

KOMPARASI METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE MENGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK ANALISIS SENTIMEN MOBIL ESEMKA

Kiki Setiawan¹, Beni Rahmatullah², Burhanuddin³, Atik Budi Paryanti⁴, Fariz Fauzi⁵

Magister Ilmu Komputer¹, Magister Ilmu Komputer², Magister Ilmu Komputer³, Manajemen⁴, Magister Ilmu Komputer⁵
STMIK Nusa Mandiri^{1,2,3}, Universitas Suryadarma⁴, STMIK Nusa Mandiri
ki2djoaz@gmail.com¹, info.benirahmatullah@gmail.com², 14002128@nusamandiri.ac.id³
atikbudiparyanti@gmail.com⁴, menyonais@gmail.com⁵

Abstrak

Esemka merupakan produk mobil hasil rakitan dari siswa Sekolah Menengah Kejuruan (SMK) yang bekerja sama dengan institusi dan industri dalam negeri, sejak awal kehadirannya, produk ini banyak menimbulkan pro dan kontra seperti yang ramai diperbincangkan di media sosial. Perlu dilakukan penelitian untuk melihat seberapa besar dampak dari kehadiran produk tersebut, diharapkan dari penelitian ini dapat bermanfaat dalam penelitian mengenai opini masyarakat lainnya. Peneliti memilih analisis sentimen karena merupakan tehnik yang tepat untuk pengolahan dataset. Dari 288 opini pengguna *youtube* yang berhasil dikumpulkan, didapat 180 opini positif dan 108 opini negatif. Kemudian, data tersebut diklasifikasikan menggunakan algoritma *Naive Bayes* dan *Support Vector Machine*, selanjutnya dilakukan optimasi dengan menggunakan *Particle Swarm Optimization*. Sehingga penelitian yang telah dilakukan mendapatkan hasil akurasi sebesar 75.04% untuk nilai akurasi *Naive Bayes* tanpa seleksi fitur, 83.33% untuk hasil *Naive Bayes* menggunakan seleksi fitur. Kemudian nilai akurasi yang didapat untuk *Support Vector Machine* tanpa seleksi fitur sebesar 78.81%, dan dengan fitur seleksi sebesar 88.19%. Ada peningkatan hasil akurasi yang signifikan, membuktikan bahwa penggunaan seleksi fitur *Particle Swarm Optimization* mempengaruhi untuk meningkatkan hasil akurasi.

Kata Kunci: Analisis Sentimen, *Naive Bayes*, *Support Vector Machine*, *Particle Swarm Optimization*.

Abstract: Esemka is an assembled car product from Vocational High School (SMK) students who work closely with domestic institutions and industries, since the beginning of its presence, this product raises many pros and cons as it is widely discussed on social media. Research needs to be done to see how big the impact of the presence of these products, it is hoped that this research can be useful in research on other public opinion. Researchers choose sentiment analysis because it is an appropriate technique for processing datasets. From 288 youtube user opinions that have been collected, obtained 180 positive opinions and 108 negative opinions. Then, the data is classified using the Naive Bayes algorithm and Support Vector Machine, then optimization is done using Particle Swarm Optimization. So that research has been done to get an accuracy of 75.04% for the accuracy value of Naive Bayes without feature selection, 83.33% for Naive Bayes results using feature selection. Then the accuracy value obtained for Support Vector Machine without feature selection is 78.81%, and with feature selection is 88.19%. There is a significant increase in accuracy results, proving that the use of Particle Swarm Optimization feature selection affects to increase accuracy results.

Keywords: Sentiment Analysis, *Naive Bayes*, *Support Vector Machine*, *Particle Swarm Optimization*

I. PENDAHULUAN

Pada awalnya esemka merupakan produk mobil hasil dari rakitan siswa-siswa Sekolah Menengah Kejuruan (SMK) yang berasal dari beberapa kota di Indonesia, bekerja sama dengan institusi dan industri dalam negeri, serta beberapa perusahaan dari lokal dan nasional. Kandungan komponen lokal (dalam negeri) berkisar antara 20% sampai dengan 80%. Berjalan seiringnya waktu, kehadiran mobil esemka ini banyak menimbulkan pro dan kontra di kalangan masyarakat,

seperti yang ramai diperbincangkan di media sosial *twitter* setelah Presiden Republik Indonesia Joko Widodo meresmikan pabrik mobil esemka pada hari jumat 6 september 2019. Peresmian pabrik yang berada dibawah naungan PT Solo Manufaktur ini berada di Boyolali, jawa tengah.

