

# Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization

Retno Sari

STMIK Nusa Mandiri Jakarta  
e-mail: retno.rnr@nusamandiri.ac.id

## Abstrak

Terdapatnya aplikasi yang memudahkan untuk mengetahui ulasan dari suatu tempat atau makanan membuat pembaca dengan mudah menentukan tempat untuk mereka berwisata kuliner. Ulasan yang diberikan terdiri dari ulasan positif dan ulasan negatif. Algoritma Naive Bayes berbasis Particle Swarm optimization dilakukan untuk mengetahui apakah terdapat peningkatan akurasi. Dataset yang digunakan berupa review restoran yang dibagi menjadi 2 class yaitu class positif dan class negatif, data diujikan menggunakan 10 Fold Cross Validation. Analisis sentimen review restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization menghasilkan akurasi sebesar 82.45%. Hasil ini lebih baik dibandingkan dengan menggunakan algoritma Naive Bayes saja yang menghasilkan akurasi sebesar 74.34%.

*Kata Kunci: analisis sentimen, review restoran, naive bayes, Particle Swarm Optimization.*

## Abstract

*The existence of an application that makes it easy to find out reviews from a place or food makes the reader easily determine the place for their culinary tour. The reviews provided consist of positive reviews and negative reviews. The Naive Bayes algorithm based on Particle Swarm optimization is done to find out whether there is an increase in accuracy. The dataset used is a restaurant review which is divided into 2 classes, positive class and negative class, the data is tested using 10 Fold Cross Validation. Sentiment analysis of restaurant reviews using Particle Swarm Optimization based on Naive Bayes Algorithm resulted in an accuracy of 82.45%. This result is better than using the Naive Bayes algorithm which produces an accuracy of 74.34%.*

**Keywords:** *sentiment analysis, restaurant review, Naive Bayes, Particle Swarm Optimization*

## 1. Pendahuluan

Terdapatnya aplikasi yang memudahkan para pecinta kuliner untuk mencari lokasi dan makanan yang sesuai harapan, membuat pecinta kuliner dengan mudah menemukan tempat dan makanan yang diinginkan dimanapun mereka berada. Aplikasi yang ada menyediakan foto, lokasi, ulasan konsumen dan rating yang diberikan oleh konsumen.

Review dan rating yang diberikan membantu pelanggan lain untuk mengevaluasi usaha atau servis yang di berikan dan mereka dapat menentukan pilihan (Govindarajan, 2014).

Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan cara otomatis mengklasifikasikan ulasan

pengguna menjadi pendapat positif atau negatif (Zhang, Ye, Zhang, & Yijun, 2011).

Terdapat beberapa penelitian sebelumnya mengenai review restoran diantaranya, penelitian yang dilakukan oleh Zhang, dkk pada tahun 2011 dengan judul klasifikasi sentimen pada review restoran di internet yang di tulis dalam bahasa Canton menggunakan pengklasifikasi Naive Bayes dan Support Vector Machine (Zhang, Ye, Zhang, & Yijun, 2011). Penelitian yang dilakukan oleh Muthia pada tahun 2017 dengan judul analisis sentimen pada review restoran dengan teks bahasa Indonesia menggunakan algoritma Naive Bayes (Muthia, 2017).

Penelitian yang dilakukan oleh Zhang, dkk pada penelitiannya

menggunakan fitur bigrams( 2 suku kata) dan trigrams(3 suku kata), pada penelitian ini lebih baik menggunakan fitur bigrams dibandingkan dengan fitur unigrams dan trigrams dan algoritma naive bayes lebih baik dibandingkan dengan SVM (Zhang, Ye, Zhang, & Yijun, 2011).

Pada penelitian di tahun 2017, Muthia meneliti menggunakan penggabungan metode pemilihan fitur Genetic Algorithm dengan menggunakan fitur trigram (3 suku kata) mengalami peningkatan akurasi klasifikasi mencapai 4% dibandingkan sebelum menggunakan metode pemilihan fitur (Muthia, 2017).

