

Optimasi Particle Swarm Optimization Pada Naive Bayes Untuk Sentiment Analysis Furniture

Rizki Aulianita ^{1,*}, Achmad Rifai ²

¹ Sistem Informasi; STMIK Nusa Mandiri Jakarta; Jl. Damai No. 8 Warung Jati Barat (Margasatwa), Pasar Minggu Jakarta Selatan 12540, (021) 78839513; e-mail: rizki.rzk@nusamandiri.ac.id

² Teknik Informatika; STMIK Nusa Mandiri Jakarta; Jl. Damai No. 8 Warung Jati Barat (Margasatwa), Pasar Minggu Jakarta Selatan 12540, (021) 78839513; e-mail: achmad.acf@gmail.com

* Korespondensi: e-mail: rizki.rzk@nusamandiri.ac.id

Diterima: 10 September 2018; Review: 13 September 2018; Disetujui: 28 September 2018.

Cara sitasi: Aulianita R, Rifai A. 2018. Optimasi Particle Swarm Optimization Pada Naive Bayes Untuk Sentiment Analysis Furniture. Information System For Educators and Professionals. 3 (1): 31 - 40.

Abstrak: Penelitian ini memfokuskan pada peningkatan akurasi produk furniture dengan menggunakan algoritma Naive Bayes yang dioptimasi PSO. Fungsi dari penerapan metode algoritma tersebut dalam mengklasifikasikan review positif dan negatif yaitu untuk memudahkan pembeli mendapatkan informasi secara mudah sebelum memilih produk yang tepat untuk dibeli. Pengolahan data uji training menggunakan 100 data review furniture positif dan 100 data review furniture negatif. Hasil eksperimen pada Naive Bayes optimasi Particle Swarm Optimization (PSO) sebesar 93,50% sehingga dapat disimpulkan bahwa Naive Bayes merupakan metode klasifikasi teks yang sederhana dan PSO salah satu teknik optimasi yang berfungsi untuk meningkatkan pembobotan atribut (*attribute weight*).

Kata kunci: *text mining, sentiment analysis, Naive Bayes, Naive Bayes optimasi PSO*

Abstract: *This study focuses on improving the accuracy of furniture products using the Naive Bayes algorithm optimized by PSO. The function of the application of the algorithm method is to classify positive and negative reviews that is to make it easier for buyers to get information easily before choosing the right product to buy. The training test data processing uses 100 positive furniture review data and 100 negative furniture review data. The experimental results on Naive Bayes Particle Swarm Optimization (PSO) optimization were 93.50% so that it can be concluded that Naive Bayes is a simple text classification method and PSO is one of the optimization techniques that serves to increase attribute weighting.*

Keywords: *text mining, sentiment analysis, Naive Bayes, PSO Optimization Naive Bayes*

1. Pendahuluan

Review produk saat ini merupakan sebuah cara yang dilakukan untuk menilai kelayakan atau kecakapan sebuah produk berdasarkan informasi atau opini yang dimiliki oleh reviewer. Beberapa situs online selain menjual produk, juga membuat kolom komentar untuk menampung review produk yang telah dibeli oleh *customer*. Bahkan tidak jarang, *customer* dapat menaikkan rating penjual ataupun membuat penjualan menurun. Tidak hanya itu, saat ini kegiatan review sudah menjadi hal yang lumrah di era disruptif ini. Berdasarkan hal tersebutlah, para pengguna internet berbondong-bondong untuk mereview atau memberikan opini terhadap sesuatu yang mereka beli atau gunakan.