Pada kesempatan tersebut Presiden Joko Widodo memberikan pernyataan bahwa mobil esemka bukanlah mobil nasional, namun mobil tersebut merupakan produk dalam negeri, tapi kehadirannya

dapat memberi efek penting untuk pergerakan ekonomi. Senada disampaikan juga oleh direktur utama PT Solo Manufaktur Eddy Wirajaya, yang menyatakan bahwa esemka bukan mobil nasional, namun mobil rakitan anak bangsa yang *sparepart* -nya 62% adalah produksi lokal.

Dari pernyataan itulah maka berdampak pada terjadinya kontroversi mengenai mobil esemka, banyak masyarakat yang memberikan opini bahwa esemka yang diproduksi sekarang bukanlah atau berbeda dari mobil esemka yang sebelumnya digadag-gadag sebagai mobil buatan anak negeri, khususnya siswa dari Sekolah Menengah Kejuruan atau SMK.

Pada penelitian analisis sentimen sebelumnya, yaitu analisis sentimen mengenai komentar pengguna aplikasi *OVO* pada *Google Play* dengan menggunakan *Naive Bayes* dan *feature selection Particle Swarm Optimization*, dimana pada penelitian tersebut bertujuan mengukur kualitas hasil analisis menggunakan tanpa *feature selection* dan dengan *feature selection*. Hasil yang didapat dari penelitian tersebut *Cross Validation Naive Bayes* tanpa *feature selection* adalah 82.30 % untuk *Accuracy* dan 0.780 untuk *AUC*. Sedangkan untuk *Naive Bayes* ditambahkan *feature selection* mengalami peningkatan menjadi 83.60 % untuk *Accuracy*, serta 0.801 untuk *AUC*.

Pada penelitian ini, peneliti menggunakan *Naive Bayes* dan *Support Vector Machine*, serta *Particle Swarm Optimization* (PSO) sebagai seleksi fitur yang diterapkan untuk meningkatkan performa klasifikasi teks.

Melalui penelitian ini diharapkan bisa mendapatkan hasil yang lebih baik dari penelitian yang sudah dilakukan sebelumnya, serta melalui penelitian ini dapat diketahui perbandingan hasil mana yang lebih baik antara algoritma *Naive Bayes* dan *Support Vector Machine*.

Pada penelitian analisis sentimen ini, dilakukan untuk melihat sebuah opini seseorang atau masyarakat yang ditujukan kepada peresmian pabrik mobil esemka, sehingga dihasilkan opini yang dapat masuk ke kategori opini negatif ataupun opini positif. Dengan adanya penilaian maupun pandangan masyarakat yang disampaikan pada media sosial, maka peneliti berinisiatif membuat penelitian ini, peneliti melihat belum ada penelitian tentang *opinion mining* sebelumnya yang mengangkat permasalahan tersebut,

sehingga kedepannya diharapkan dapat bermanfaat membantu untuk melakukan riset atas opini masyarakat yang mengandung sentimen positif, negatif ataupun netral lainnya.

II. LITERATUR DAN METODE

A. Analisis Sentimen.

Sentiment Analysis, atau yang biasa disebut juga *Opinion Mining*, adalah bidang studi yang bertujuan untuk menganalisis opini, sentimen, penilaian, evaluasi, sikap dan emosi publik terhadap suatu entitas dari produk, pelayanan, suatu permasalahan, organisasi, peristiwa tertentu, topik yang hangat dibicarakan dan atributnya [1].

Ada beberapa macam metode yang bisa digunakan pada analisis sentimen, yaitu antara lain, NB (*Naive Bayes*), *Decision Tree*, KNN (*K-Nearest Neighbor*), *Neural Networks* dan juga SVM (*Support Vector Machines*).

B. Particle Swarm Optimization

Particle Swarm Optimization (PSO) pertama kali diusulkan oleh Eberhart dan Kennedy pada tahun 1995, *Particle swarm optimization* (PSO) adalah jenis algoritma kecerdasan yang berasal dari perilaku kawanan burung mencari makan [2]. *Particle Swarm Optimization* dapat diasumsikan sebagai kelompok burung secara mencari makanan disuatu daerah. Burung tersebut tidak tahu dimana makanan tersebut berada, tapi mereka tahu seberapa jauh makanan itu berada, strategi terbaik untuk menemukan makanan tersebut adalah dengan mengikuti burung yang terdekat dari makanan. *Particle swarm optimization* banyak digunakan untuk memecahkan masalah optimasi, serta sebagai masalah seleksi fitur [3].

Menurut Basari bahwa *Particle Swarm Optimization* merupakan teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter. Dalam *Particle Swarm Optimization* terdapat beberapa teknik untuk melakukan pengoptimasian diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, menseleksi atribut (*attribute selection*) dan *feature selection* [4].

C. Algoritma Naive Bayes

Metode ini menggunakan teorema *Bayes*, yang ditemukan oleh Thomas Bayes abad ke-18 [5]. Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi

probabilitas keanggotaan suatu *class*. Menurut Wu dan Kumar bahwa *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam *data mining*. *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training* [6].