Bayes merupakan salah satu klasifikasi probabilistik yang sederhana yang didasarkan pada teorema Bayes dengan asumsi naives yang kuat (Ahmed, Guan, & Chung, 2014). Keuntungan dari klasifikasi Bayes yaitu bahwa Naive Bayes hanya membutuhkan jumlah yang kecil untuk data training untuk memperkirakan parameter yang dibutuhkan untuk klasifikasi (Korada, Kumar, & Deekshitulu, 2012). Namun Naive Bayes memiliki kekurangan yaitu sangat sensitif dalam pemilihan fitur (Chen, Tian, & Qu, 2009).

PSO merupakan teknik optimasi yang sangat mudah diterapkan dan terdapat beberapa parameter untuk dimodifikasi (Basari, Hussin, Ananta, & Zeniarja, 2012).

Pada penelitian ini untuk mengetahui seberapa pengaruhnya Particle Swarm Optimization pada algoritma Naive Bayes untuk analisis sentimen review restoran dalam meningkatkan akurasi.

Tujuan penelitian ini menggunakan Particle Swarm Optimization dalam meningkatkan akurasi pada analisis sentimen review restoran menggunakan algoritma Naive Bayes.

## 2. Metode Penelitian

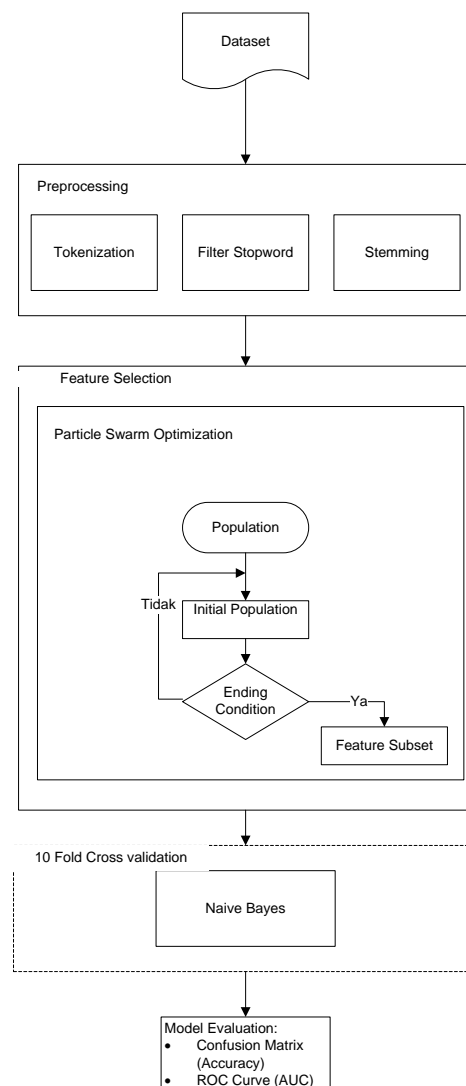
Metode penelitian yang penulis lakukan adalah metode eksperimen, dengan tahapan berikut:

- a. Pengumpulan Data  
Pengumpulan data menggunakan data review restoran yang didapat dari web yelp yang terdiri dari 100 review positif dan 100 review negatif.
- b. Pengolahan Data  
Dataset ini dalam tahap *preprocessing* harus melalui 3 proses, yaitu:
  - a) *Tokenization*

Pada proses ini, kata yang memiliki tanda baca dan symbol yang bukan huruf dihilangkan.

- b) *Filter Stopword*  
Proses penghapusan atau pembuangan kata-kata yang sering ditampilkan dalam dokumen.
- c) *Stemming*  
Proses pembuangan prefix dan suffix sehingga membentuk menjadi kata dasar.

- c. Metode yang diusulkan  
Untuk mengetahui pengaruh pengklasifikasi Naive Bayes berbasis Particle Swarm Optimization dalam analisis sentimen review restoran. Model yang diusulkan pada penelitian ini dapat dilihat dari gambar 1.



Gambar 1. Model yang di usulkan  
Sumber: Sari (2018)

- 1) Algoritma Naive Bayes  
Bayesian klasifikasi adalah pengklasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *Decision Tree* dan *Neural Network*. *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan kedalam *database* dengan data yang besar (Kusrini & Luthfi, 2009).