Dengan adanya review produk inilah, memudahkan *customer* dalam menentukan pilihan untuk membeli produk. *Customer* dapat dengan mudah mengetahui informasi produk sebelum membelinya. Di sisi lain, hal ini juga menjadi sebuah keuntungan bagi pemilik perusahaan untuk mengetahui komentar pasar terhadap produk yang dijual perusahaan sehingga berdasarkan hasil review tersebut dapat diketahui kekurangan dan kelebihan sebuah produk. Beberapa situs online yang menyediakan fasilitas review produk yaitu www.yelp.com salah satunya ulasan mengenai restoran. www.gsmarena.com merupakan situs yang menyediakan review atau ulasan mengenai handphone. www.amazon.com merupakan sebuah *marketplace online e-commerce* yang memiliki fasilitas rating dan kolom komentar sehingga dapat memudahkan para pembeli lain untuk mendapatkan informasi terhadap produk yang akan dibeli karena banyak pembeli yang memeriksa pendapat ataupun opini sebelum membeli produk untuk membuat pilihan produk yang tepat. [Taylor dalam Indrayuni, 2016]

Banyak text review dalam segala format seperti text, video, image dan lainnya namun belum dapat menghasilkan sebuah informasi. Untuk itulah diperlukan adanya sentiment analisis yang bermanfaat untuk mengelompokkan review positif dan negatif menggunakan metode algoritma sehingga secara otomatis dapat menghasilkan informasi. Analisis sentimen dapat dianggap sebagai proses klasifikasi dengan tiga klasifikasi utama tingkat di SA: dokumen-tingkat, kalimat-tingkat, dan aspek-tingkat SA. Dokumen-tingkat SA bertujuan untuk mengklasifikasikan dokumen opini sebagai mengekspresikan pendapat atau sentimen positif atau negatif [Aulianita, 2015].

“Terdapat enam metode dalam klasifikasi seperti Rocchio, k-NN, Regression Model, Naive Bayes dan Decision Tree dan Decision Rules.” [Rasjid and Setiawan, 2017]. “Beberapa algoritma yang sering digunakan dalam *machine learning* yaitu SVM, Naive Bayes dan Random Forest” [Tjahyanto and Sisephaputra, 2017] “Naive Bayes membutuhkan sebuah data training untuk mengestimasi parameter yang diperlukan dalam klasifikasi teks”. [Al-ghuribi and Alshomrani, 2013]. “Naive Bayes memiliki fakta menarik seperti mudah diimplementasikan, dapat menghasilkan akurasi yang terbaik dalam dataset dengan volume besar serta dapat mengklasifikasikan dokumen dengan akurasi optimal.” Naive Bayes dan SVM termasuk ke dalam metode yang paling banyak digunakan untuk klasifikasi dalam klasifikasi teks.

“Penggolongan Naive Bayes adalah pengklasifikasi linier yang dikenal sederhana namun sangat efisien. Model probabilistik penggolongan Klasifikasi Bayes didasarkan pada Teorema Bayes, dan sifatnya berasal dari asumsi bahwa fitur dalam kumpulan data saling independen.” [Raschka, 2014]

Pada dasarnya Naive Bayes bekerja dengan teorema bersyarat dimana teorema tersebut dimodifikasi sesuai dengan aplikasi teks klasifikasi. Untuk mengevaluasi teks tersebut maka Naive Bayes memiliki dua pendekatan dua faktor metric, precision, recall untuk menemukan efektivitas melalui *True Positive (tp)*, *True Negative (tn)*, *False Positive (fp)* dan *False Negative (fn)* [Vidhya and Aghila, 2010]. Penggolongan Naive Bayes didasarkan pada prinsip statistik. Di sini, ada atau tidaknya sebuah kata dalam dokumen tekstual menentukan hasil dari prediksi. Dengan kata lain, setiap teks yang diproses adalah menetapkan probabilitas bahwa kata tersebut termasuk kategori tertentu. Probabilitas ini dihitung dari kejadian dalam data latih dimana kategori tersebut sudah dikenal. Ketika semua probabilitas ini dihitung, dokumen baru bisa diklasifikasikan sesuai dengan jumlah probabilitas untuk setiap kategori. Namun, penggolong ini tidak mengambil jumlah kejadian pada data training, yang merupakan sumber informasi tambahan yang berpotensi bermanfaat. Mereka disebut "bayes" karena algoritma mengasumsikan bahwa semua istilah terjadi independendari satu sama lain.” [Bijalwan et al., 2014]

