

# IJCIT

(Indonesian Journal on Computer and Information Technology)

Journal Homepage: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit>

---

---

## Analisis Sentiment Masyarakat Menggunakan Penggabungan Algoritma Naive Bayes Dan Particle Swarm Optimization

Sopian Aji<sup>1</sup>, Ina Maryani<sup>2</sup>, Elly Muningsih<sup>3</sup>

<sup>1,2</sup>Sistem Informasi, Universitas Nusa Mandiri  
Jakarta, Indonesia

Email: [sopian.sop@nusamandiri.ac.id](mailto:sopian.sop@nusamandiri.ac.id)<sup>1</sup>, [ina.maryani@nusamandiri.ac.id](mailto:ina.maryani@nusamandiri.ac.id)<sup>2</sup>

<sup>3</sup>Sistem Informasi Kampus Kota Yogyakarta, Universitas Bina Sarana Informatika  
Yogyakarta, Indonesia  
Email: [elly.emh@bsi.ac.id](mailto:elly.emh@bsi.ac.id)

---

### ABSTRAK

IMDb adalah sumber informasi paling populer untuk informasi film, TV, dan selebritas, Tonton trailer, dapatkan jadwal tayang, dan beli tiket untuk film mendatang serta melihat ulasan guna dalam mengambil keputusan yang akan diambil. Dengan menggunakan Aplikasi IMDb informasi akan lebih cepat lagi karena kini informasi ada digenggaman tangan yang telah terinstal pada handphone yang digunakan dengan jumlah lebih dari 100 juta unduhan untuk platform tersebut pada play store. Pada ulasan pengujian kali ini Pengujian akan menggabungkan Algoritma *Naive Bayes* dengan *Particle Swarm Optimization* (PSO) untuk meningkatkan nilai akurasi dan *ROC Curve* yang didapatkan. Data ulasan analisis sentimen yang digunakan 1000 ulasan dalam bahasa indonesia yang masing-masing terdiri dari 500 ulasan negatif dengan sifat yang tidak mendukung serta 500 ulasan positif dengan sifat yang mendukung, dengan sumber data ulasan diambil dari *Google Play* di Aplikasi IMDb. Hasil yang didapat dari kedua model ini yakni Algoritma *Naive Bayes* dengan *Particle Swarm Optimization* (PSO) mendapatkan hasil *accuracy* pengujian terbaik yakni nilai *accuracy*: 80.00% eksperimen dilakukan dengan menggunakan Aplikasi RapidMiner Studio.

**Kata Kunci:** naive bayes, particle swarm optimization, ulasan aplikasi imdb

---

### ABSTRACTS

*IMDb is the most popular source of information for movie, TV and celebrity information, Watch trailers, get showtimes and buy tickets for upcoming movies and view reviews to make informed decisions. By using the IMDb application, information will be even faster because now the information is in the palm of the hand that has been installed on the cellphone used with more than 100 million downloads for that platform on the play store. In this review, the test will combine the Naive Bayes Algorithm with Particle Swarm Optimization (PSO) to increase the accuracy and ROC Curve values obtained. Sentiment analysis review data used 1000 reviews in Indonesian, each of which consisted of 500 negative reviews with unsupportive characteristics and 500 positive reviews with supportive characteristics, with the source of review data taken from Google Play in the IMDb Application. The results obtained from these two models, namely the Naive Bayes Algorithm with Particle Swarm Optimization (PSO) get the best test accuracy results, namely the accuracy value: 80.00% experiments were carried out using the RapidMiner Studio Application.*

**Keywords:** naive bayes, particle swarm optimization, imdb app review



## 1. PENDAHULUAN

Menonton film merupakan aktivitas yang sangat digemari oleh banyak orang, bukan saja sebagai hiburan semata namun melalui film juga seseorang dapat belajar banyak hal (Keni Hasanah Wiguna, 2021)). Tak jarang ketika ingin menonton sebuah film manusia membutuhkan informasi yang valid terkait rekomendasi film yang menarik untuk ditontonnya.

IMDb (*Internet Movie Database*) merupakan media hiburan berupa informasi seperti film, acara televisi, video rumahan atau permainan, biografi kru produksi dan personil, acara internet, termasuk daftar pemeran, trailer movie. Menurut (Prasetyo et al., 2022) tercatat jutaan data film pada platform IMDb ini.

