kata atau opini yang sifatnya positif dan tidak menimbulkan permusuhan, penghinaan, perdebatan, dan perselisihan di dunia maya.

Seseorang yang terlalu mencintai artis akan berkomentar positif pada artis tertentu dan seseorang yang lain yang kurang menyukai artis tersebut akan berkomentar negatif dan hal ini akan berdampak tindakan pidana sebagaimana di atur dalam UU ITE tahun 2008 Pasal 27 ayat (3) yaitu setiap orang dengan sengaja dan tanpa hak mendistribusikan dan/atau mentransmisikan dan/atau membuat dapat diaksesnya informasi elektronik dan/atau dokumen elektronik yang bermuatan penghinaan dan/atau pencemaran nama baik. Seseorang yang terbukti dengan sengaja menyebarkan informasi elektronik yang bermuatan pencemaran nama baik seperti yang dimaksudkan dalam Pasal 27 ayat (3) UU ITE akan dijera dengan Pasal 45 Ayat (1) UU ITE, sanksi pidana penjara maksimum 6 tahun dan/atau denda maksimum 1 Milyar Rupiah. Maka dari itu diperlukan suatu sistem yang dapat memfilter atau menyaring kata-kata yang tidak seharusnya di postingkan.

“Meluasnya penggunaan internet telah meningkatkan jumlah informasi yang disimpan dan diakses melalui web dalam kecepatan yang sangat cepat” [19]. Beberapa tahun terakhir, pengguna internet telah berkembang sangat pesat. “Banyak forum, blog, jejaring sosial, situs web e-commerce, dan laporan berita berfungsi sebagai bentuk untuk mengekspresikan pendapat, yang dapat dimanfaatkan untuk memahami pendapat masyarakat umum dan konsumen pada peristiwa sosial, politik, strategi perusahaan, preferensi produk, dan reputasi pemantauan” [15], “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” [20].

Berdasarkan penjelasan di atas, analisis tersebut disebut sentiment analysis, yang secara umum dapat didefinisikan “sebagai studi komputasional dari opini-opini orang, sentimen

dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks” [14]. “Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral” [17].

“Analisis sentimen yang banyak diteliti yaitu analisis sentimen terhadap konten teks di facebook dan twitter” [10], “sentimen analisis terhadap social media” [7], “sentimen analisis terhadap dokumen teks” [1], “sentimen analisis terhadap konten berita” [12], sedangkan penulis akan melakukan penelitian terhadap sentimen analisis berita artis. “Machine Learning yang memperkenalkan klasifikasi teks seperti Naive Bayes, K-NN, SVM dan Rocchio Classification” [18]. Naive Bayes. “Eksperimental serta evaluasi menunjukkan bahwa SVM, KNN dan NB merupakan tradisional teks klasifikasi. Eksperimen dan evaluasi menunjukkan teks klasifikasi yang valid” [22]. Untuk itulah penelitian ini menggunakan metode Support Vector Machine (SVM) untuk klasifikasi teks.

## II. KAJIAN LITERATUR

### 1. Analisa Sentimen (Sentiment Analysis)

“Sentiment Analysis atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan text mining yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk layanan, organisasi, individu, ataupun kegiatan tertentu” [13]. “Tujuan dari analisa sentimen adalah untuk menentukan perilaku atau opini dari seorang penulis dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan” [2]. Sentiment analysis juga dapat menyatakan perasaan emosional sedih, gembira, atau marah.

Langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen adalah [16]:

#### a. Definisikan 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.

#### f. Interpretation/Evaluation

Tahap evaluasi biasanya menghitung akurasi, recall, precision, dan F-1.

### 2. Text Mining

“Text mining atau text analytics adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan data mining dimana data mining mengolah data yang sifatnya terstruktur. Pada dasarnya, text mining merupakan bidang interdisiplin yang mengacu pada perolehan informasi (information retrieval), data mining, pembelajaran mesin (machine learning), statistik, dan komputasi linguistik” [9].

“Text mining umumnya mencakup kategorisasi informasi atau teks, mengelompokkan teks, ekstraksi entitas atau konsep, pengembangan dan perumusan taksonomi umum. Text mining berkenaan dengan informasi terstruktur atau tekstual ekstraksi informasi yang bermakna dan pengetahuan dari jumlah besar teks” [8].

“Text mining adalah penambahan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda” [6].

“Text mining merupakan teknik yang digunakan untuk menangani masalah klasifikasi, clustering, information extraction dan information retrieval” [4].

