

## **Implementasi *Particle Swarm Optimization* (PSO) Pada *Analysis Sentiment Review* Aplikasi Halodoc Menggunakan *Algoritma Naïve Bayes***

Nuzuliarini Nuris<sup>1</sup>, Eka Rini Yulia<sup>2</sup>, Kusmayanti Solecha<sup>3</sup>  
Universitas Bina Sarana Informatika<sup>1,3</sup>, Universitas Nusa Mandiri<sup>2</sup>  
eka.erl@nusamandiri.ac.id

### **Abstrak**

Kesehatan sangat penting bagi manusia, jika mengalami gejala atau merasakan sakit maka sewajarnya kita memeriksakan kesehatan dan mendatangi rumah sakit atau klinik, namun jika tidak memungkinkan untuk keluar rumah maka aplikasi konsultasi kesehatan secara online dianggap dapat membantu. Namun sebelum dapat menggunakan dan memanfaatkan aplikasi tersebut perlu diketahui review dari konsumen berdasarkan opini positif dan opini negatif. Penelitian ini menerapkan algoritma naive bayes untuk melakukan klasifikasi teks dan memilih fitur seleksi particle swarm optimization untuk mendukung peningkatan akurasi yang didapatkan. Evaluasi dan validasi klasifikasi dilakukan menggunakan confusion matriks dan Kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi yang sebelumnya 88,50% dan AUC 0.535, mengalami peningkatan menjadi 90.50% dan AUC 0,525. Dapat diambil kesimpulan bahwa pemilihan fitur seleksi particle swarm optimization berhasil meningkatkan akurasi.

**Kata Kunci** : Fitur seleksi, naïve bayes, particle swarm optimization.

### **Abstract**

*Health is very important for humans, if you experience symptoms or feel pain, it is appropriate for us to have a health check and go to a hospital or clinic, but if it is not possible to leave the house, an online health consultation application is considered to be helpful. But before you can use and take advantage of these applications, it is necessary to know reviews from consumers based on positive opinions and negative opinions. This study applies the Naive Bayes algorithm to perform text classification and selects the particle swarm optimization selection feature to support the increased accuracy obtained. Classification evaluation and validation are performed using confusion matrix and ROC curves. The results showed an increase in accuracy previously 88.50% and AUC 0.535, increased to 90.50% and AUC 0.525. It can be concluded that the selection of the particle swarm optimization feature has succeeded in increasing the accuracy.*

**Keywords**: selection features, naïve bayes, particle swarm optimization.

## Pendahuluan

Kesehatan merupakan hal yang terpenting dalam kehidupan manusia untuk dapat menjalani hari-hari dengan ringan dan tanpa beban yang mengganjal dari kesehatan tubuh, jika kita merasakan gejala sakit atau telah merasakan sakit pada tubuh maka sebaiknya kita mengobati sakit tersebut. Sudah tidak diragukan lagi bahwa pergi ke rumah sakit atau ke klinik untuk menemui dokter dapat mengobati sakit yang kita derita, namun bagaimana jika kita tidak dapat memiliki waktu luang yang cukup atau jarak ke rumah sakit maupun klinik terhitung jauh apalagi dengan kondisi saat ini yang memaksa kita untuk tetap berada di rumah dan tidak mengunjungi tempat yang ramai. Saat ini telah ada aplikasi yang dapat memungkinkan kita untuk melakukan *Telekonsultasi Medis* melalui sebuah smartphone menggunakan aplikasi Halodoc.

Terdapat beberapa penelitian yang telah melakukan analisis sentimen review atau opini dari konsumen diantaranya, (Hendra, 2021) menerapkan Naive Bayes dalam melakukan analisis sentimen review halodoc, (Ernawati, 2016) menerapkan particle swarm optimization untuk seleksi fitur algoritma Naive Bayes dalam melakukan analisa sentimen pada review perusahaan penjualan online, (Wati, 2020) menerapkan naive bayes dan particle swarm optimization untuk mengklasifikasikan berita hoax pada media sosial, (Indrayuni, 2019) melakukan review produk kosmetik untuk teks bahasa indonesia dengan menerapkan algoritma Naive Bayes dengan fitur N-Gram, (Chandra et al., 2016) melakukan klasifikasi berita lokal Radar Malang dengan menambah fitur N-gram,

Naive Bayes Classification memiliki beberapa keunggulan seperti algoritma klasifikasi yang sederhana namun memiliki akurasi yang tinggi, (Indrayuni, 2019) (Wati, 2020) (Hendra, 2021), meskipun begitu Naive Bayes juga memiliki kekurangan sangat sensitive dalam pemilihan fitur yang dapat

mempengaruhi hasil akurasi (Wati, 2020).