Untuk itu, dalam penelitian ini, digunakan metode Naive Bayes untuk meningkatkan performa serta akurasi yang lebih baik dengan menggunakan dataset furniture pada www.amazon.com. Dataset yang digunakan terdiri dari 100 review positif dan 100 review negatif yang dioptimasi menggunakan PSO. Particle Swarm Optimization merupakan salah satu teknik optimasi yang sangat sederhana untuk memodifikasi beberapa parameter serta digunakan untuk memecahkan masalah optimasi pada *feature selection* [Kristiyanti and Wahyudi]

Particle Swarm Optimization (PSO) sebagai teknik evolusi pertama dikemukakan oleh J Kennedy dan R. Eberhart pada tahun 1995. Seorang Psikolog dan Insinyur Listrik di Amerika Serikat [Liu et al., 2014]

Dalam PSO, setiap individu diperlakukan sebagai partikel, yang merupakan solusi dalam masalah ini. Partikel didefinisikan sebagai titik point $X_i(X_{i1}, X_{i2}, \dots, X_{id}, \dots, X_{iD})$ dalam ruang dimensi. Disini ruang dimensi hanya sebagai lingkungan pencarian. Sementara itu setiap partikel memiliki kecepatan sendiri seperti $V_i(V_{i1}, V_{i2}, \dots, V_{id}, \dots, V_{iD})$. Dengan demikian memungkinkan partikel untuk bergerak dalam pencarian ruang. Dari penjelasan di atas, *array solution* diinisialisasi secara acak dan kemudian memperbaharui posisi kecepatan mereka masing-masing [Liu et al., 2014]

2. Metode Penelitian

Dalam penelitian ini, peneliti menggunakan metode algoritma Naïve Bayes dengan optimasi dengan PSO (Particle Swarm Optimization). Metode Naive Bayes merupakan metode yang sangat sederhana, efisien dan sangat sensitif terhadap seleksi fitur sehingga dapat menghasilkan akurasi yang tinggi.

Teknik pengumpulan data yang digunakan *Simple Random Sampling* yang berfungsi untuk mendapatkan sample yang langsung dilakukan pada unit sampling. Dengan demikian setiap unit sampling sebagai unsur populasi yang terpengcil memperoleh peluang sama untuk menjadi sampel atau untuk mewakili populasi [Aulianita, 2015].

Dataset yang digunakan bersumber dari www.amazon.com. Data sampling yang diambil yaitu review furniture sebanyak 100 data review positif dan 100 data review negatif. Preprocessing yang digunakan pada data latih menggunakan *Tokenize*, *Stem* dan *Filter* sedangkan *feature weight metode* peneliti menggunakan *TF-IDF (Term Frequency Inverse Document Frequency)* dan *feature selection* menggunakan *Particle Swarm Optimization (PSO)*.

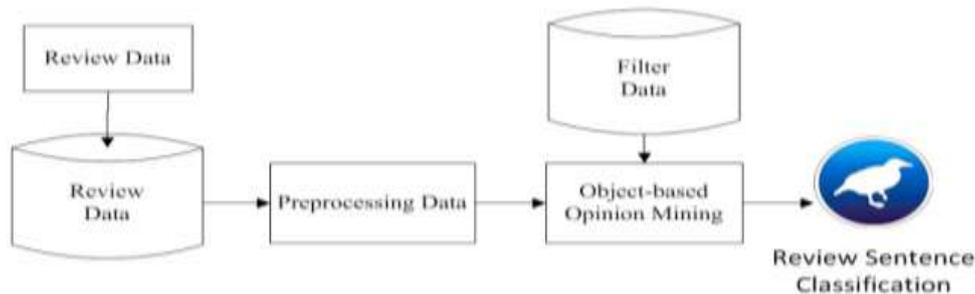
Tujuan dari penelitian ini yaitu membuktikan performa atau peningkatan akurasi Particle Swarm Optimization pada metode Naive Bayes untuk klasifikasi review positif dan review negatif sentimen analisis review furniture sehingga akan menghasilkan akurasi untuk menentukan konsumen dalam memilih furniture sebelum memutuskan untuk membelinya.

Langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentiment [Moraes et al., 2013] yaitu Pengumpulan dataset yang melingkupi suatu domain, misalnya dataset review film, dataset review produk kosmetik, sentiment buku, seniment mengenai restaurant, sentiment tentang tweet pada twitter dan lain sebagainya.