Masalah yang muncul adalah polemic ulasan positif dan negatif terhadap suatu film. Tak jarang masyarakat berkomentar atau memberikan ulasan yang kurang pantas bahkan sampai memberikan kalimat yang kasar dan bersifat ujaran kebencian. Ulasan dari masyarakat pada aplikasi IMDb ini dapat dijadikan bahan analisis sentiment untuk mengetahui pendapat masyarakat tentang aplikasi IMDb.

Analisis sentiment adalah teknik yang digunakan untuk mengetahui informasi sentiment yang bernilai positif, bernilai netral atau bernilai negative pada suatu teks (Legiawati et al., 2022). Analisis sentiment ini banyak dilakukan oleh para peneliti seperti dalam penelitian (Jatmiko, dkk, 2022 ) yaitu tentang analisis sentiment komentar di twitter terkait event formula E di Jakarta. Sejumlah algoritma telah digunakan untuk proses analisis sentiment seperti penggunaan algoritma naïve bayes yang digunakan oleh (Ramadhani & Wahyudin, 2022) untuk analisis jenis vaksin covid-19, selain itu dalam penelitian lainnya yaitu (Pramukti et al., 2022) mengenai kebijakan perpanjangan PPKM dimasa pandemi. Algoritma naïve bayes sangat populer dan sering digunakan oleh para peneliti karena kemampuannya dapat menganalisis data dan menghasilkan nilai akurasi yang tinggi.

Dalam penelitian ini peneliti menggunakan algoritma PSO (*Particle Swarm Optimization*). Dalam (Setiawan et al.,

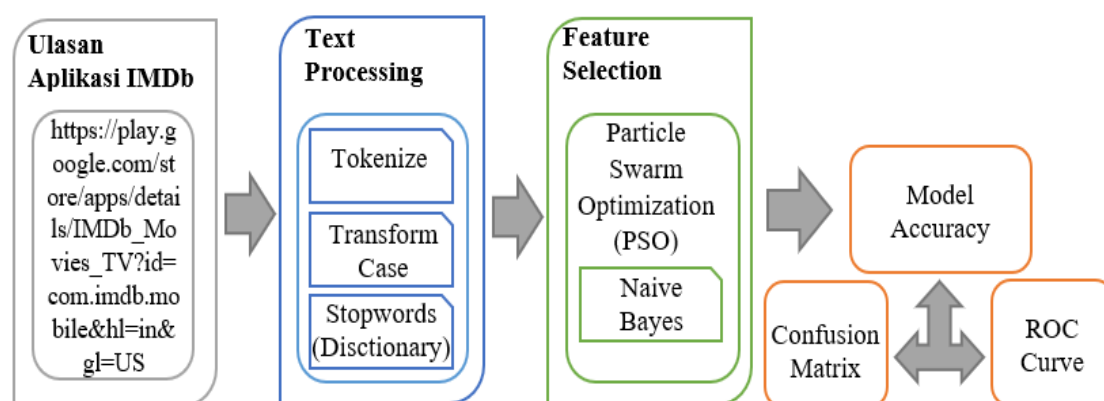
2019) mengatakan bahwa Algoritma PSO cukup baik dibandingkan algoritma genetika lainnya, selain itu melalui algoritma PSO ini dapat mengambil keputusan yang baik. Penggunaan penggabungan kedua metode yakni Naive bayes dan PSO juga dilakukan oleh (Sudriyanto et al., 2021) untuk memprediksi mahasiswa yang lulus tepat waktu dan terbukti dapat menghasilkan nilai akurasi yang tinggi. Berikutnya dalam penelitian lainnya juga PSO dianggap lebih unggul untuk menentukan posisi robot (Istighfarin et al., 2020). Selain itu masi berbicara mengenai keunggulan PSO yang terbukti dalam penelitian (Saprudin, 2017) tentang peningkatan akurasi proses klasifikasi menggunakan Algoritma C45.

Dengan berdasarkan literatur yang sudah disampaikan maka peneliti melakukan penelitian untuk proses analisis data pada IMDb.

## 2. METODE PENELITIAN

Metode yang digunakan oleh peneliti untuk mengolah data atau ulasan dalam analisa sentiment ini eksperimen menggunakan penggabungan algoritma yakni *Naive Bayes* dengan *Particle Swarm Optimization* (PSO), penggabungan algoritma ini dapat meningkatkan nilai akurasi dan *ROC Curve pada algoritma Naive Bayes*. Adapun Pengolahan data tersebut didapatkan dari ulasan atau komentar google play store pada penggunaan aplikasi IMDb. Data tersebut dibagi kedalam 2 (dua) kelompok yaitu ulasan yang bersifat positif dengan makna mendukung dan ulasan yang bersifat negatif dengan makna yang menolak, yang tersimpan dalam 1 (satu) dokumen excel dengan ekstension *xlsx*.