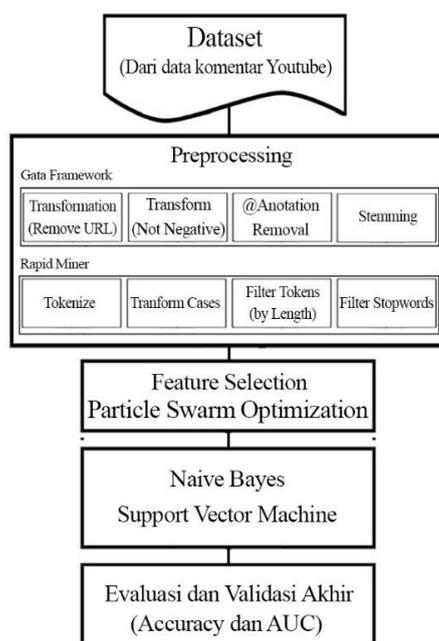
D. Algoritma Support Vector Machine

Support Vector Machines telah menjadi metode klasifikasi dan regresi yang sering digunakan untuk masalah linear dan nonlinear. Kelebihan dari algoritma *Support Vector Machines* adalah dari kemampuannya untuk menerapkan pemisahan linear pada input data non linear berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi *kernel* yang diperlukan.

Efektivitas *Support Vector Machine* sangatlah dipengaruhi oleh jenis fungsi kernel yang dipilih dan diterapkan berdasarkan karakteristik data [7]. *Support Vector Machine* dioptimalkan dengan *Particle Swarm* [8].

III. METODE

Berikut adalah langkah – langkah metode penelitian yang dijelaskan pada gambar dibawah ini :



Gambar 1. Metode Penelitian

A. Pengumpulan dan Pengolahan Data

Penelitian ini menggunakan data yang didapat dari komentar video *youtube* milik akun resmi Sekretariat Presiden yang diposting pada tanggal 6 September 2019 lalu, berikut *source link* video <https://youtu.be/A8FtqVhUE>. Proses pengambilan data dilakukan secara manual satu persatu pada tanggal 27 Maret 2020, data yang digunakan dalam penelitian ini sebanyak 288 data yang terdiri dari 180 data *review* positif dan 108 data *review* negatif.

Setelah melakukan pengumpulan data, dilakukan proses pengolahan data atau *preprocessing*, tahapan ini mencakup kegiatan membangun data dan dilanjutkan kegiatan pembersihan data agar dapat dilakukan pengelolaan ketahap selanjutnya. Berikut tahapan *preprocessing* data, antara lain :

1. *Remove URL*.
Yaitu proses menghilangkan *Uniform Resource Locator* yang ada pada dataset hasil pengumpulan data. Biasanya digunakan untuk merujuk pada *source* atau sumber dari suatu berita/informasi.
2. *Transform (Not Negative)*.
Proses mengubah kata – kata yang bermakna negatif yang akan disatukan dengan tanda garis bawah (_).
3. *Anotation Removal*.
Menghapus tanda Mention (@) beserta teks yang ada dibelakangnya. Pada media sosial biasanya berfungsi untuk memanggil user/pengguna lainnya.
4. *Stemming*.
Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.
5. *Tokenize*.
Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.
6. *Transforms Cases*.
Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.
7. *Filter Tokens (By Length)*.

Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.

8. Filter Stopword.

Membuang kata – kata yang diabaikan pada sentimen analis, biasanya yang berupa kata sambung dan kata keterangan.

Preprocessing pertama dilakukan menggunakan gataframework, yaitu pengolahan data teks berbasis web, dengan menggunakan bahasa pemrograman php. Gataframework dapat diakses di alamat link www.gataframework.com/textmining. Banyak yang tersedia di Gataframework, antara lain *annotation removal*, *remove url* dan lainnya. Peneliti menggunakan Gataframework karena memiliki keunggulan dalam melakukan proses *Stemming* bahasa Indonesia. Pengolahan data selanjutnya dilakukan menggunakan tools *Rapidminer*.

B. Seleksi Fitur dan Algoritma

Peneliti mengusulkan menggunakan metode pemilihan seleksi fitur *Particle Swarm Optimization* (PSO) yang digunakan untuk meningkatkan akurasi pada pengklasifikasian algoritma. Algoritma yang akan dijadikan perbandingan atau komparasi pada penelitian ini adalah Naive Bayes dan Support Vector Machine. Peneliti menggunakan Naive Bayes dan Support Vector Machine karena keduanya merupakan algoritma terbaik dalam data mining, dan sangat baik dalam melakukan proses pengklasifikasian data.