Tahap pemrosesan awal yang umumnya dilakukan dengan proses *Tokenization*, *Stopwords removal*, dan *Stemming* kemudian proses representasi angka yang dihitung dari data tekstual dengan menghasilkan kata yang muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual misalnya *bad*, *abort*, *good* dan lain-lain.

Dibuatnya seleksi fitur atau pemilihan fitur berfungsi agar pengklasifikasian lebih efisien/efektif dengan mengurangi jumlah data untuk dianalisis dengan mengidentifikasi fitur yang relevan untuk pemrosesan selanjutnya.

Metode pemilihan fitur yang biasanya digunakan adalah *Expert. Knowledge*, *Minimum Frequency*, *Information gain*, *Chi-Square*, dan lain sebagainya. *Naïve Bayes*, *Support Vector Machine*, *k-NN* merupakan metode algoritma untuk klasifikasi. Adapun Tahap evaluasi biasanya yang menentukan hasil dari akurasi, recall, precision, dan F-1. Berikut merupakan gambaran umum mengenai object berdasarkan opinion mining yaitu:



Sumber: Tjahyanto dan Bonda Sisephaputra (2017)

Gambar 1. Object Based Opinion Mining

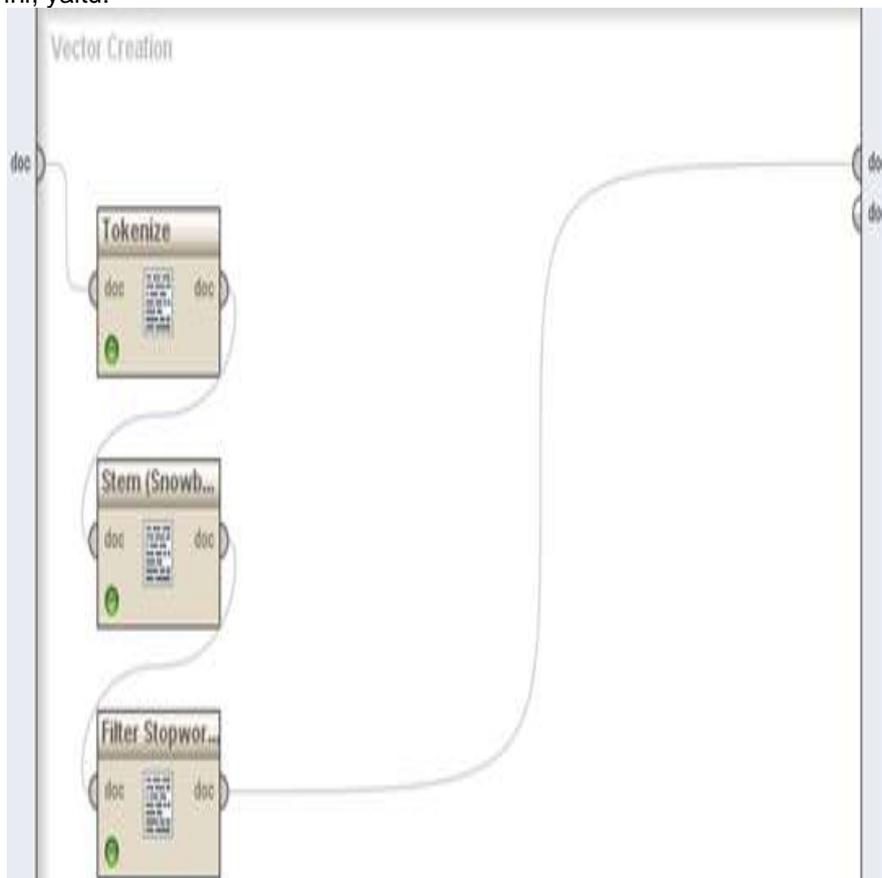
3. Hasil dan Pembahasan

Fokus permasalahan pada penelitian ini yaitu tersebarnya website yang menawarkan pembelian furniture secara *online*, sehingga dibutuhkan review untuk memudahkan pembeli

dalam menentukan pilihan barang yang tepat dan sesuai sebelum memutuskan pembelian furniture. Penelitian ini menggunakan metode Naive Bayes untuk menghitung nilai akurasi pada sentiment analisis furniture, kemudian hasil dari akurasi tersebut dioptimasi oleh Particle Swarm Optimization (PSO) sebagai pembobotan atribut sehingga akurasi yang dihasilkan lebih optimal.