Data yang akan diolah untuk penelitian yakni ulasan yang terdiri dari 1000 data ulasan dari pengguna aplikasi IMDb, terdiri dari 500 ulasan yang bersifat positif atau mendukung dan 500 ulasan yang bersifat negative atau menolak atau tidak mendukung. Hasil dari pengelompokan data ulasan tersebut diproses menggunakan Aplikasi RapidMiner dengan pengolahan pada text processing yang peneliti gunakan *Tokenize, Transform Cases, Stopwords (dictionary)*.



**Gambar 1.** Kerangka Pemikiran

Pada gambar 1 menjelaskan bahwa kerangka kerja dalam penelitian ini dimulai dengan mengambil data dari ulasan pengguna aplikasi IMDb dengan menggunakan bahasa Indonesia yakni pada *google play store* dengan alamat tautan sebagai berikut ini [https://play.google.com/store/apps/details/IMDb\\_Movies\\_TV?id=com.imdb.mobile&hl=in&gl=US](https://play.google.com/store/apps/details/IMDb_Movies_TV?id=com.imdb.mobile&hl=in&gl=US) kemudian hasil dari pengumpulan data tersebut dilakukan pemrosesan teks menggunakan *tokenize*, *Transform Case*, *Stopwords (direktori)*. Pada *Stopwords (dictionar)* ini peneliti menambahkan kosa kata atau kata yang mengandung unsur kebencian. Sedangkan pada *feature selection* terlihat jelas terjadinya penggabungan 2 (dua) algoritma dimana *Naive Bayes* berada dalam *Particle Swarm Optimization*, sehingga hasil eksperimen tersebut didapatkan hasil keakuratan dari memproses data tersebut.

### 3. HASIL DAN PEMBAHASAN

Penelitian dalam melakukan eksperimen menggunakan metode *Naive Bayes* dan *Particle Swarm Optimization* untuk menganalisa sentimen pengguna aplikasi IMDb, dimana pengguna memberikan pengalamannya dalam menggunakan aplikasi IMDb pada ulasan *google play store*. Untuk mengukur akurasi data eksperimen menggunakan Aplikasi RapidMiner Studio 9.10.011 dengan menggunakan spesifikasi CPU Intel Core i3 1.90GHz, RAM 10 GB, dan system operasi Microsoft Windows 10 Professional 64-bit.

Berikut salah satu ulasan yang mengandung ujaran kebencian yakni pada data ulasan yang bersifat negatif, data akan diproses menggunakan Text Processing yang terlihat pada

tabel 1 yang dilalui terdiri dari Tokenize, Transform Cases, dan Stopwords(dicstionary). Hasil pengolahan sebagai berikut:

**Tabel 1.** Text Processing menggunakan Tokenize, Transform Cases, Stopwords (Dictionar)

Ulasan	Tokenize	Transform Cases	Stopwords (Dictionar)
aplikasi	aplikasi	aplikasi	aplikasi
apaan	apaan	apaan nih	apaan nih
nih! gak	nih gak	gak bisa	gak bisa
buat	bisa	buat	buat
buat	buat	nonton	nonton
nonton	nonton	film	film
film,	film	payah	
PAYAH!!	PAYAH	bego	
BEGO!	BEGO		

Pada tabel 1 terlihat bagaimana proses ulasan ujaran kebencian yang bersifat negatif setelah melewati beberapa proses menjadi ulasan yang menjadi lebih baik bersifat positif. Dimana ulasan setelah melalui *text processing* seluruh simbol karakter atau tanda baca akan dihilangkan seperti tanda seru (!) koma (,) dan symbol karakter lainnya. Kemudian pada *transform cases* seluruh text yang menggunakan huruf kapital atau huruf besar akan berubah menjadi huruf kecil, karena ungkapan text dalam huruf kapital adalah ungkapan dimana seseorang mengungkapkan ungkapan yang tinggi, keras atau marah. Pada tahap terakhir yakni *Stopwords (dicstionary)* berfungsi untuk menghilangkan kosa kata atau penggunaan kata yang memiliki arti ujaran kebencian dengan kata-kata kasar, Seperti yang kita lihat pada tabel 1 “BEGO” dalam Bahasa Indonesia artinya “sangat

bodoh”, sehingga setelah melewati tahap *stopwords (disctionary)* ini akan dihilangkan.