C. Evaluasi dan Validasi

Pada tahap pengujian metode, peneliti menggunakan *RapidMiner* dalam melakukan proses eksperimen, data *training* yang digunakan adalah komentar video *youtube* mengenai pro dan kontra pembangunan pabrik mobil esemka. Kemudian *dataset* dikelompokkan menjadi 2 bagian yaitu opini positif dan negatif.

Peneliti menggunakan *10 Fold Cross Validation* dalam melakukan validasi, dan untuk mengukur akurasi dilakukan dengan *Confusion Matrix*, yaitu membandingkan akurasi *Naive Bayes* dengan *Support Vector Machine* tanpa dan dengan menggunakan fitur seleksi *Particle Swarm Optimization* (PSO). Kemudian Kurva *ROC* digunakan untuk mengukur nilai *AUC*.

1. *Confusion Matrix*

Matriks yang menginformasikan hasil prediksi secara keseluruhan dari nilai akurasi dan untuk melihat

kinerja pengklasifikasi, yaitu seberapa sering kasus *class X* yang benar diklasifikasikan sebagai *class X* atau kesalahan klasifikasi *class* yang lainnya. Ketika *dataset* hanya memiliki dua kelas, yaitu *class* positif dan *class* negatif, maka dapat dibuatkan seperti gambar dibawah ini :

Tabel 1. *Confusion Matrix*

Correct Classification	Classified As	
	+	-
+	True Positive	False Negative
-	False Positive	True Negative

True positif (TP) merupakan jumlah *record* positif dalam *dataset* yang diklasifikasikan positif. *True* negatif (TN) merupakan jumlah *record* negatif dalam *dataset* yang diklasifikasikan negatif. *False* positif merupakan jumlah *record* negatif dalam *dataset* yang diklasifikasikan positif. *False* negatif (fn) merupakan jumlah *record* positif dalam *dataset* yang diklasifikasikan negatif.

2. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. Kurva ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false* positif sebagai garis horizontal dan *true* positif sebagai garis vertikal. Pedoman umum untuk mengklasifikasikan keakuratan pengujian menggunakan *AUC* [9].

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. Kurva ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false* positif sebagai garis horizontal dan *true* positif sebagai garis vertikal. Pedoman umum untuk mengklasifikasikan keakuratan pengujian menggunakan *AUC* [9].

- 0.90 - 1.00 = *Excellent Classification*;
- 0.80 - 0.90 = *Good Classification*;
- 0.70 - 0.80 = *Fair Classification*;
- 0.60 - 0.70 = *Poor Classification*;
- 0.50 - 0.60 = *Failure*

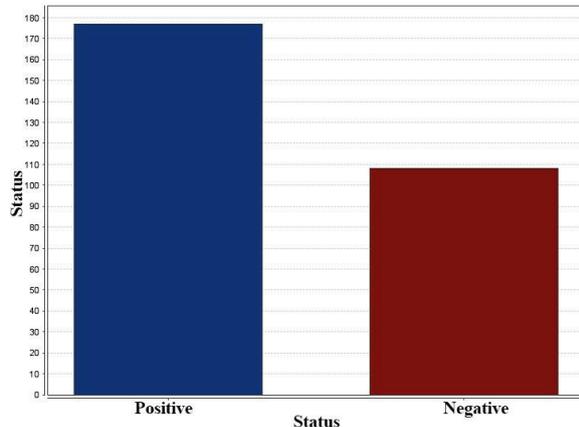
IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Hasil pengumpulan data yang didapat dari komentar video *youtube* mengenai peresmian pabrik mobil esemka, peneliti mengumpulkan 288 data opini pengguna *youtube*. Kemudian data tersebut kumpulkan dalam format file xls, setelah itu data

dikelompokkan menjadi 2 bagian yaitu data positif dan data negatif.

Pengelompokan data ini dilakukan oleh beberapa koresponden, sehingga kemudian dihasilkan 180 opini Positif dan 108 opini negatif.



Gambar 2. Statistik hasil pengumpulan data.

B. Pengolahan Data

Setelah pengumpulan data dan membagi data menjadi komentar positif dan komentar negatif, selanjutnya dilakukan proses pengolahan data atau *preprocessing*. Berikut ini adalah tahapan *preprocessing* :

1. Remove URL.

Yaitu proses menghilangkan *Uniform Resource Locator* yang ada pada dataset hasil pengumpulan data. Berikut contoh perbedaan sebelum dan sesudah proses.

Tabel 2. Hasil *Remove URL*.

Sebelum	Sesudah
Nih kompas juga bilang mirip https://youtu.be/iYPELEaHi3U	Nih kompas juga bilang mirip kalo kaya gini gimana tuh

2. Transform (Not Negative).

Proses mengubah kata – kata yang bermakna negatif yang akan disatukan dengan tanda garis bawah (_).

Tabel 3. Hasil *Transform Negative*.