“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” [21].

Dari ke lima pendapat ahli diatas, maka dapat disimpulkan bahwa text mining adalah informasi terstruktur yang digunakan untuk menganalisis atau mengelompokkan dokumen atau teks dari sejumlah besar dokumen atau teks.

### 3. Review Berita Artis

Ulasan atau review yang terdapat di internet sangat banyak namun tidak diolah menjadi sebuah informasi yang bermanfaat. Sentimen publik bisa dijadikan sebagai indikator

untuk melihat apakah berita tersebut berkualitas atau tidak. Media sosial merupakan media yang sering digunakan untuk menuangkan sentimen atau opini publik mengenai berita tersebut. Situs berita TRIBUNnews.com dikelola PT. Tribun

Digital Online, Divisi Koran Daerah Kompas Gramedia (Group of Regional Newspaper). Berkantor pusat di Jakarta, situs berita ini menyajikan berita-berita nasional, regional, internasional, olahraga, ekonomi dan bisnis, serta seleb dan lifestyle. TRIBUNnews.com juga mengelola forum diskusi, dan komunitas online melalui Facebook, Twitter, serta Google+. Selain didukung reporter yang bertugas di Jakarta, TRIBUNnews.com didukung tidak saja oleh jaringan 28 koran daerah atau Tribun Network, tapi juga didukung oleh hampir 500 wartawan di 28 kota penting di Indonesia. Situs berita TRIBUNnews.com merupakan induk bagi lebih dari 20 laman daerah Tribun Network ([www.tribunnews.com](http://www.tribunnews.com)).

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

“SVM merupakan algoritma klasifikasi yang memiliki tujuan untuk menemukan fungsi pemisah (hyperplane) dengan margin paling besar, sehingga dapat memisahkan dua kumpulan data secara optimal” [9].

“SVM pada awalnya digunakan untuk klasifikasi data numerik, tetapi ternyata SVM juga sangat efektif dan cepat untuk menyelesaikan masalah-masalah data teks. Data teks cocok untuk dilakukan klasifikasi dengan algoritma SVM karena sifat dasar teks yang cenderung mempunyai dimensi yang tinggi, dimana terdapat beberapa fitur yang tidak relevan, tetapi akan cenderung berkorelasi satu sama lain dan umumnya akan disusun dalam kategori yang terpisah secara linear” [23].

“SVM memiliki kelebihan yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda” [5]. “Namun SVM memiliki kekurangan terhadap masalah pemilihan parameter atau fitur yang sesuai” [3]. “Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi” [24].

#### 5. Tinjauan Studi Penelitian Terkait

Berikut merupakan ringkasan dari penelitian terkait yang dijadikan peneliti sebagai panduan untuk penelitian ini:

TABEL 1. TINJAUAN STUDI PENELITIAN TERKAIT

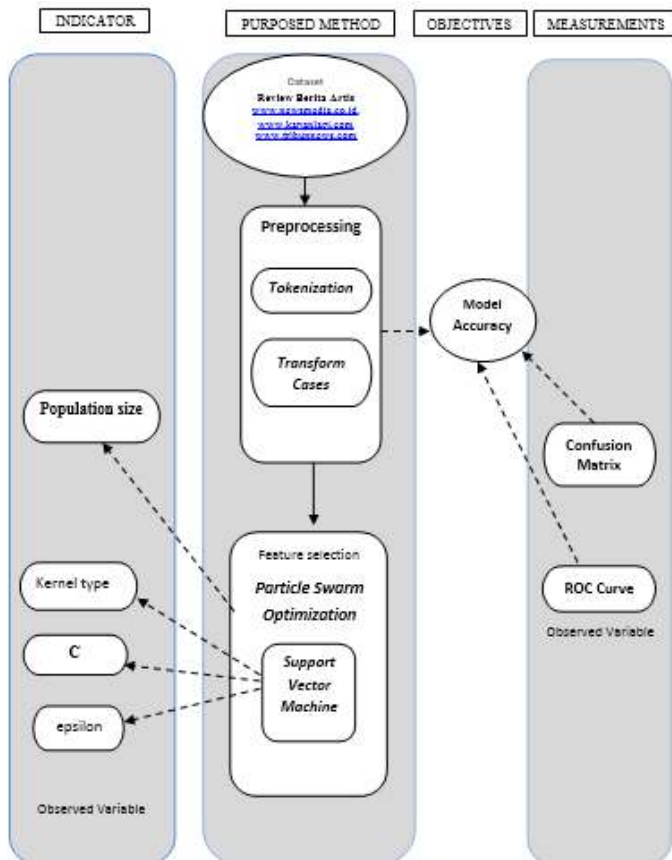
Judul	Peneliti	Classifier and Feature Selection	Akurasi	Hasil
An advanced multi class instance selection based support vector machine for text classification	Ramesh	AMCISSVM	67.1224% (Glas), 78.9042% (Diabetes), 94.7869% (Lonosphere)	Peneliti menghasilkan bahwa AMCISSVM memiliki kerja yang optimal dan akurasi yang tinggi
The role of text pre-processing in sentiment analysis	Haddi	SVM, (TF-IDF, FF, FP)	78.33%(TF-IDF), 76.33%(FF), 82.7%(FP)	Peneliti menghasilkan pengujian data dengan SVM +(TF-IDF, FF, FP) memiliki akurasi tertinggi
Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization	Basari et al	Hybrid SVM PSO	77%	Peneliti menggunakan PSO sebagai optimasi dengan 10 Fold-Cross Validation