Pada pengujian pertama yakni pada ulasan aplikasi IMDb menggunakan algoritma *Naive Bayes* mendapatkan hasil *accuracy*: 74.28%.

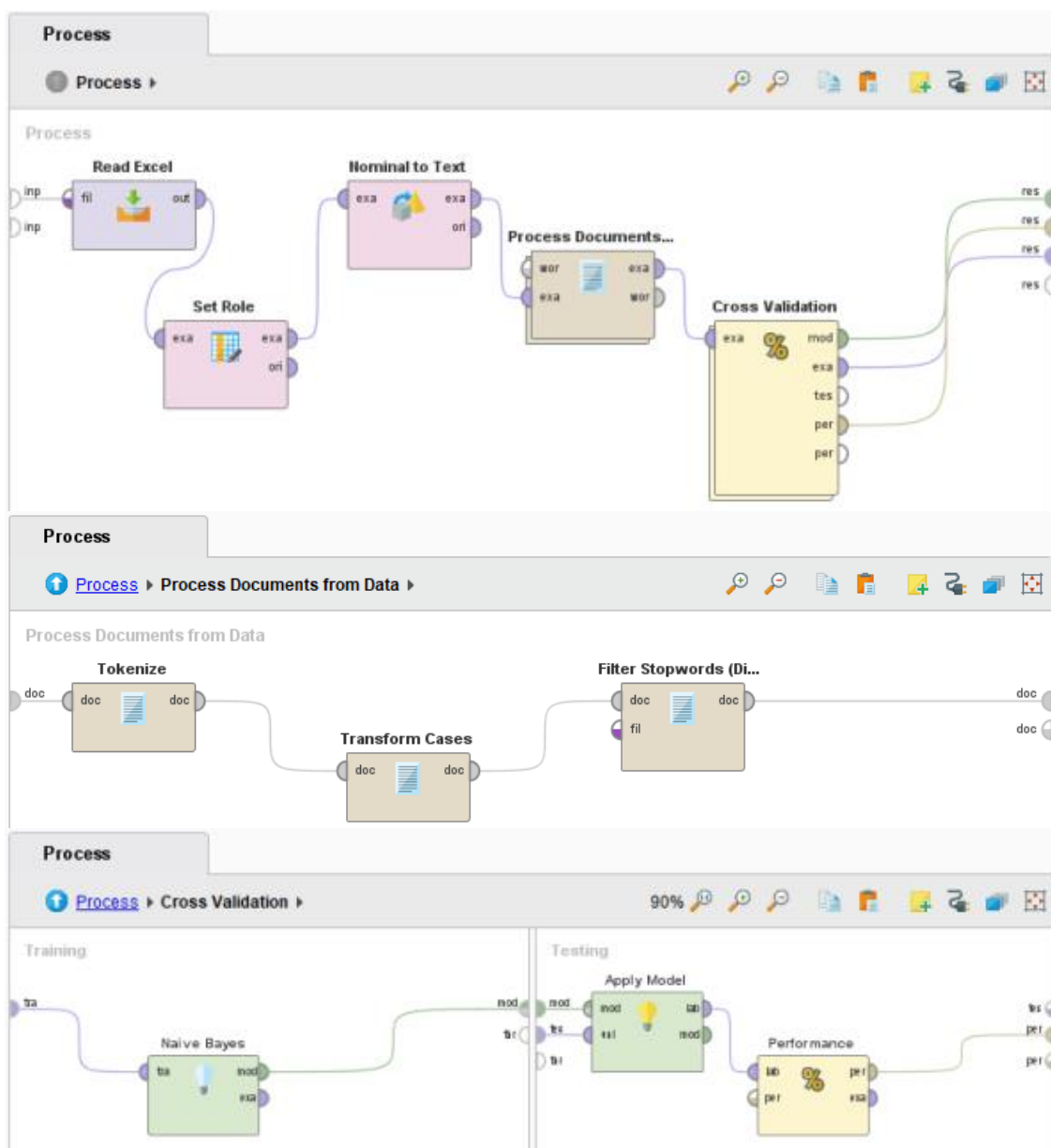
**Tabel 2.** Accuracy Algoritma *Naive Bayes*