Tahapan dalam algoritma Naive Bayes, sebagai berikut (Han & Kamber, 2006):

- a) Perhatikan D adalah *record training* dan ditetapkan label-label kelasnya dan
- b) masing-masing *record* dinyatakan n atribut (n field)  $X = (X_1, X_2, \dots, X_n)$
- c) Misalkan terdapat m kelas  $C_1, C_2, \dots, C_m$
- d) Klasifikasi adalah diperoleh maximum posteriori yaitu maximum  $P(C_i|X)$
- e) Ini diperoleh dari teorema Bayes

$$P(C_i|X) = P(X|C_i)P(C_i)P(X) \dots \dots \dots (2.1)$$

Karena  $P(X)$  adalah konstan untuk semua kelas, hanya perlu dimaksimumkan

$$P(C_i|X) = P(X|C_i)P(C_i) \dots \dots \dots (2.2)$$

- 2) Particle Swarm Optimization (PSO)  
Particle Swarm Optimization merupakan sebuah evolusi algoritma yang efisien. Particle Swarm Optimization adalah populasi berdasarkan metode optimasi stokastik berbasis populasi yang didasarkan pada perilaku burung flock. Particle Swarm optimization dimulai dengan satu set acak dan mencapai optima global dengan memperbaharui generasinya (Gupta, Reddy, Shweta, & Ekbal, 2015).

- d. Eksperimen dan Pengujian Metode  
Proses eksperimen pada penelitian ini menggunakan RapidMiner 5.3 untuk pengujian model dilakukan dengan menggunakan dataset. Tahapan pengujian untuk analisis sentimen review restoran sebagai berikut:
  - 1) Menyiapkan dataset untuk eksperimen yang sudah diketahui classnya
  - 2) Mendesain arsitektur algoritma Naive Bayes pada analisis sentimen review restoran dan mendesain arsitektur menggunakan Particle Swarm Optimization untuk optimasi pada algoritma Naive Bayes pada analisis sentimen review restoran
  - 3) Melakukan *training* dan *testing* terhadap metode Naive Bayes dan penggunaan algoritma Naive Bayes berbasis Particle Swarm Optimization .
- e. Evaluasi dan Validasi Data  
Validasi dilakukan menggunakan *10 fold cross validation*. Dimana dengan menggunakan teknik ini dengan membagi secara acak ke dalam tiap bagian dimana terdiri dari 10 bagian an setiap bagian akan dilakukan proses klasifikasi. Sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC (*Receiver Operating Characteristics*) untuk mengukur nilai AUC.

Tabel 1 Confusion Matrix

Classification Observed Class	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 -TFN)

Sumber : (Gorunescy, 2011)

### 3. Hasil dan Pembahasan

#### 3.1. Hasil

Data yang digunakan dalam pengklasifikasian teks terdiri dari 100 review positif dan 100 review negatif. Sebelum diklasifikasikan, data tersebut harus melewati beberapa tahapan proses agar dapat diklasifikasikan ke proses

selanjutnya, berikut ini adalah tahapan prosesnya:

- 1) Pengumpulan data  
Data review restoran positif dijadikan didalam satu folder dengan nama review positif dan data review restoran negatif dijadikan didalam satu folder dengan nama negatif.
- 2) Pengolahan Awal Data  
Proses pengolahan awal data terdiri dari *tokenization*, *filter stopwords* dan *steeming*.
- 3) Klasifikasi  
Dalam proses ini review restoran diklasifikasikan untuk menentukan class untuk setiap reviewnya, *class* review restoran terbagi dua yaitu *class* positif dan *class* negatif. Penentuan class untuk setiap review ditentukan melalui perhitungan probabilitas dari rumus algoritma Naive Bayes.

Data yang diujikan menggunakan 10 *fold cross validation*. Dimana dengan menggunakan teknik ini dengan membagi secara acak ke dalam tiap bagian dimana terdiri dari 10 bagian an setiap bagian akan dilakukan proses klasifikasi.