More than words: social networks' text mining for consumer brand sentiments	Mostafa	Sentiment Analysis	SP (N 46%, P 54%), SC (N 29, P 34)	Peneliti ini menggunakan random sample 3516 tweets untuk mengevaluasi sentimen konsumen terhadap merk-merk terkenal dan menggunakan metodologi kualitatif dan kuantitatif.
A text mining framework for advancing sustainability indicators	Rivera et al	SI, K-NN, SVM, NB	K-NN(85%), SVM(53%), NB(68%)	Peneliti menghasilkan pengujian data dengan sustainability indicators (SI) dengan menganalisis terstruktur digital artikel berita dengan metode text mining.
Analisis sentimen berita artis dengan menggunakan algoritma support vector machine dan particle swarm optimization	Peneliti	SVM PSO	?	Peneliti melakukan pengujian terhadap review berita artis pada SVM berbasis PSO

Sumber: (Hasil Penelitian, 2016)

#### 6. Kerangka Pemikiran

Berdasarkan pada latar belakang, maka penelitian ini metode Support Vector Machine dalam pengklasifikasian. Dari klasifikasi tersebut akan dioptimasi kembali oleh fitur seleksi PSO agar nilai akurasi yang didapat menjadi lebih optimal dan baik. Peneliti mengambil data dari sumber [www.newsmedia.co.id](http://www.newsmedia.co.id), [www.kapanlagi.com](http://www.kapanlagi.com) dan [www.tribunnews.com](http://www.tribunnews.com) yang terdiri dari beberapa review atau pendapat masyarakat mengenai artis. Peneliti mengambil sample data secara simple random sebanyak 150 review positif dan 150 review negatif. Sebelum data diklasifikasi, terlebih dahulu dilakukan preprocessing antara lain: Tokenization dan Transform Cases. Dalam pembobotan yang peneliti lakukan adalah Term Frequency Invers Document Frequency (TF-IDF) dan pemilihan seleksi fitur menggunakan Particle Swarm Optimization (PSO). Sedangkan klasifikasi yang digunakan adalah Support Vector Machine. Pengujian 10 fold cross validation akan dilakukan, akurasi algoritma diukur dengan Confusion Matrix dan hasil olahan akan ditampilkan dalam bentuk kurva ROC dan accuracy. Software yang digunakan untuk mengolah data klasifikasi adalah RapidMiner sebagai alat bantu dalam mengukur akurasi data eksperimen.



Gbr 1. Kerangka Pemikiran

### III. METODE PENELITIAN

#### 1. Perancangan Penelitian

Pada dasarnya, penelitian merupakan suatu investigasi yang terorganisasi, yang dilakukan untuk menyajikan suatu informasi dan memecahkan masalah. Metode penelitian yang digunakan penulis menggunakan metode penelitian eksperimen. Adapun metode penelitian yang penulis gunakan melalui beberapa tahapan sebagai berikut:

##### a. Pengumpulan Data

Data yang digunakan untuk melakukan eksperimen dikumpulkan melalui website newsmedia.co.id, kapanlagi.com dan tribunnews.com, kemudian data opini publik berita artis tersebut diseleksi dan dikumpulkan ke dalam notepad untuk diolah dalam pengujian data.