Sebelum	Sesudah
---------	---------

Kalo ente itu nyinyir bukan dukung	kalo ente itu nyinyir bukan dukung
------------------------------------	------------------------------------

3. Anotation Removal.

Menghapus tanda *Mention* (@) beserta teks yang ada dibelakangnya.

Tabel 4. Hasil *Anotation Removal*.

Sebelum	Sesudah
Nunggu reviewnya di @autonetmagz aja ah	Nunggu reviewnya di aja ah

4. Stemming.

Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.

Tabel 5. Hasil *Stemming*.

Sebelum	Sesudah
Saya tertarik Alhamdulillah kalo harga lebih murah Mudah mudah saya beri rejeki oleh Allah untuk membeli ESEMKA tipe pickup untuk usaha saya Amin ya Allah	tarik alhamdulillah kalo harga murah mudah mudah rejeki allah beli esemka tipe pickup usaha amin allah

5. Tokenize.

Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.

6. Transforms Cases.

Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.

Tabel 6. Hasil *Transforms Cases*.

Sebelum	Sesudah
Mantap bisa buat mobil sendiri. Harusnya dihadirkan siapa saja insinyur yang merancang mesinya mobilnya diatas 1000 cc	Mantap bisa buat mobil sendiri harusnya dihadirkan siapa saja insinyur yang merancang mesinya mobilnya diatas 1000 cc

7. Filter Tokens (By Lenght).

Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.

8. Filter Stopword.

Membuang kata – kata yang diabaikan pada sentimen analis, biasanya yang berupa kata sambung dan kata keterangan.

Tabel 7. Hasil Filter Stopword.

Sebelum	Sesudah
semua butuh proses tidak ada yang berubah kalau semua berfikir negatif ayo dukung semua produk indonesia dimulai dari sekarang	butuh proses berubah berfikir negatif ayo dukung produk indonesia

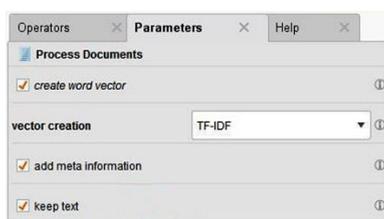
Berikut gambar proses pengolahan data pada tools *rapidminer*.



Gambar 3. Proses Pengolahan Data Rapidminer.

Data hasil *preprocessing* yang berupa kata akan diubah ke dalam bentuk angka dengan dilakukan proses pembobotan kata yang bertujuan untuk menghitung bobot pada masing-masing kata yang akan digunakan sebagai fitur, semakin banyak dokumen yang akan diproses maka semakin banyak fitur. Pada tahapan ini terdapat dua bagian proses yaitu *TF* (*Term Frequency*) dan *IDF* (*Inverse Document Frequency*),

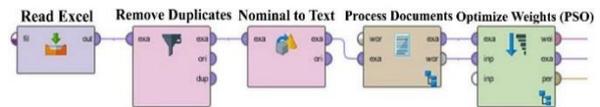
TF adalah jumlah kemunculan tiap kata pada sebuah dokumen semakin banyak kata muncul pada tiap dokumen maka semakin besar nilai *TF*. *Df* adalah jumlah nilai dokumen pada tiap kata yang berbanding terbalik yaitu apabila suatu kata jarang muncul pada sebuah dokumen maka nilai *IDF* lebih besar daripada kata yang sering muncul.



Gambar 4. Proses *Tf Idf*.

C. Fitur seleksi *Particle Swarm Optimization*.

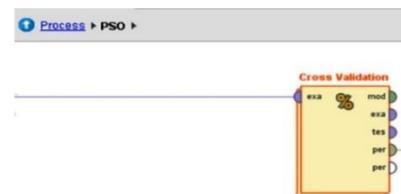
Peneliti mengusulkan menggunakan *Particle Swarm Optimization* untuk meningkatkan akurasi dari pengklasifikasi Algoritma *Naive Bayes* dan *Support Vector Machine*. Penelitian ini nantinya menghasilkan akurasi dan nilai *AUC* dengan menggunakan aplikasi *RapidMiner* untuk hasil evaluasi.



Gambar 5. Model Pengujian *PSO* dengan *RapidMiner*.

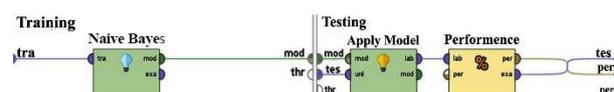
D. Proses Validasi Algoritma.

Selanjutnya adalah penggunaan algoritma *Naive Bayes* dan *Support Vector Machine* untuk melakukan klasifikasi data yang dihubungkan dengan pengujian *10-fold cross validation* dimana proses ini untuk mengevaluasi proses kerja algoritma tersebut dengan membagi data secara acak ke dalam *10 fold* untuk mendapatkan 10 data yang sama, kemudian data tersebut digunakan 9 *fold* untuk data *training* dan 1 *fold* untuk data *testing*.