	true Negatif	true Positif
pred. Negatif	311	69
pred. Positif	188	431

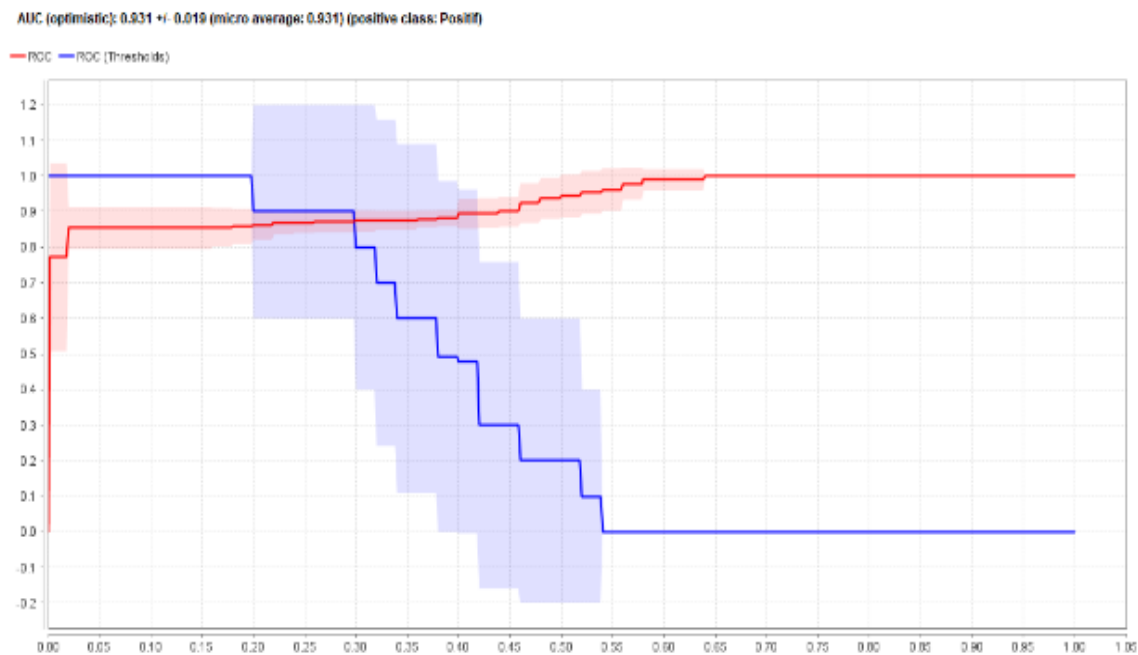
Pada tabel 2 diatas dapat lihat hasil prediksi yang didapat yaitu 431 diklasifikasikan ke

dalam ulasan positif sesuai dengan prediksi yang dilakukan dengan metode *Naive Bayes*, 69 data diprediksi ulasan positif tetapi ternyata hasil prediksinya ulasan negatif, 311 diklasifikasikan review negatif sesuai dengan prediksi yang dilakukan dengan metode *Naive Bayes* dan 188 data diprediksi ulasan negatif ternyata hasil prediksinya ulasan positif.

Dalam menentukan hasil penelitian menggunakan algoritma *Naive Bayes* dengan rancangan pada RapidMiner Studio seperti terlihat pada gambar 2. Sedangkan hasil dari ROC Curve pada penelitian *Naive Bayes* dapat kita lihat pada gambar 3.



**Gambar 2.** Modelling Validation Testing *Naive Bayes*



Gambar 3. ROC Curve Algoritma Naive Bayes

Pada gambar 3 terlihat dengan jelas bahwa hasil yang didapat ROC Curve Algoritma Naive Bayes dengan nilai AUC (optimistic): 0.931 +/- 0.019 (micro average: 0.931) (positive class: Positif). Kemudian pengujian kedua dilakukan yakni menggabungkan kedua metode Naive Bayes dengan Particle Swarm Optimization (PSO) seperti yang terlihat pada tabel 3.