Metode klasifikasi berbasis fitur yang dikembangkan dalam penelitian-penelitian tersebut menghasilkan akurasi yang baik (Laksana Utama, 2018). Particle Swarm Optimization (PSO) adalah algoritma populer dan bionik berdasarkan perilaku sosial yang terkait dengan burung berkelompok untuk masalah optimasi (Yan & Jiao, 2016). Particle Swarm Optimization dapat meningkatkan akurasi pengklasifikasi Naive Bayes. (Ernawati, 2016) (Wati, 2020) (Hendra, 2021) Penggunaan N-Gram mampu menambah jenis kata sebelum masuk ke proses stemming akan sangat membantu proses klasifikasi Naive Bayes menjadi lebih efektif dan akurat. (Chandra et al., 2016)

Dengan demikian pada penelitian ini akan menerapkan Naive Bayes Classification dengan optimasi particle swarm optimization dengan menggunakan fitur N-Gram pada proses stemming analisa sentimen terhadap review konsumen yang telah melakukan *Telekonsultasi Medis* pada aplikasi Halodoc.

### 1. Analisis Sentiment

Analisis sentiment merupakan bentuk pengolahan bahasa alami yang berguna melacak mood masyarakat sesuai dengan topik atau produk tertentu. Analisis sentimen yang dapat disebut sebagai tambang pendapat, terlibat dalam membangun sistem sebagai media yang mengumpulkan dan meneliti pendapat tentang produk yang dibangun dalam komentar maupun ulasan. Beberapa cara analisis sentiment dapat dipergunakan. Misalnya, pada bidang pemasaran dapat membantu keputusan g keberhasilan sebuah iklan atau peluncuran produk baru, menentukan versi produk atau jasa yang sedang tren hingga mengidentifikasi demografi suka atau tidak suka terhadap fitur tertentu (Salappa, A., Doumpos, M., & Zopounidis, 2007).

2. Algoritma Naïve Bayes  
Naive Bayes merupakan kasifikasi algoritma sederhana tetapi dapat menghasilkan akurasi paling tinggi (Rodiyanah & Winarko, 2012).
3. Particle Optimization Swarm (PSO), *Particle Swarm Optimization* ibarat sekelompok burung yang mencari kebutuhan makanannya pada suatu daerah. Mereka tidak tahu keberadaan dimana letaknya makanan berada, namun mereka dapat mengetahui jauh dekatnya jarak makanan itu berada, maka mereka membuat strategi terbaik agar bisa menemukan keberadaan makanan itu dengan cara mengikuti satu burung paling dekat dengan makanan tersebut (Salappa, A., Doumpos, M., & Zopounidis, 2007).
4. Evaluasi dan Validasi Klasifikasi
  - a. Confusion Matrix  
Terdapat beberapa metode yang bisa digunakan dalam memvalidasi suatu model berdasarkan data yang didapat, seperti holdout, cross validation, random sub-sampling, dan lain-lain. *Confusion matrix* mampu menghasilkan keputusan dalam proses training dan testing, juga dapat memberikan penilaian dalam performan cek klasifikasi objek dengan hasil benar atau salah (Gorunescu, 2011).

Berikut adalah persamaan model Confusion Matrix:

Nilai *Accuracy* merupakan proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

## METODE

### 1. Pengumpulan Data

Penelitian ini mengambil data yang berasal dari situs web yaitu dan banyak review yang tersedia di situs tersebut mengenai berbagai macam komenar tentang larangan mudik. Peneliti

menggunakan data sebanyak 200 data yang terdiri dari 100 data positif dan 100 data negative.

### 2. Pengolahan data awal

Dataset yang digunakan sebanyak 200 data, 100 review positif dan 100 review negative. Dataset ini dalam tahap Processing harus dengan 3 tahap yaitu :

- a. Tokenization  
Tokenize.
- b. Stopword Removal
- c. Stemming

### 3. Metode yang diusulkan

Metode yang peneliti gunakan untuk klasifikasi ini yaitu dengan menggunakan Particle Swarm Optimization (PSO) untuk meningkatkan akurasi.

### 4. Eksperimen dan Pengujian Metode

Peneliti melakukan proses pengujian menggunakan RapidMiner. Data training yang digunakan adalah dataset review dari situs dan dimana dataset telah dikelompokkan menjadi 2 terdiri dari 100 dataset review positif, dan 100 dataset review negative.