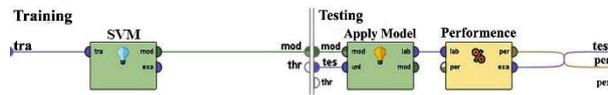
Gambar 6. Proses *Cross Validation* dengan *Rapidminer*.

Tahapan selanjutnya adalah proses Evaluasi, ialah mengevaluasi kinerja terhadap permodelan dan perhitungan yang telah digunakan dengan menerapkan algoritma *Naive Bayes* dan *Support Vector Machine* yang ditambahkan *Particle Swarm Optimization* agar dihasilkan peningkatan akurasi pada penghitungan. Untuk mengestimasi performa dari model algoritma yang telah dipilih digunakan *Cross Validation* sehingga dapat menghasilkan nilai akurasi (Proses *Cross Validation* dapat dilihat pada Gambar 6). Dibawah ini adalah model pengujian *Naive Bayes* dengan menggunakan tools *Rapidminer*.



Gambar 7. Model Pengujian *Naive Bayes* dengan *RapidMiner*

Dibawah ini adalah model pengujian *Support Vector Machine* dengan menggunakan tools *Rapidminer*.



Gambar 8. Model Pengujian *Support Vector Machine* dengan *RapidMiner*.

E. Evaluasi

Penelitian ini menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* untuk melakukan analisis sentimen. Kemudian dalam menentukan evaluasinya peneliti menggunakan *Accuracy* dan *AUC (Area Under Curve)*. Selanjutnya peneliti akan membandingkan hasil akurasi dari *Naive Bayes* dan *Support Vector Machine*, dengan atau tanpa menggunakan seleksi fitur *Particle Swarm Optimization*.

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 288 komentar opini pengguna *Youtube* mengenai video peresemian pabrik mobil esemka, maka hasil Akurasi Algoritma *Naive Bayes* tanpa menggunakan Fitur Seleksi *Particle Swarm Optimization* yaitu sebesar 75.04 %, sedangkan untuk nilai *AUC* sebesar 0.500.

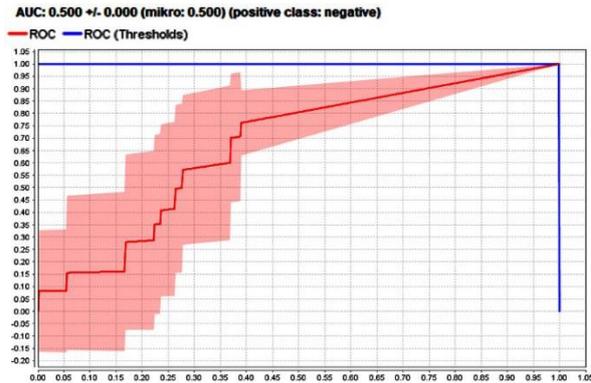
Berikut ini adalah Tabel *Confusion Matrix* Algoritma *Naive Bayes* tanpa fitur seleksi.

Tabel 8. *Confusion Matrix* Algoritma NB.

	True Positive	True Negative	
Pred Positive	141	33	81.03%
Pred Negative	39	75	65.79%
	78.33%	69.44%	

Pada tabel 8 *Confusion Matrix* dapat dilihat, sebanyak 141 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 39 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 33 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 75 data di prediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 9. Grafik *Area Under Curve (AUC) NB*.

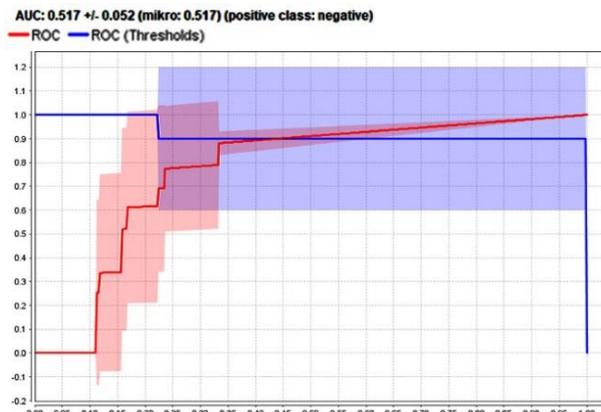
Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 288 komentar opini pengguna *Youtube* mengenai video peresemian pabrik mobil esemka, maka hasil Akurasi Algoritma *Naive Bayes* dengan menggunakan Fitur Seleksi *Particle Swarm Optimization* yaitu sebesar 83.33 %, sedangkan untuk nilai *AUC* sebesar 0.517.

Berikut ini adalah Tabel *Confusion Matrix* Algoritma *Naive Bayes* dengan fitur seleksi.