Tabel 3. Accuracy Algoritma Naive Bayes dan Particle Swarm Optimization (PSO)

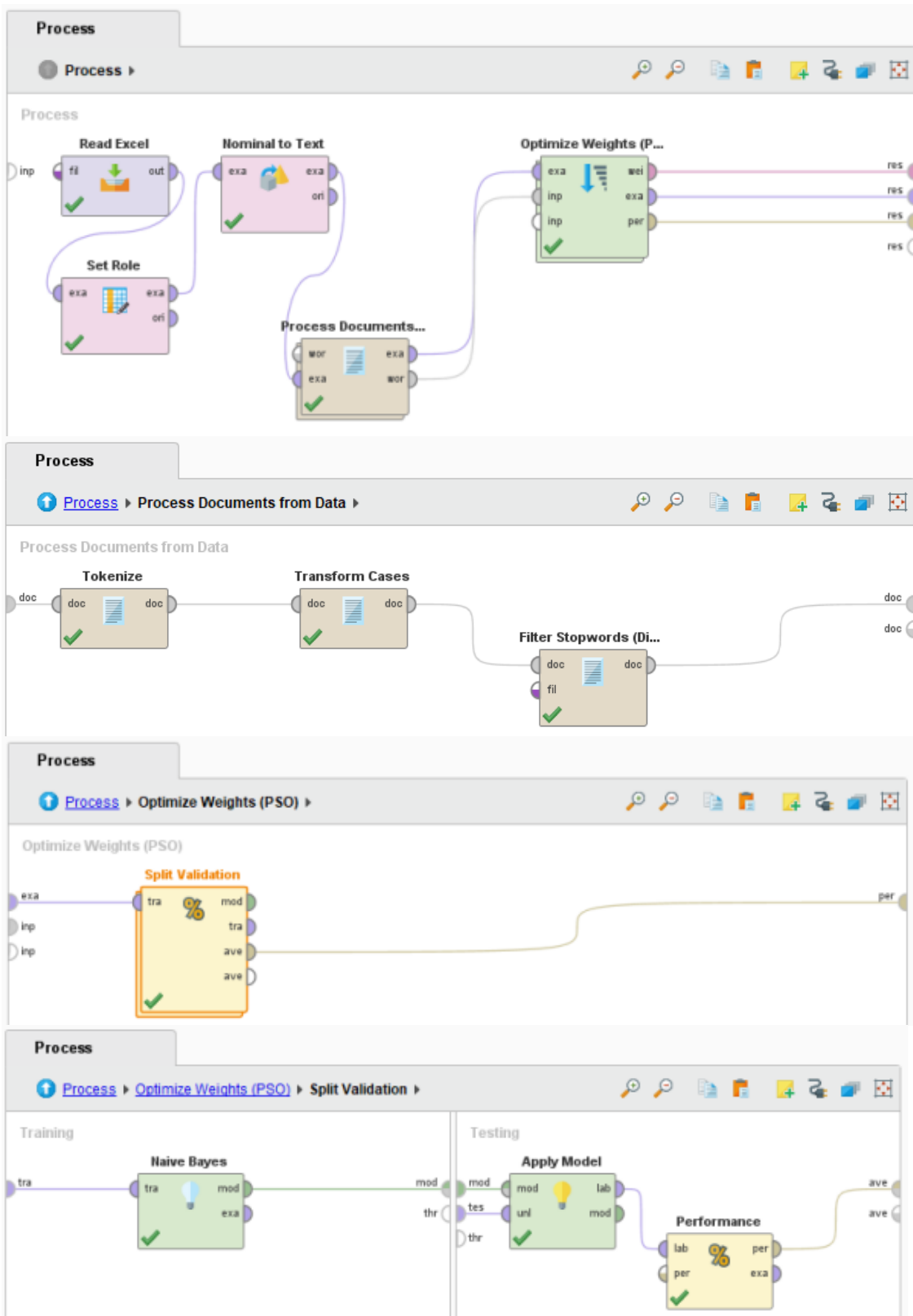
	true Negatif	true Positif
pred. Negatif	84	8
pred. Positif	52	156

Pada tabel 3 dapat dilihat hasil prediksi yang didapat yaitu 156 diklasifikasikan ke dalam ulasan positif sesuai dengan prediksi yang dilakukan dengan metode Naive Bayes dan Particle Swarm Optimization (PSO), 8 data diprediksi ulasan positif tetapi ternyata hasil