### 5. Evaluasi dan Validasi Hasil Penelitian

Pada penelitian ini validasi yang dilakukan dengan menggunakan data 10fold Validation. Akurasi dapat diukur dengan confusion matrix yaitu membandingkan akurasi Naïve Bayes sebelum menggunakan pemilihan fitur dengan Naïve Bayes dengan Naïve Bayes berbasis Particle Swarm Optimization (PSO) sebagai pemilihan fitur. Kurva ROC digunakan untuk mengukur nilai AUC.

## HASIL

### 1. Pengumpulan Data

Penulis mengumpulkan Dataset yang digunakan berasal dari situs <https://play.google.com/store/apps/details?id=com.linkdokter.halodoc.android&hl=in> sebanyak 200 dataset terdiri

dari 100 dataset review positif dan 100 dataset review negative. Data tersebut merupakan kumpulan teks yang terpisah dalam bentuk dokumen. Ada 2 folder yaitu 1 folder dengan nama positif yaitu kumpulan dari berbagai macam komentar positif dan 1 folder dengan nama negative yaitu kumpulan dari berbagai macam komentar negative.

a. Pengolahan Data Awal

1. Tokenization

Dalam proses awal tokenization setiap kata yang ada di dalam dokumen dihilangkan tanda bacanya.

Tabel 1.1  
Perbandingan kata sebelum dan sesudah tokenization

Sebelum menggunakan tokenization	Ad yg nlp ke no sya katanya dari halodoc,katanya sya mendapatkan hadiah dari halodoc..saya mau konfirmasi apakah benar apk halodoc mengadakan hadiah kalau tidak kenapa dia bisa tau email dan no hp saya ?
Sesudah menggunakan tokenization	Ad yg nlp ke no sya katanya dari halodoc katanya sya mendapatkan hadiah dari halodoc saya mau konfirmasi apakah benar apk halodoc mengadakan hadiah kalau tidak kenapa dia bisa tau email dan no hp saya Ampun Ampun Kalau belum siap ya jangan di munculin di aplikasi nya

Sumber : Peneliti

2. Stem (Snowball)

Proses stem hampir sama dengan tokenization menghilangkan tanda baca, namun pada proses snowball ini akan menghapus kata-kata yang tidak relevan.

Tabel 1.2  
Perbandingan kata sebelum dan sesudah snowball

Sebelum menggunakan snowball	Ad yg nlp ke no sya katanya dari halodoc,katanya sya mendapatkan hadiah dari halodoc..saya mau konfirmasi apakah benar apk halodoc mengadakan hadiah kalau tidak kenapa dia bisa tau email dan no hp saya ?
Sesudah menggunakan snowball	ad yg nlp ke no sya katanya dari halodoc katanya sya mendapatkan hadiah dari halodoc saya mau konfirmasi apakah benar apk halodoc mengadakan hadiah kalau tidak kenapa dia bisa tau email dan no hp saya

Sumber : Peneliti

3. Stopwords

Pada proses ini, proses stopwords mengubah token yang berlebihan menjadi kata dasar, kemudian menghilangkan semua imbuhan tersebut.

Tabel 1.3  
Perbandingan kata sebelum dan sesudah stopwords

Sebelum menggunakan stopwords	Udah 2 kali di telepon mengatas nama kan halodoc, aneh nya penelpon bisa tau email kita selalu bilang kita dapat uang 2 juta
-------------------------------	--

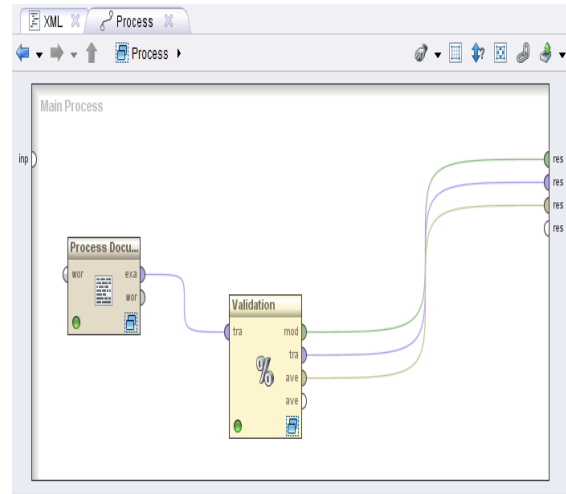
	<p>rupiah, apa aplikasi halodoc tidak memprivasi email sehingga ada saja pihak yg ingin mengambil keuntungan dgn menipu,tolong privasi lbh d perketat</p>
<p>Sesudah menggunakan stopwords</p>	<p>udah kali di telepon mengata nama kan halodoc aneh nya penelpon bisa tau email kita selalu bilang kita dapat uang juta rupiah apa aplikasi halodoc tidak memprivasi email sehingga ada saja pihak yg ingin mengambil keuntungan dgn menipu tolong privasi lbh d perketat</p>