Tabel 9. *Confusion Matrix* Algoritma NB dan PSO.

	True Positive	True Negative	
Pred Positive	149	17	89.76%
Pred Negative	31	91	74.59%
	82.78%	84.26%	

Pada tabel 9 *Confusion Matrix* dapat dilihat, sebanyak 149 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 31 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif. Dan sebanyak 17 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 91 data di prediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 10. Grafik Area Under Curve (AUC) NB dan PSO.

Berikutnya adalah pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 288 komentar opini pengguna Youtube mengenai video peresemian pabrik mobil esemka, maka hasil Akurasi Algoritma Support Vector Machine tanpa menggunakan Fitur Seleksi Particle Swarm Optimization yaitu sebesar 78.81 %, sedangkan untuk nilai AUC sebesar 0.887.

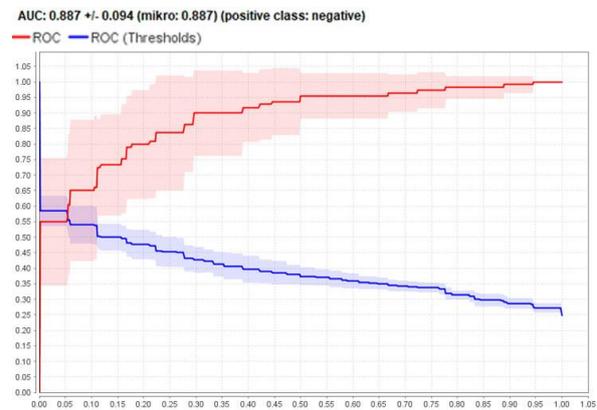
Dibawah ini adalah Tabel Confusion Matrix Algoritma Support Vector Machine tanpa fitur seleksi.

Tabel 10. Confusion Matrix Algoritma SVM

	True Positive	True Negative	
Pred Positive	155	36	81.15%
Pred Negative	25	72	74.23%
	86.11%	66.67%	

Pada tabel 10 Confusion Matrix dapat dilihat, sebanyak 155 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 25 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 36 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 72 data di prediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 11. Grafik Area Under Curve (AUC) SVM.

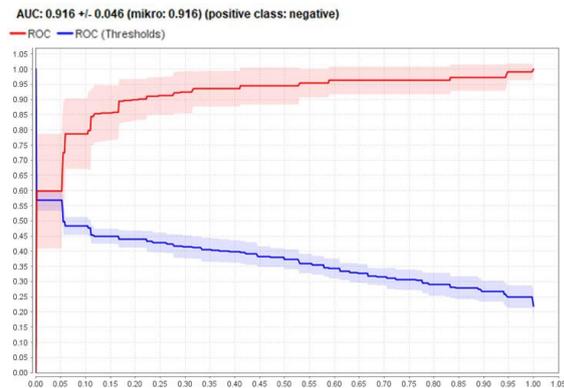
Yang terakhir pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 288 komentar opini pengguna Youtube mengenai video peresemian pabrik mobil esemka, maka hasil Akurasi Algoritma Support Vector Machine dengan menggunakan Fitur Seleksi Particle Swarm Optimization yaitu sebesar 88.19 %, sedangkan untuk nilai AUC sebesar 0.916.

Berikut ini adalah Tabel Confusion Matrix Algoritma Support Vector Machine dengan fitur seleksi.

Tabel 11. Confusion Matrix Algoritma SVM dan PSO

	True Positive	True Negative	
Pred Positive	175	29	85.78%
Pred Negative	5	79	94.05%
	97.22%	73.15%	

Pada tabel 11 Confusion Matrix dapat dilihat, sebanyak 175 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 5 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif. Dan sebanyak 29 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 79 data di prediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 12. Grafik Area Under Curve (AUC) SVM dan PSO.

Dari hasil pengujian diatas, didapat perbedaan hasil akurasi dan AUC dari Naive bayes dengan Support Vector Machine. Nilai akurasi Support Vector Machine memiliki hasil yang lebih baik dari Naive Bayes, dengan atau tanpa menggunakan fitur seleksi Particle Swarm Optimization.

Nilai akurasi Naive Bayes tanpa menggunakan fitur seleksi adalah 75.04% dan nilai AUC 0.500, sedangkan Naive Bayes dengan menggunakan fitur seleksi Particle Swarm Optimization memiliki akurasi sebesar 83.33% dan nilai AUC 0.517, selisih peningkatan akurasi tanpa dan dengan menggunakan fitur seleksi adalah 8.29% dan selisih peningkatan AUC 0.017.