Dalam melakukan pengumpulan data. Peneliti mendapatkan dataset berdasarkan review yang diperoleh dari website www.amazon dengan mengambil sample review masing-masing sebanyak 100 review positif dan 100 review negatif. Dataset yang digunakan total 200 review dengan cara mengambil setiap review kemudian disalin di Notepad lalu dikategorikan menurut label positif dan negatif.

Dataset yang sudah dipilih kemudian dilakukan *preprocessing*. Hal ini berfungsi untuk menghilangkan *noise* pada data berdimensi tinggi serta meminimalisir struktur teks yang tidak baik. 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. Proses memotong setiap kata dalam teks dan mengubah huruf dalam dokumen menjadi huruf kecil atau biasa disebut dengan *stemming*. 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. Penghapusan kata-kata yang tidak relevan seperti *the, on, of, with, and* dan sebagainya. Berikut merupakan Preprocessing dalam penelitian ini, yaitu:



Sumber: Hasil Pengolahan Data (2018)

Gambar 2. Preprocessing Metode Naïve Bayes

Pada Gambar 3 diperoleh Vector label Class Hasil Klasifikasi hasil perhitungan RapidMiner untuk menghasilkan nilai pada label yang mengandung dataset review positif dan review negatif yang nantinya akan menentukan nilai akurasi dari metode yang diusung yaitu Naive Bayes.

class positif lebih kecil dari pada *class* negatif, maka kalimat tersebut termasuk ke dalam *class* negatif.

accuracy: 51.50% +/- 3.20% (mikro: 51.50%)			
	true negatif_furniture	true positif_furniture	class precision
pred. negatif_furniture	20	17	54.05%
pred. positif_furniture	80	83	50.92%
class recall	20.00%	83.00%	

Sumber: Hasil Penelitian (2018)

Gambar 5. Accuracy Pengujian Metode Naïve Bayes

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 Naive Bayes. Kurva ROC Naive Bayes memiliki **AUC** sebesar **0.500**. Kurva tersebut menggambarkan bahwa prediksi positif 80 dianggap sebagai review negatif, prediksi negatif 20 termasuk ke dalam review negatif, sedangkan prediksi negatif 17 termasuk ke dalam review positif dan prediksi positif 83 benar termasuk ke dalam review positif. Hasil tersebut menunjukkan bahwa Metode Naive Bayes termasuk ke dalam kategori *failure classification*.



Sumber: Hasil Penelitian (2018)

Gambar 6. Kurva ROC Naïve Bayes

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 1 Confusion Matrix Naive Bayes

CLASSIFICATION	PREDICTED CLASS	
	Class=Yes	Class=No
Class=Yes	A (True Positive-TP)	B (False Negative-FN)
Class=No	C (False Positive-FP)	D (True Negative-TN)

Sumber: Hasil Penelitian (2018)

Akurasi yang diperoleh yaitu 51.50% dari 100 data review positif furniture dan 100 review negatif. Data review yang dihasilkan RapidMiner dengan model Naive Bayes menunjukkan bahwa klasifikasi untuk review positif yang sesuai prediksi yaitu **20**. Data review negatif yang termasuk ke dalam prediksi positif yaitu **80**. Data review positif yang termasuk ke dalam prediksi negatif yaitu **17** dan data review negatif yang termasuk ke dalam prediksi positif yaitu **83**.