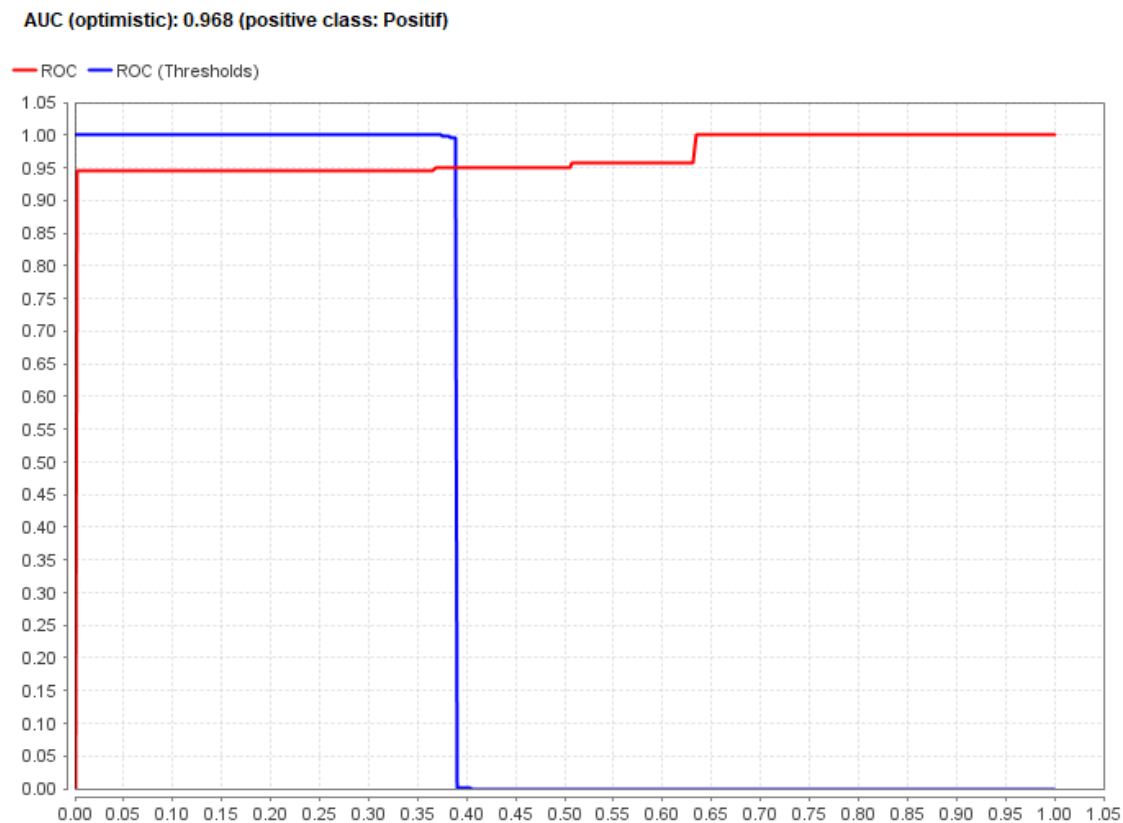
prediksinya ulasan negatif, 84 diklasifikasikan review negatif sesuai dengan prediksi yang dilakukan dengan metode Naive Bayes dan Particle Swarm Optimization (PSO), dan 52 data diprediksi ulasan negatif ternyata hasil prediksinya ulasan positif. Dengan accuracy yang didapat pada pengujian kedua menggunakan Naive Bayes dan Particle Swarm Optimization (PSO) nilai accuracy: 80.00%.

Dalam menentukan hasil penelitian menggunakan algoritma Naive Bayes dan Particle Swarm Optimization (PSO) dengan rancangan pada RapidMiner Studio tampak pada gambar 4. Sedangkan hasil dari ROC Curve pada penelitian Naive Bayes dan Particle Swarm Optimization (PSO) dapat dilihat pada gambar 5.

Pada gambar 5 bahwa hasil dari penelitian yang kedua ini didapat ROC Curve Algoritma Algoritma Naive Bayes dan Particle Swarm Optimization (PSO) dengan nilai AUC (optimistic): AUC (optimistic): 0.968 (positive class: Positif).