##### b. Pengolahan Data awal

Memilih metode yang akan digunakan pada saat pengujian data. Metode yang dipilih, berdasarkan penelitian yang terdahulu. Penulis menggunakan Metode Algoritma Support Vector Machine.

##### c. Metode yang Diusulkan

Metode yang diusulkan penulis ditambahkan optimasi agar dapat meningkatkan nilai akurasi. Optimasi yang digunakan yaitu Particle Swarm Optimization (PSO).

#### d. Eksperimen dan Pengujian Metode

Eksperimen yang dilakukan peneliti, menggunakan framework RapidMiner 6.4 untuk mengolah data sehingga menghasilkan nilai akurasi yang akurat dan untuk pengujian metode penulis membuat aplikasi menggunakan bahasa pemrograman PHP dan HTML.

#### e. 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).

### 2. Pengolahan Data Awal

Text Mining adalah suatu proses yang bertujuan untuk menemukan informasi atau tren terbaru yang sebelumnya tidak terungkap, dengan memproses dan menganalisa data dalam jumlah besar. "Dalam menganalisa sebagian atau keseluruhan unstructured text, text mining mencoba untuk mengasosiasikan satu bagian teks dengan yang lainnya berdasarkan aturan-aturan tertentu" [11]. Teks yang belum diolah biasanya memiliki karakteristik dimensi yang tinggi, terdapat noise pada data dan terdapat struktur teks yang tidak baik. Untuk itu, dalam pengolahan data awal, teks mining harus melalui beberapa tahapan yang disebut dengan preprocessing. Tahapan-tahapan tersebut yaitu:

#### a. Tokenization

Proses memotong setiap kata dalam teks dan mengubah huruf dalam dokumen menjadi huruf kecil. Hanya huruf yang diterima, sedangkan karakter khusus atau tanda baca akan dihilangkan. Jadi hasil dari proses tokenization adalah kata-kata yang merupakan penyusun kalimat atau string yang dimasukan tanpa ada tanda baca.

#### b. Transform Cases

Merubah seluruh huruf menjadi huruf kecil atau kapital semua.

### IV. HASIL DAN PEMBAHASAN

#### 1. Deskripsi Hasil Penelitian

Data training yang digunakan pada saat pengujian data diambil dari newsmedia.co.id, kapanlagi.com dan tribunnews.com. Pengujian data, dilakukan dengan menggunakan review berita artis (300 data training, yang terdiri dari 150 review negatif dan 150 review positif) kemudian dilakukan testing dan training dataset sehingga didapatkan accuracy dan AUC. Berikut akan dijelaskan lebih rinci mengenai hasil penelitian yang diperoleh.

Berikut merupakan tahapan-tahapan dalam melakukan pengolahan data yaitu:

##### a. Pengumpulan Data

Review berita artis, masing-masing dikelompokkan dengan cara disimpan ke dalam satu folder yaitu folder positif dan



folder negatif, kemudian tiap dokumennya diberikan ekstensi .txt sehingga dapat dibuka dengan aplikasi Notepad.

#### b. Pengolahan Data Awal (Preprocessing)

Berikut merupakan tahapan yang dilakukan dalam preprocessing:

##### a) Tokenization

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

TABEL 2. PERBANDINGAN TEKS SEBELUM DAN SESUDAH DILAKUKAN PROSES TOKENIZATION

<b>Teks sebelum dilakukan proses tokenization</b>	Nice posting.... Saya paling gedek sama (maaf) orang2 islam fanatik yg langsung main telan pemberitaan2 yg terjadi, lalu langsung di posting di account sosial FB, patt, dll....yg belum tentu kebenarannya, apalagi langsung di replay dengan berbagai macam ucapan selamat versi islami...(Mungkin itu typical orang islam indonesia yg fanatik kali yaa.....), giliran dikasi tau kebenarannya....nyolotnya minta ampun.....merasa paling benar sendiri
<b>Teks setelah dilakukan proses tokenization</b>	Nice posting Saya paling gedek sama maaf orang islam fanatik yg langsung main telan pemberitaan yg terjadi lalu langsung di posting di account sosial FB patt dll yg belum tentu kebenarannya apalagi langsung di replay dengan berbagai macam ucapan selamat versi islami Mungkin itu typical orang islam indonesia yg fanatik kali yaa giliran dikasi tau kebenarannya nyolotnya minta ampun merasa paling benar sendiri

##### b) Transform Cases

Dalam proses transform cases ini, semua huruf dirubah menjadi huruf kecil semua atau huruf kapital semua.