Berikut merupakan persamaan model *Confusion Matrix* untuk menghitung akurasi pada percobaan Naive Bayes:

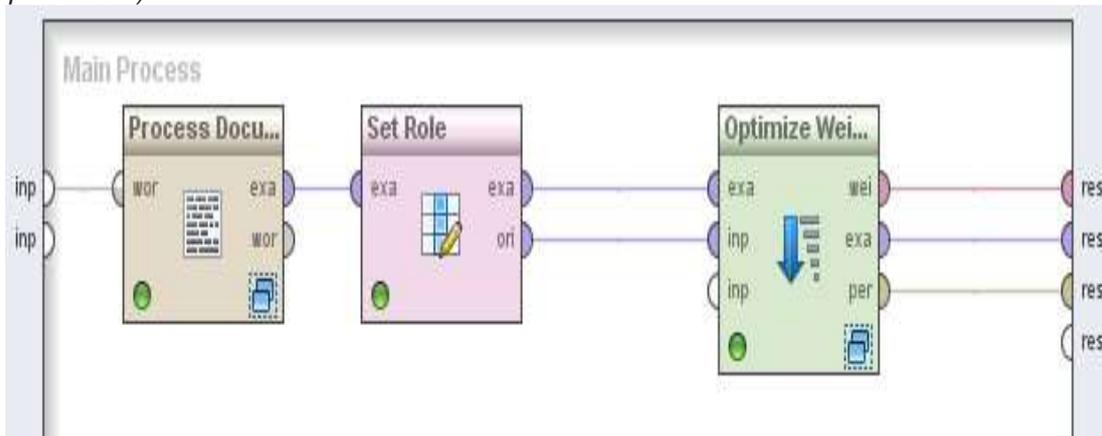
Tabel 2 Confusion Matrix dengan Model Naive Bayes

Accuracy : 51,50% +/- 3.20% (mikro: 51.50%)			
	True Positif	True Negatif	Class Precision
Prediksi Positif	20	17	80.88%
Prediksi Negatif	80	83	81.12%
Class Recall	20.00%	83.00%	

Sumber: Hasil Penelitian (2018)

$$Acc (Accuracy) = \frac{TP+TN}{TP+TN+FP+FN} = \frac{20+83}{20+17+80+83} = \frac{103}{200} = 0.515$$

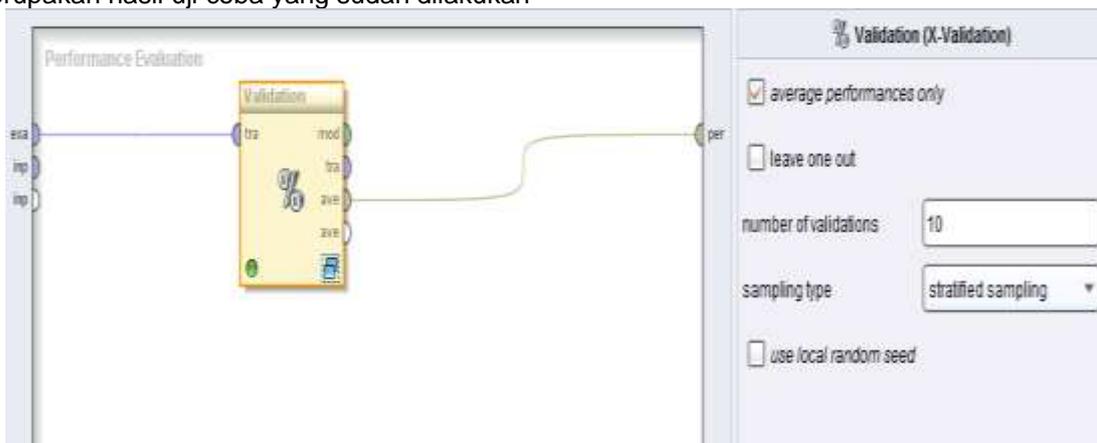
Kemudian peneliti melakukan uji coba pada data latih yang sama dengan menambahkan *PSO*. Berikut hasil tampilan Naive Bayes dioptimasi *PSO (Particle Swarm Optimization)*



Sumber: Hasil Penelitian (2018)