### 3.2. Pembahasan

Model klasifikasi teks pada review restoran membuat pembaca dapat dengan mudah mengidentifikasi mana review yang yang positif dan negatif. kata-kata dipisahkan lalu diberikan bobot pada masing-masing kata untuk review restoran yang sudah ada. Kata yang ada dilihat dari hubungan dengan sentimen yang sering muncul dan memiliki bobot paling tinggi untuk setiap dokumennya. Dengan demikian dapat diketahui review tersebut positif atau negatif.

Hasil pengujian model ini dibahas melalui *confusion matrix* yang menunjukkan seberapa baik model yang terbentuk. pengujian dengan menggunakan algoritma Naive Bayes memiliki akurasi 74.34%. Akurasi tersebut masih perlu ditingkatkan. PSO ditambahkan pada algoritma Naive Bayes untuk analisis sentimen review restoran dan akurasi meningkat menjadi 82.45%.

Dengan menggunakan algoritma Naive Bayes saja menghasilkan akurasi sebesar 74.34% dan nilai AUC 0.985. *Confusion matrix* algoritma Naive Bayes sebelum menambahkan PSO.

Tabel 2 *Confusion Matrix* Algoritma Naive Bayes

Accuracy : 74.34% +/- 10.67% (mikro: 74.37%)		
	True negatif	True Positif
Pred Negatif	97	48
Pred Positif	3	51

Sumber : Sari (2018)

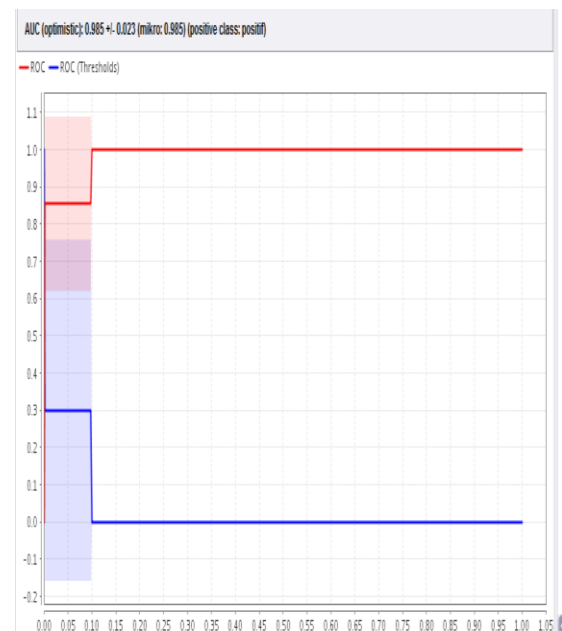
Sedangkan algoritma Naive Bayes berbasis Particle Swarm Optimization menghasilkan akurasi sebesar 82.45% dan nilai AUC 0.993. *Confusion matrix* algoritma Naive Bayes sesudah menambahkan PSO.

Tabel 3 *Confusion Matrix* Algoritma Naive Bayes berbasis PSO

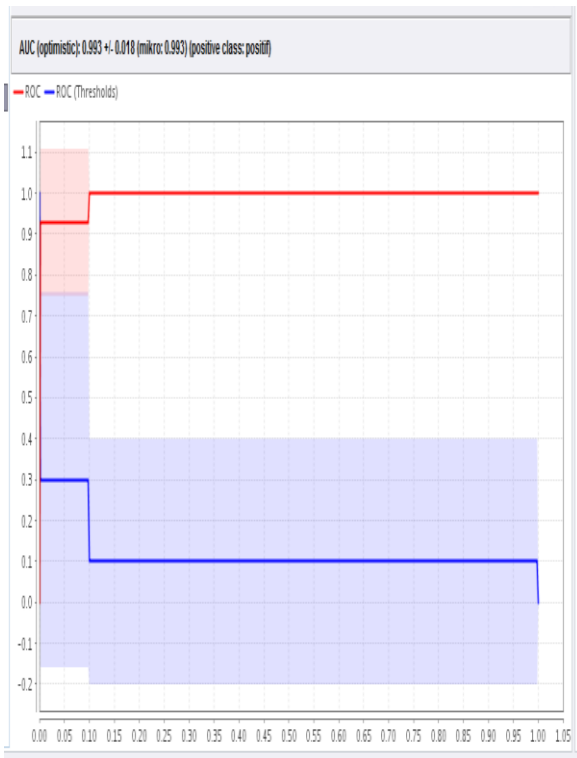
Accuracy : 82.45% +/- 10.51% (mikro: 82.41%)		
	True negatif	True Positif
Pred Negatif	98	33
Pred Positif	2	66