TABEL 3. PERBANDINGAN TEKS SEBELUM DAN SESUDAH DILAKUKAN PROSES TRANSFORM CASES

<b>Teks sebelum dilakukan proses transform cases</b>	Mau sejelek apapun ibu kandung kita...kita tetep gk boleh lupa aplgi smpai durhaka...toh dy yg mengandung dan melahirkan kita....penuh pengorbanan....
<b>Teks setelah dilakukan proses transform cases</b>	mau sejelek apapun ibu kandung kita kita tetep gk boleh lupa aplgi smpai durhaka toh dy yg mengandung dan melahirkan kita penuh pengorbanan

#### 2. Analisis Evaluasi Hasil dan Validasi Model

Validation adalah proses untuk mengevaluasi keakuratan prediksi dari model.

##### a. Support Vector Machine (SVM)

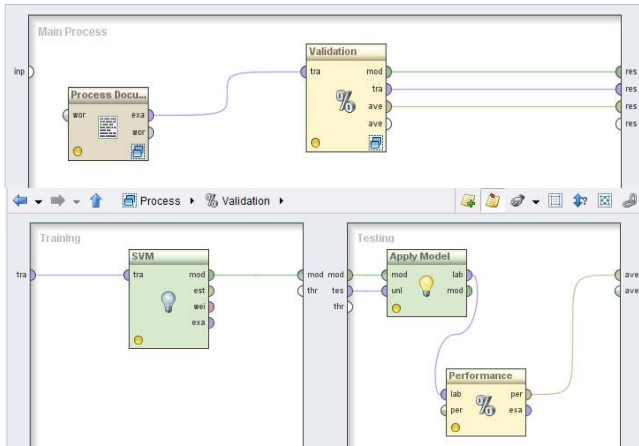
Nilai training cycles dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C, epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai training cycles.

TABEL 4. EKSPERIMEN PENENTUAN NILAI TRAINING CYCLES SVM

C	Epsilon	SVM	
		Accuracy	AUC
0.0	0.0	73.33%	0.774
0.1	0.1	71.00%	0.765
0.2	0.2	67.67%	0.739
0.3	0.3	71.33%	0.764
0.4	0.4	70.33%	0.764
0.5	0.5	73.33%	0.770
0.6	0.6	70.00%	0.768
0.7	0.7	71.33%	0.766
0.8	0.8	72.33%	0.762
0.9	0.9	71.33%	0.763
0.0	1.0	50.00%	0.500
1.0	1.0	50.00%	0.500
1.0	0.0	70.00%	0.762

Hasil pengujian menunjukkan bahwa penerapan metode Support Vector Machine pada Tabel IV.3 dengan C = 0.0 dan Epsilon E = 0.0 dihasilkan Accuracy= 73.33% dan AUC= 0.774.

Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan review berita artis negatif dan review opini publik berita artis positif menggunakan algoritma Support Vector Machine (SVM) pada framework RapidMiner dengan desain model berikut ini:



Gbr.2 Desain Model Validasi Support Vector Machine

## a) Confusion Matrix

Memberikan keputusan yang diperoleh dalam training dan testing, confusion matrix memberikan penilaian performance klasifikasi berdasarkan objek benar atau salah. Confusion matrix berisi informasi aktual (actual) dan prediksi (predicted) pada sistem klasifikasi.

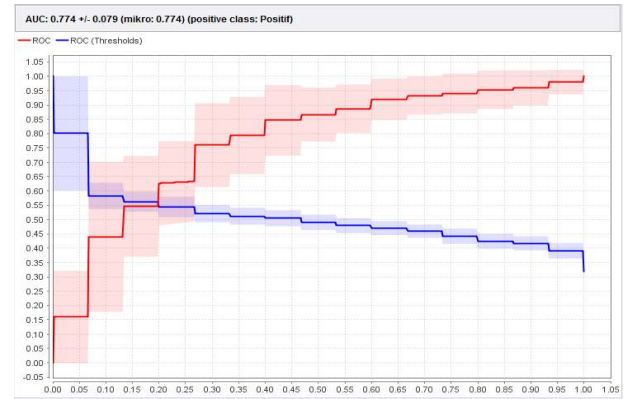
TABEL 5. CONFUSION MATRIX SUPPORT VECTOR MACHINE

Accuracy: 73.33% +/- 9.55% (mikro: 73.33%)			
	True Positif	True Negatif	Class Precision
Prediksi Negatif	91	21	81.25%
Prediksi Positif	59	129	68.62%
Class Recall	60.67%	86.00%	

$$\text{Acc (Accuracy)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{91 + 129}{91 + 21 + 59 + 129} = \frac{220}{300} = 0.73$$