Gambar 7. Model *Training Naive Bayes Optimasi Particle Swarm Optimization*

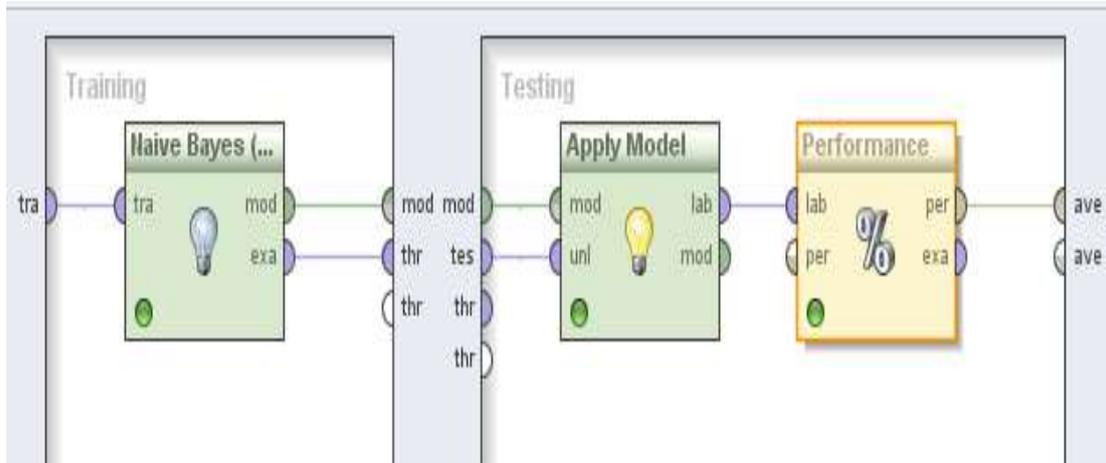
Untuk Validation, peneliti menggunakan *X-Validation* dengan *10 number of sampling* atau *ten cross X-Validation* serta *type* sampling yang digunakan *stratified sampling*. Berikut merupakan hasil uji coba yang sudah dilakukan



Sumber: Hasil Penelitian (2018)

Gambar 8. Penggunaan *X Validation* dalam Uji Data *Training Naive Bayes Optimasi PSO*

Untuk data training *Naive Bayes Optimasi PSO* menggunakan *Performance (Clasification)*. Klasifikasi disini digunakan untuk melakukan klasifikasi terhadap label positif dan negatif. Adapun bentuknya disajikan pada gambar 9.



Sumber: Hasil Penelitian (2018)

Gambar 9. Data Uji Training dan Testing Naive Bayes Optimasi PSO

Hasil pengujian data training metode Naive Bayes berbasis Particle Swarm Optimization menggunakan *Set Role* yang berfungsi untuk menentukan field pada kelas kemudian diberikan optimasi menggunakan *Particle Swarm Optimization* agar akurasi yang dihasilkan lebih tinggi. Pengukuran akurasi tersebut, akan dijabarkan melalui Kurva ROC dan Confusion Matrix disajikan pada tabel 3.

Tabel 3 Confusion Matrix dengan Model Naive Bayes Optimasi PSO

Accuracy : 93,50% +/- 3.91% (mikro: 93.50%)			
	True Positif	True Negatif	Class Precision
Prediksi Positif	93	6	93,94%
Prediksi Negatif	7	94	93,07%
Class Recall	93.00%	94.00%	

Sumber: Hasil Penelitian (2018)

$$Acc (Accuracy) = \frac{TP+TN}{TP+TN+FP+FN} = \frac{93+94}{93+94+6+7} = \frac{187}{200} = 0.935$$

Berikut merupakan hasil uji coba metode Naive Bayes Optimasi PSO dengan menghasilkan akurasi sebesar 93,50%, naik sebesar +/- 40,00% dari ujicoba metode Naive Bayes.

accuracy: 93.50% +/- 3.91% (mikro: 93.50%)			
	true negatif_furn	true positif_furn	class precision
pred. negatif_furn	93	6	93.94%
pred. positif_furn	7	94	93.07%
class recall	93.00%	94.00%	

Sumber: Hasil Pengolahan Data (2018)

Gambar 10. Hasil Akurasi Naive Bayes Optimasi PSO

Analisa dan Validasi Model, berdasarkan penelitian yang telah dilakukan melalui ujicoba data latih dan training terhadap review furniture. Berikut peneliti rangkum hasil penelitian Naive Bayes Optimasi PSO (*Particle Swarm Optimization*)