Sumber : Sari (2018)

Data uji diatas akan dinilai hasil prediksi dengan menggunakan grafik ROC. Berikut grafik ROC untuk algoritma Naive Bayes yang dapat dilihat pada gambar 2. Dan grafik ROC untuk algoritma Naive Bayes berbasis Particle Swarm Optimization dapat dilihat pada gambar 3



Gambar 2. Kurva ROC model Naive Bayes  
Sumber : Sari (2018)



Gambar 3. Kurva ROC model Naive Bayes berbasis PSO  
Sumber : Sari (2018)

Berdasarkan seluruh hasil eksperimen dapat disimpulkan bahwa PSO telah dapat meningkatkan kinerja algoritma Naive Bayes pada analisis sentimen review restoran. Perbandingan model algoritma Naive Bayes sebelum dan sesudah menggunakan PSO dapat di lihat pada tabel 4.

Tabel 4 Perbandingan algoritma Naive Bayes sebelum dan sesudah menggunakan PSO

	Algoritma Naive Bayes	Algoritma Naive Bayes + PSO
Review positif	51	66
Review Negatif	97	98
	74.34%	82.45%

#### 4. Kesimpulan

Kesimpulan yang diperoleh setelah melakukan eksperimen untuk memperoleh akurasi yang lebih akurat pada analisis sentimen review restoran menggunakan algoritma Naive Bayes berbasis Particle

Swarm memiliki akurasi sebesar 82.45%, sedangkan analisis sentimen review restoran dengan menggunakan algoritma Naive Bayes didapatkan akurasi sebesar 74.34%.

Dapat dilihat dari hasil eksperimen algoritma Naive Bayes berbasis Particle Swarm Optimization terjadi peningkatan yang sangat signifikan, meningkatkan akurasi mencapai 8.11%.

#### Referensi

- Ahmed, I., Guan, D., & Chung, T. (2014). SMS Classification based on Naive Bayes classifier and Apriori Alhoritm Frequent Itemset. *International Journal of Machine Learning and Computing Vol 4 No 2*.
- Basari, A. S., Hussin, B., Ananta, I. G., & Zeniarja, J. (2012). Opinion Mining of Movie Revie using Hybrid Method of SUpport Vector Machine and Particle Swarm Optimization. *Malaysian Technical Universities Conference on Engineering & Technology 2012, MUCET 2012* (hal. 453-462). Malaysia: Elsevier.
- Chen, J. H., Tian, S., & Qu, Y. (2009). Feature Selection for text classification with Naive Bayes. *Expert Systems with Application 36* (3), 5432-5435.
- Gorunesy, F. (2011). *Data Mining Concept Model Technique*. Verlag Berlin Heidelberg: Springer.
- Govindarajan, M. (2014). Sentimen Analysisi of Restaurant Review Using Hybrid Classification. *2nd IRF International Conference*. India.
- Gupta, D. K., Reddy, K. S., Shweta, & Ekbal, A. (2015). PSO-ASent: Feature Selection Using Particle Swarm Optimization for Aspect Based Sentiment Analysis. *Natural Language Processing and Information Systems Volume 91(1)* (hal. 220-233). Switzerland: Springer International.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques*. San Francisco: Elsevire.
- Korada, N., Kumar, N., & Deekshitulu, Y. (2012). Implementation of Naive Bayesian Classifier and Ada-Boost Algorithm Using Maize Expert

- System. *International Journal of Information Sciencesan Techniques*.
- Kusrini, & Luthfi, E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Muthia, D. A. (2017). Analisis Sentimen pada Review Restoran Dengan teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, 39-45.
- Zhang, Z., Ye, Q., Zhang, Z., & Yijun, L. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 7674-7682.