## b) Kurva ROC

Kurva ROC (Receiver Operating Characteristic) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah pada sumbu X dan positif benar pada sumbu Y. Hasil perhitungan pada kurva ROC, menggambarkan kurva ROC untuk algoritma Support Vector Machine. Dapat disimpulkan bahwa satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Kurva ROC Support Vector Machine dengan nilai AUC (Area Under Curve) sebesar 0.774 dimana diagnosa hasilnya fair classification. Berikut dapat dilihat Kurva ROC Support Vector Machine pada Gambar 2.



Gbr.3. Kurva ROC Support Vector Machine

## b. Support Vector Machine (SVM) berbasis Particle Swarm Optimization (PSO)

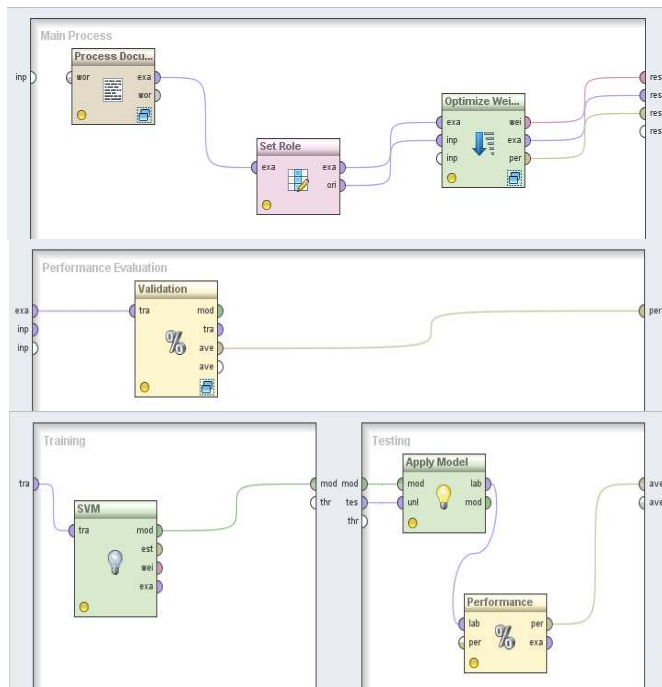
Nilai training cycles dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C, epsilon dan population size. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai training cycles.

TABEL 6. EKSPERIMEN PENENTUAN NILAI TRAINING CYCLES SVM BERBASIS PSO

C	Epsilon	SVM		Population Size	C	Epsilon	SVM + PSO	
		Accuracy	AUC				Accuracy	AUC
0.0	0.0	73.33%	0.774	5	0.0	0.0	76.00%	0.794
0.1	0.1	71.00%	0.765	5	0.1	0.1	72.67%	0.783
0.2	0.2	67.67%	0.739	5	0.2	0.2	73.00%	0.771
0.3	0.3	71.33%	0.764	5	0.3	0.3	74.67%	0.781
0.4	0.4	70.33%	0.764	5	0.4	0.4	75.00%	0.792
0.5	0.5	73.33%	0.770	5	0.5	0.5	75.00%	0.794
0.6	0.6	70.00%	0.768	5	0.6	0.6	73.67%	0.773
0.7	0.7	71.33%	0.766	5	0.7	0.7	75.00%	0.796
0.8	0.8	72.33%	0.762	5	0.8	0.8	75.67%	0.798
0.9	0.9	71.33%	0.763	5	0.9	0.9	72.67%	0.788
0.0	1.0	50.00%	0.500	5	0.0	1.0	50.00%	0.500
1.0	1.0	50.00%	0.500	5	1.0	1.0	50.00%	0.500
1.0	0.0	70.00%	0.762	5	1.0	0.0	73.33%	0.790

Hasil terbaik pada eksperimen SVM berbasis PSO di atas adalah C=0.0 dan Epsilon E=0.0 serta population size=5 yang dihasilkan accuracy=76.00% dan AUC=0.794. Hal ini menunjukkan bahwa dengan menggunakan optimasi Particle Swarm Optimization dapat meningkatkan akurasi yang lebih baik.

Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan review berita artis negatif dan review opini publik berita artis positif menggunakan algoritma Support Vector Machine (SVM) berbasis Particle Swarm Optimization (PSO) pada framework RapidMiner dengan desain model berikut ini:



Gbr.4 Model Pengujian Support Vector Machine berbasis Particle Swarm Optimization

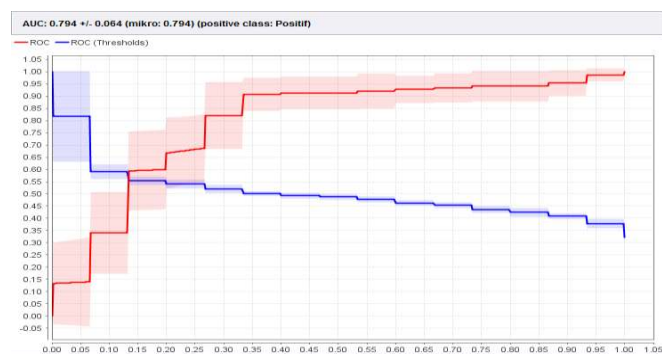
## a) Confusion Matrix

TABEL 7. CONFUSION MATRIX SUPPORT VECTOR MACHINE  
BERBASIS PARTICLE SWARM OPTIMIZATION

Accuracy: 76.00% +/- 4.67% (mikro: 76.00%)			
	True Positif	True Negatif	Class Precision
Prediksi Negatif	97	19	83.62%
Prediksi Positif	53	131	71.20%
Class Recall	64.67%	87.33%	

$$\text{Acc (Accuracy)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{97 + 131}{97 + 19 + 53 + 131} = \frac{228}{300} = 0.76$$

## b) Kurva ROC



Gbr.5 Kurva ROC Support Vector Machine berbasis Particle Swarm Optimization

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukkan bahwa ada peningkatan pada **akurasi** menggunakan Support Vector Machine berbasis Particle Swarm Optimization sebesar **76.00%** dan AUC sebesar **0.794**

## 3. User Interface

Berdasarkan rekomendasi hasil dari analisis sistem, maka analisis sistem harus memikirkan bagaimana membentuk dan mengembangkan sistem tersebut. Berikut ini adalah graphical user interface (GUI) yang dirancang untuk analisis sentiment berita artis berbasis website, yaitu:

Sumber: Hasil Penelitian (2016)

Gbr.6 Tampilan Preprocessing Tokenization

Sumber: Hasil Penelitian (2016)

Gbr.7 Tampilan Hasil Tokenization

Sumber: Hasil Penelitian (2016)

Gbr.8 Tampilan Transform Cases

Sumber: Hasil Penelitian (2016)

Gbr.9 Tampilan Hasil Transform Cases

## 4. Implikasi Penelitian

Adapun implementasi dari penelitian ini antara lain:

## a. Implikasi terhadap sistem

Berdasarkan hasil pengujian data, terbukti bahwa penerapan Particle Swarm Optimization terhadap seleksi fitur dapat meningkatkan akurasi pada algoritma Support Vector Machine.

## b. Implikasi terhadap aspek manajerial

Memudahkan pengembang sistem untuk mereview berita artis baik dari situs penyedia review atau ulasan produk tertentu maupun dari sosial media lainnya.

## c. Implikasi terhadap aspek penelitian lanjutan

Penelitian selanjutnya selain seleksi fitur, pada saat preprocessing juga bisa menggunakan berbagai teknik seperti n-gram, bi-gram dan seterusnya, serta dapat menggunakan dataset dari domain seperti film, perguruan tinggi, kamera dan sebagainya.

## V. KESIMPULAN

Berdasarkan pemaparan yang telah dijelaskan pada bab sebelumnya, penelitian ini menghasilkan akurasi dalam bentuk Confusion Matrix dan Kurva ROC. Adapun akurasi yang dihasilkan pada algoritma Support Vector Machine sebesar 73.33% dan AUC sebesar 0.774, sedangkan Support Vector Machine dan Particle Swarm Optimmmization dengan akurasi 76.00% dan AUC sebesar 0.794. Dengan demikian dapat disimpulkan bahwa penerapan optimasi dapat meningkatkan akurasi. Model di Support Vector Machine dan Particle Swarm Optimization dapat memberikan solusi terhadap permasalahan klasifikasi review berita artis agar lebih akurat dan optimal.

Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Menggunakan metode pemilihan fitur yang lain, seperti Chi Square, Infomation Gain, Mutual Information, dan lain-lain agar hasilnya bisa dibandingkan.
2. Menggunakan pengklasifikasi lain yang mungkin di luar Supervised learning. Sehingga bisa dilakukan penelitian yang berbeda dari umumnya yang sudah ada.
3. Menggunakan data review dari domain yang berbeda, misalnya review produk, review film, review saham dan lain sebagainya

## REFERENSI

- [1] Andini (2013). Klasifikasi Dokumen Text Menggunakan Algoritma Naive Bayes Dengan Bahasa Pemrograman Java. *Jurnal Teknologi Informasi & Pendidikan*. 2086-4981
- [2] Basari et al. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453-462.
- [3] Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J.(2013). Opinion Mining of Movie Review using HybridMethod of Support Vector Machine and Particle SwarmOptimization. *Procedia Engineering*, 53, 453-462.doi:10.1016/j.proeng.2013.02.059.
- [4] Berry, M.W. & Kogan, J. 2010. Text Mining Aplication and theory. WILEY : United Kingdom.
- [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. doi:10.1016/j.eswa.2013.12.035.
- [6] Feldman, Ronen and Sanger, James. 2007. The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, New York.Francisco: Diane Cerra.
- [7] Habernal, Ptáček, Steinberger. (2015), Reprint of “Supervised sentiment analysis in Czech social media”. *Information Processing & Management*, 50, 693-707.
- [8] Hashimi, Hussein, Alaaeldin Hafez, & Hassan Mathkour. (2014). Selection criteria for text mining approaches. *Computers in Human Behavior*. 729-733.
- [9] Jiawei, H., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques Third Edition*. Waltham, MA: Morgan Kaufmann.
- [10] Kaplan, A., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53, 59–68.
- [11] Kunaifi,Aang.(2009).Klasifikasi Email Berbahasa Indonesia Menggunakan Text Mining Dan Algoritma KMeans. Surabaya: Politeknik Elektronika Negeri Surabaya.
- [12] Kurniawan et al (2012), Klasifikasi Konten Berita Dengan Metode Text Mining. *Teknologi Informasi*. 14-19.
- [13] Liu, Bing. 2012. *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & ClaypoolPublisher.
- [14] Liu, H., Tian, H., Chen, C., & Li, Y. (2013). Electrical Power and Energy Systems An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization. *International Journal of Electrical Power & Energy Systems*, 52, 161–173.
- [15] M.R. Saleh, M.T. Martín-Valdivia, A. Montejó-Ráez, L.A. Ureña-López, Experiments with SVM to classify opinions in different domains, *Expert Syst. Appl.* 38 (2011) 14799–14804.
- [16] Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document-level sentiment.
- [17] Pang, B. & Lee, L. 2008. Subjectivity Detection and Opinion Identification. *Opinion Mining and Sentiment Analysis*. Now Publishers Inc. [Online]. Tersedia di: <http://www.cs.cornell.edu/home/llee/opinion-mining-sentiment-analysisurvey.html>.
- [18] Ramesh (2015). An Advanced Multi Class Instance Selection Based Support Vector Machine for Text Classification. *Procedia Computer Science*. 1124-1130.
- [19] Rocha, Leonardo et al (2013). Temporal contexts: effective text classification in evolving document collection. *Information Systems*. 388-409.
- [20] Rozi, Hadi, Achmad. (2012), Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS Vol. 6, No. 1, Juni 2012*. *Systems with Applications*, 40(2), 621–633. doi:10.1016/j.eswa.2012.07.059
- [21] Statsoft. (2015). Naive Bayes Classifier Introductory Overview. Retrieved April 22, 2015, from Statsoft Web Site: <http://www.statsoft.com/textbook/naivebayes-classifier>.
- [22] Yao, Zhi-Min. (2012), An Optimized NBC Approach in Text Classification. *Physics Procedia*, 24, 1910-1914



- [23] Zhai, C., & Aggarwal, C. C. (2012). Mining Text Data. New York: Springer.
- [24] Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. Expert Systems with Applications, 38(5), 5197–5204. doi:10.1016/j.eswa.2010.10.041.



Norma Yunita, M. Kom Tahun 2015 lulus dari Program Pascasarjana (S2) Program Studi Magister Ilmu Komputer Pascasarjana STMIK Nusa Mandiri Jakarta. Selain itu juga bekerja sebagai tenaga pengajar di AMIK BSI Pontianak.