Tabel 4. Hasil Eksperimen Naive Bayes dan Naive Bayes Optimasi PSO

	Naive Bayes	Naive Bayes + PSO
Accuracy	51,50%	93,50%
AUC	0,5	0,9

Sumber: Hasil Penelitian (2018)

Penjelasan tabel di atas menunjukkan bahwa penerapan PSO dalam metode Naive Bayes mampu meningkatkan akurasi yang signifikan. Pengujian sebelumnya yaitu Naive Bayes menghasilkan akurasi sebesar 51.50% dan AUC sebesar 0.5 sedangkan saat penambahan PSO hasil akurasi meningkat menjadi 93.50% dan AUC sebesar 0.9.

Penulis mengusulkan model dalam review furniture dengan menerapkan Naive Bayes dan berbasis Particle Swarm Optimization, pengujian pada Naive Bayes termasuk ke dalam kategori failure accuracy namun setelah dilakukan optimasi dengan Particle Swarm Optimization, nilai akurasi yang dihasilkan meningkat menjadi *Excellent Classification*. Sedangkan algoritma Naive Bayes pada saat pengujian data dihasilkan akurasi yang cukup baik, namun penulis mencoba meningkatkan akurasi pada hasil perhitungannya dengan menggunakan optimasi Particle Swarm Optimization, pada optimasi PSO ini, percobaan population size sangat ditentukan agar akurasi yang dihasilkan meningkat sehingga terbentuklah struktur algoritma yang ideal dalam memecahkan masalah klasifikasi teks.

4. Kesimpulan

Kemudahan akan mendapatkan review secara otomatis dengan menggunakan algoritma bagi pembeli merupakan salah satu manfaat dari penelitian ini sehingga pembeli dapat menentukan pilihan barang secara cepat dan tepat sebelum membeli barang yang diinginkan terutama furniture. Hasil akurasi penelitian yang telah dilakukan pada metode Naive Bayes review furniture yaitu 51,50% dan AUC sebesar 0,5 sedangkan pada saat penambahan Particle Swarm Optimization pada dataset training Naive Bayes dihasilkan akurasi sebesar 93,50% dan AUC sebesar 0,9. Hal ini membuktikan bahwa PSO merupakan salah satu teknik optimasi untuk seleksi fitur dan penambahan bobot atau meningkatkan akurasi.

Referensi

- Al-ghuribi SM, Alshomrani S. 2013. Algorithms for Arabic and English Languages.
- Aulianita R. 2015. Komparasi Algoritma Naive Bayes dan K-Nearest Neighbors Pada Sentiment Analysis Review Restaurant. Thesis M.Kom STMIK Nusa Mandiri Jakarta 5–27.
- Bijalwan V, Kumar V, Kumari P, Pascual J. 2014. KNN based machine learning approach for text and document mining. *Int. J. Database Theory Appl.* 7: 61–70.
- Indrayuni E. 2016. Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization.
- Kristiyanti DA, Wahyudi M. Feature Selection Based on Genetic Algorithm, Particle Swarm Optimization and Principal Component Analysis for Opinion Mining Cosmetic Product Review.
- Liu R, Chen Y, Jiao L, Li Y. 2014. A particle swarm optimization based simultaneous learning framework for clustering and classification. *Pattern Recognit.* 47: 2143–2152.
- Moraes R, Valiati JF, Gavião Neto WP. 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Syst. Appl.* 40: 621–633.
- Raschka S. 2014. Naive Bayes and Text Classification I - Introduction and Theory. 1–20.

- Rasjid ZE, Setiawan R. 2017. Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques. *Procedia Comput. Sci.* 116: 107–112.
- Tjahyanto A, Sisehaputra B. 2017. The Utilization of Filter on Object-based Opinion Mining in Tourism Product Reviews. *Procedia Comput. Sci.* 124: 38–45.
- Vidhya KA, Aghila G. 2010. A Survey of Naïve Bayes Machine Learning approach in Text Document Classification. *Int. J. Comput. Sci. Inf. Secur.* 7: 206–211.