Kemudian pada pengujian menggunakan Support Vector Machine tanpa fitur seleksi, nilai akurasi 78.81%, dan nilai AUC 0.887. sedangkan Support Vector Machine dengan menggunakan fitur seleksi Particle Swarm Optimization memiliki hasil akurasi sebesar 88.19% dan nilai AUC 0.916. Selisih peningkatan akurasi tanpa dan dengan menggunakan fitur seleksi adalah 9.38% dan selisih peningkatan AUC 0.029. Berikut adalah tabel perbandingannya.

Tabel 12. Perbandingan Hasil Pengujian

	NB	NB + PSO	SVM	SVM + PSO
Accuracy	75.04%	83.33%	78.81%	88.19%
AUC	0.500	0.517	0.887	0.916

V. KESIMPULAN

Berdasarkan dari penelitian yang sudah dilakukan, dapat disimpulkan bahwa hasil akurasi dan AUC dengan algoritma Support Vector Machine lebih baik dari algoritma Naive Bayes, bahkan hasil pengujian pengklasifikasian AUC nilai keakuratannya dapat dikategorikan sebagai Excellent Classification. Begitu juga pengaruh penggunaan fitur seleksi Particle Swarm Optimization terhadap kedua algoritma tersebut juga sangat besar, dapat dilihat dari hasil pengujian diatas, terjadi peningkatan nilai akurasi dan AUC sangat signifikan. Dan pada penelitian ini juga, hasil akurasi dan AUC dengan menggunakan algoritma Naive Bayes dengan dan tanpa menggunakan fitur seleksi Particle Swarm Optimization, memiliki hasil yang lebih baik dari penelitian sebelumnya yang menjadi referensi.

Untuk mendukung penelitian ini, peneliti akan mengembangkan aplikasi opini dari komentar youtube mengenai mobil Esemka untuk mengklasifikasikan opini negatif dan positif menggunakan bahasa pemrograman php.

REFERENSI

- [1] Liu, B. (2012). Sentiment Analysis and Opinion Mining Mining. Synthesis Lectures on Human Language Technologies. [Draft]. In *Synthesis Lectures on Human Language Technologies* (Vol. 5, Issue 1). <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- [2] Yun L, Cao Q, and HZ-CI. (2011). Undefined Application of the PSO-SVM model for Credit Scoring. *ieeexplore.ieee.org*
- [3] Abraham, A., Grosan, C., & Ramos, V. (2006). *Swarm Intelligence in Data Mining*. London: Verlag Berlin Heidelberg, Springer.
- [4] Basari, Hussin B., Ananta, & Zeniarja, J. (2013), Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*,53,453-462. doi:10.1016/j.proeng. 2013.02.059.
- [5] Barde, N. C., & Patole, M. (2013). Classification and Forecasting of Weather using ANN, k-NN

- and Naïve Bayes Algorithms. In *International Journal of Science and Research* (Vol. 5). www.ijsr.net
- [6] Mukminin, A., Riana, D. (2017). Komparasi Algoritma C4. 5, Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah. *Jurnal Informatika*, 4(1), 21–31.
<https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/1002>
- [7] Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *First International Conference on Information Technology and Quantitative Management*, 17, 26–32.
<https://doi.org/10.1016/j.procs.2013.05.05>
- [8] Al-Amrani, Y., Lazaar, M., El Kadiri. (2018). Sentiment Analysis Using Hybrid Method of Support Vector Machine And Decision Tree. *Journal of Theoretical and Applied Information Technology* 96(7), pp.1886-1895.
- [9] Zhang, L., & Liu, B. (2016). Sentiment Analysis and Opinion Mining. In *Encyclopedia of Machine Learning and Data Mining* (pp. 1–10).
https://doi.org/10.1007/978-1-4899-7502-7_907-1
- [10] El Houby, E. M. F., Yassin, N. I. R., & Omran, S. (2017). A hybrid approach from ant colony optimization and K-nearest neighbor for classifying datasets using selected features. *Informatika (Slovenia)*, 41(4), 495–506.
- [11] Koncz, P., & Paralic, J. (2011). An approach to feature selection for sentiment analysis. *INES 2011 - 15th International Conference on Intelligent Engineering Systems, Proceedings*, 357–362.
<https://doi.org/10.1109/INES.2011.5954773>
- [12] Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Evolusi Volume 4 Nomor 2 - 2016*, 4(2), 20–27.
- [13] Lu, Y., Liang, M., Ye, Z., & Cao, L. (2015). Improved particle swarm optimization algorithm and its application in text feature selection. *Applied Soft Computing Journal*, 35, 629–636.
<https://doi.org/10.1016/j.asoc.2015.07.005>
- [14] Indriati, I., & Ridok, A. (2016). Sentiment Analysis For Review Mobile Application Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN). *Journal of Environmental Engineering and Sustainable Technology*, 3(1), 23–32.
<https://doi.org/10.21776/ub.jeest.2016.003.01.4>