Sumber : Peneliti

**2. Model dengan Metode Klasifikasi Naïve Bayes**

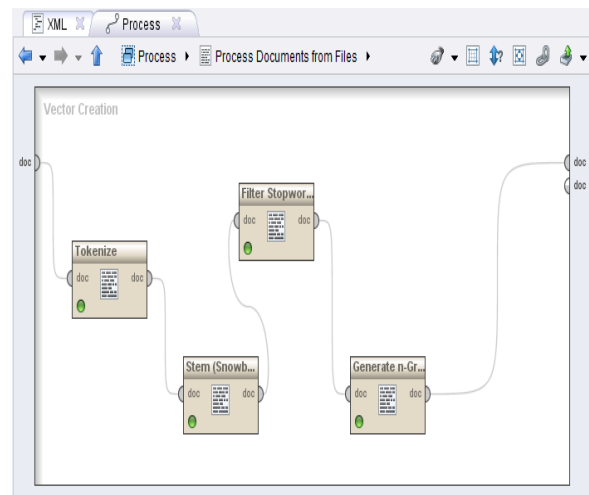
Dalam Proses klasifikasi ini, untuk menentukan kalimat yang beranggotakan class positif dan class negative dengan menggunakan rumus Naïve Bayes. Peneliti hanya menggunakan dataset sebanyak 200 data terdiri dari 100 dataset positif dan 100 dataset negative.

Berikut Proses Pengujian pertama gambar Naïve Bayes menggunakan RapidMiner.



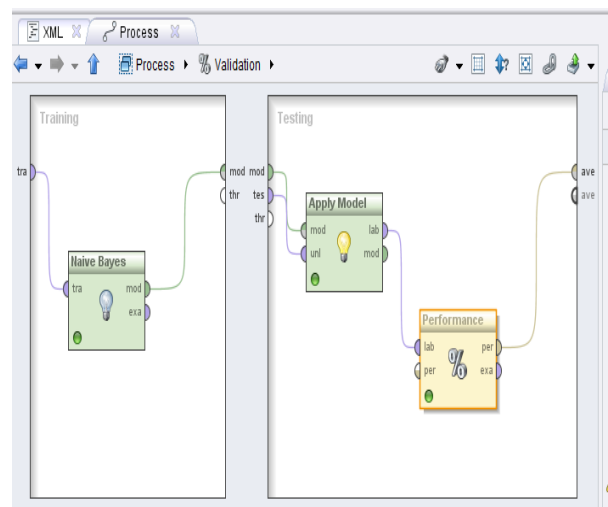
Sumber : Peneliti

Gambar 1.1  
Proses Pemasukan data melalui NB



Sumber : Peneliti

Gambar 1.2  
Proses Document dari File data



Sumber : Peneliti

Gambar 1.3  
Proses Validasi

Hasil akurasi menggunakan algoritma Naïve Bayes dengan fitur N-gram

Tabel 1.4  
Hasil Naïve Bayes

Accuracy : 88.50%+/-5.50% (micro average : 88.50%)			
	True Negatif	True Positif	Class Precision
Pred Negatif	94	17	84.68%
Pred Positif	6	83	93.26%
Class Recall	94.00%	83.00%	

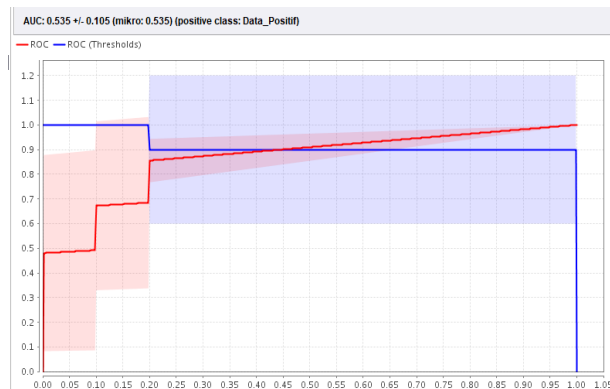
Sumber : Peneliti

Nilai Accuracy dari confusion matrix tersebut adalah sebagai berikut :