Gambar 4. Modelling Validation Testing Naive Bayes dan Particle Swarm Optimization (PSO)



**Gambar 5.** ROC Curve Algoritma Naive Bayes dan Particle Swarm Optimization (PSO)

#### 4. KESIMPULAN

Pengujian model dengan menggunakan Algoritma Naive Bayes dan Particle Swarm Optimization (PSO) dengan menggunakan data ulasan analisis sentimen dari 1000 data ulasan dalam bahasa Indonesia yang terdiri dari 500 ulasan positif dengan sifat yang mendukung dan 500 review negative dengan sifat yang tidak mendukung, Model yang dihasilkan mendapatkan hasil *accuracy* pengujian pertama dilakukan dengan menggunakan Algoritma Naive Bayes dengan nilai *accuracy*: 74.28% yang kemudian diuji kembali dengan pengujian yang kedua dengan menggabungkan 2 (dua) Algoritma Naive Bayes dan Particle Swarm Optimization (PSO) dengan nilai yang terbaik dengan *accuracy* 80.00%. Setelah kita melihat hasil nilai *accuracy* pengujian pertama dan kedua ini kita dapat simpulkan bahwa pengujian kedua jauh lebih baik, dimana Algoritma Naive Bayes akan lebih baik lagi hasilnya jika menggabungkan Algoritma Particle Swarm Optimization (PSO) sehingga dapat meningkatkan nilai *accuracy*. Dengan adanya ulasan atau komentar yang disampaikan oleh pengguna aplikasi atau produk melalui kolom komentar,

kita dapat menentukan pilihan dengan tepat dari pengalaman yang telah menggunakan aplikasi atau produk tersebut. Tentu saja penilaian yang diberikan atau ulasan ada yang bersifat mendukung atau menolak hal ini merupakan hal yang wajar yang sering terjadi, permasalahannya adalah pada ulasan yang disampaikan mengandung unsur kebencian. Tentu saja hal ini yang tidak boleh terjadi dengan memberikan ulasan yang baik walaupun bersifat tidak mendukung atau menolak akan tetap bersifat positif bagi pengembang aplikasi atau produk tertentu sehingga dapat meningkatkan kualitas oleh pengembang aplikasi atau produk tertentu dari ulasan atau komentar yang diberikan oleh pengguna aplikasi atau produk.

#### 5. REFERENSI

- H. B. Jatmiko, N. T. Kurniadi, D. M. (2022). *Optimasi Naive Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta*. 2(1), 22–30.
- Istighfarin, N. F., Rahmastati, R. A., & Nugroho, H. (2020). Penerapan Metode Particle

- Swarm Optimization (PSO) Dan Genetic Algorithm (GA) Pada Sistem Optimasi Visible Light Communication (VLC) Untuk Menentukan Posisi Robot. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 11(1), 279–286. <https://doi.org/10.24176/simet.v11i1.4052>
- Keni Hasanah Wiguna. (2021). *Film Bukan Hanya Sekedar Hiburan*. <https://www.djkn.kemenkeu.go.id/>.
- Legiawati, N., Hermanto, T. I., & Ramadhan, Y. R. (2022). Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naive Bayes Berbasis PSO. 9(4), 930–937. <https://doi.org/10.30865/jurikom.v9i4.4629>
- Pramukti, S. D., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Masyarakat Dengan Metode Naive Bayes dan Particle Swarm Optimization. *Techno.Com*, 21(1), 61–74. <https://doi.org/10.33633/tc.v21i1.5332>
- Prasetyo, V. R., Mercifia, M., Averina, A., Sunyoto, L., & Budiarto. (2022). Film Rating Prediction on Imdb Website Using Neural Network. *Jurnal Ilmiah NERO*, 7(1), 60293.
- Ramadhani, S. H., & Wahyudin, M. I. (2022). Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naive Bayes dan K-NN. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(4), 526–534. <https://doi.org/10.35870/jtik.v6i4.530>
- Saprudin, S. (2017). Penerapan Particle Swarm Optimization (PSO) untuk Klasifikasi dan Analisis Kredit dengan Menggunakan Algoritma C4.5. *Jurnal Informatika Universitas Pamulang*, 2(4), 214. <https://doi.org/10.32493/informatika.v2i4.1488>
- Setiawan, A., Santoso, L. W., & Adipranata, R. (2019). Penerapan Algoritma Particle Swarm Optimization ( PSO ) untuk Optimisasi Pembangunan Negara dalam Turn Based Strategy Game. *Jurnal Infra*, 7(1), 249–255.
- Sudriyanto, S., Rizaldi, R., & Hariri, M. A. R. (2021). Implementasi Particle Swarm Optimization (PSO) untuk Optimisasi Algoritma Naive Bayes dalam Memprediksi Mahasiswa Lulus Tepat Waktu. *COREAI: Jurnal Kecerdasan Buatan, Komputasi Dan Teknologi Informasi*, 2(1), 62–68. <https://www.ejournal.unuja.ac.id/index.php/core/article/view/2181>