$$Accuracy = \frac{(TN+TP)}{(TN+FN+TP+FP)}$$

$$Accuracy = \frac{(94+83)}{(94+6+17+83)}$$

$$Accuracy = \frac{177}{200} = 0.885 = 88.50\%$$



Sumber : Peneliti

Gambar 1.5  
Kurva ROC Naïve Bayes

### 3. Hasil Pengujian Kedua Algoritma Naïve Bayes berbasis Particle Optimization Swarm (PSO) dengan fitur n-gram

Dari hasil pembahasan dengan algoritma Naïve bayes mendapatkan akurasi yang

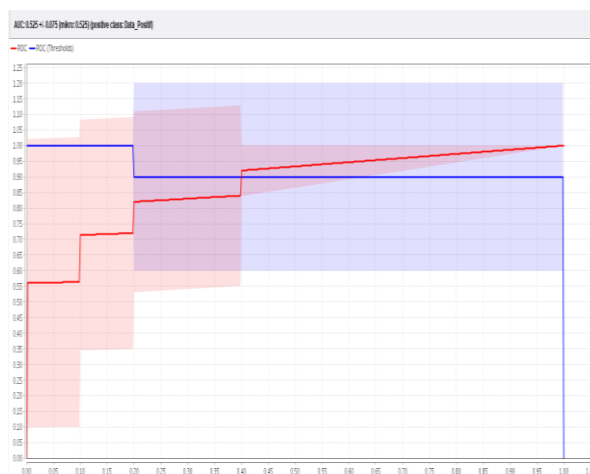
kecil sehingga peneliti menggunakan PSO untuk mengoptimalkan akurasi. Maka hasil yang didapatpun meningkat menjadi 90.50% dengan AUC 0.525.

Tabel 1.5

Hasil akurasi menggunakan algoritma Naïve Bayes berbasis PSO

Accuracy : 90.50%+/-9.34% (micro average : 90.50%)			
	True Negatif	True Positif	Class Precision
Pred Negatif	92	11	89.32%
Pred Positif	8	89	91.75%
Class Recall	92.00%	89.00%	

Sumber : Peneliti



Sumber : Peneliti

Gambar 1.6.

Kurva ROC Naïve Bayes berbasis PSO

### KESIMPULAN

Dari pengolahan data yang sudah dilakukan, penggunaan metode fitur pemilihan PSO dapat meningkatkan akurasi dengan algoritma Naïve Bayes. Data review aplikasi halodoc yang didapat dari playstore dapat diklasifikasikan menjadi data positif dan data negative. Akurasi Naïve Bayes dengan fitur n-gram sebelum menggunakan penggabungan pemilihan fitur mendapatkan hasil 88.50% dengan AUC 0.535. Dan meningkat setelah dioptimalkan dengan PSO nilai akurasinya 90.50%

dengan AUC 0.525. Peningkatan akurasi mencapai 2%. Sehingga penggunaan metode Naïve Bayes berbasis PSO pada klasifikasi review review aplikasi halodoc yang didapat dari playstore terbukti memberikan hasil akurasi yang lebih akurat.

#### DAFTAR PUSTAKA

1. Chandra, D. N., Indrawan, G., & Sukajaya, I. N. (2016). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram. *Jurnal Ilmiah Teknologi Informasi Asia*, 10(1), 11–19.
2. Ernawati, S. (2016). Penerapan Particle Swarm Optimization Untuk Seleksi Fitur Pada Analisis Sentimen Review Perusahaan Penjualan Online Menggunakan Naïve Bayes. *Evolusi*, 4(1), 45–54.
3. Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*.
4. Hendra, A. (2021). Analisis Sentimen Review Halodoc Menggunakan Naïve Bayes Classifier. 6(2), 78–89.
5. Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1), 29–36. <https://doi.org/10.31294/jki.v7i1.1>
6. Laksana Utama, P. K. (2018). Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning. *Widya Duta: Jurnal Ilmiah Ilmu Agama Dan Ilmu Sosial Budaya*, 13(1), 69. <https://doi.org/10.25078/wd.v13i1.436>
7. Rodiyansyah, S. F., & Winarko, E. (2012). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *FMIPA UGM*, 6(1), 91–100.
8. Salappa, A., Doumpos, M., & Zopounidis, C. (2007). *Feature Selection Algorithms in Classification Problems: An Experimental Evaluation. Systems Analysis, Optimization and Data Mining in Biomedicine*. 199–212.
9. Wati, R. (2020). Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 159–164. <https://doi.org/10.33480/jitk.v5i2.1034>