



## Perbandingan Optimasi *Feature Selection* pada *Naïve Bayes* untuk Klasifikasi Kepuasan *Airline Passenger*

Yoga Religia<sup>1</sup>, Amali<sup>2</sup><sup>1,2</sup>Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa<sup>1</sup>yoga.religia@pelitabangsa.ac.id, <sup>2</sup>amali@pelitabangsa.ac.id**Abstract**

The quality of an airline's services cannot be measured from the company's point of view, but must be seen from the point of view of customer satisfaction. Data mining techniques make it possible to predict airline customer satisfaction with a classification model. The Naïve Bayes algorithm has demonstrated outstanding classification accuracy, but currently independent assumptions are rarely discussed. Some literature suggests the use of attribute weighting to reduce independent assumptions, which can be done using particle swarm optimization (PSO) and genetic algorithm (GA) through feature selection. This study conducted a comparison of PSO and GA optimization on Naïve Bayes for the classification of Airline Passenger Satisfaction data taken from [www.kaggle.com](http://www.kaggle.com). After testing, the best performance is obtained from the model formed, namely the classification of Airline Passenger Satisfaction data using the Naïve Bayes algorithm with PSO optimization, where the accuracy value is 86.13%, the precision value is 87.90%, the recall value is 87.29%, and the value is AUC of 0.923.

**Keywords:** Data Mining, Classification, Naïve Bayes, Particle Swarm Optimization, Genetic Algorithm

**Abstrak**

Kualitas dari pelayanan suatu maskapai tidak dapat diukur dari sudut pandang perusahaan, melainkan harus dilihat dari sudut pandang kepuasan pelanggan. Teknik *data mining* memungkinkan untuk melakukan prediksi kepuasan pelanggan maskapai penerbangan dengan model klasifikasi. Algoritma *Naïve Bayes* telah menunjukkan akurasi klasifikasi yang luar biasa, namun saat ini asumsi bebas jarang dibahas. Beberapa literatur menyarankan penggunaan pembobotan atribut untuk mengurangi asumsi bebas, dimana dapat dilakukan menggunakan *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) melalui *Feature Selection*. Penelitian ini melakukan perbandingan optimasi PSO dan GA pada *Naïve Bayes* untuk klasifikasi data *Airline Passenger Satisfaction* yang diambil dari [www.kaggle.com](http://www.kaggle.com). Setelah dilakukan pengujian, diperoleh performa yang paling bagus dari model yang dibentuk, yaitu klasifikasi data *Airline Passenger Satisfaction* menggunakan algoritma *Naïve Bayes* dengan optimasi PSO, dimana diperoleh nilai akurasi sebesar 86.13%, nilai presisi sebesar 87.90%, nilai *recall* sebesar 87.29%, dan nilai AUC sebesar 0.923.

**Kata kunci:** Data Mining, Klasifikasi, Naïve Bayes, Particle Swarm Optimization, Genetic Algorithm

**1. Pendahuluan**

Indonesia sebagai negara kepulauan membutuhkan sarana transportasi yang dapat mempermudah masyarakat berpindah dari satu pulau ke pulau yang lain, salah satunya melalui jalur udara. Hal ini dilihat sebagai suatu potensi yang besar yang dapat diambil oleh perusahaan maskapai penerbangan [1]. Perusahaan maskapai perlu menyediakan pelayanan yang berkualitas kepada masyarakat agar masyarakat tidak berpindah menggunakan maskapai yang lain [2]. Kualitas dari pelayanan suatu maskapai tidak dapat diukur dari sudut pandang perusahaan, melainkan harus

dilihat dari sudut pandang kepuasan pelanggan [3]. Salah satu cara yang dapat digunakan untuk mengetahui kepuasan pelanggan adalah menggunakan teknik *data mining* [4].

Data mining dapat digunakan untuk melakukan evaluasi model konseptual yang dapat dijadikan sebagai jembatan antara teori dan praktik berdasarkan kumpulan data yang besar [5]. Salah satu cara yang dapat digunakan untuk memprediksi kepuasan pelanggan dengan teknik *data mining* adalah menggunakan model klasifikasi. Model klasifikasi dapat diterapkan pada data *supervised learning* [6]. Pada website [www.kaggle.com](http://www.kaggle.com) saat ini

menyediakan data *Airline Passenger Satisfaction* yang terdiri dari 25976 *instance*, 22 atribut dan 1 label yang termasuk dalam kategori data *supervised learning* [7], sehingga dapat digunakan untuk membuat model klasifikasi.

Pembuatan model klasifikasi yang optimal membutuhkan suatu algoritma yang baik, salah satunya menggunakan *Naïve Bayes*. Berdasarkan beberapa penelitian terdahulu, algoritma *Naïve Bayes* mampu memberikan performa klasifikasi yang lebih baik dibandingkan algoritma klasifikasi yang lain [8] [9] [10]. Selain dapat memberikan performa klasifikasi yang baik, algoritma *Naïve Bayes* juga dapat digunakan pada data *imbalance* [11] [12], sehingga cocok digunakan untuk mengklasifikasikan data *Airline Passenger Satisfaction*.

Meskipun *Naïve Bayes* telah menunjukkan akurasi klasifikasi yang luar biasa, namun saat ini asumsi bebas jarang dibahas pada klasifikasi *Naïve Bayes*. Asumsi bebas adalah anggapan bahwa setiap sampel data yang digunakan terhadap nilai rata-rata sampel pada setiap kelompok bersifat saling bebas. Salah satu cara untuk mencoba asumsi bebas pada algoritma *Naïve Bayes* adalah dengan pembobotan atribut [13]. Hal tersebut didukung pula oleh Liangxiao Jiang (2019) yang menyebutkan bahwa perlu diusulkan metode pembobotan atribut untuk mengurangi asumsi bebas [14]. Pembobotan atribut dapat dilakukan menggunakan *particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) melalui *Feature Selection* [15].

PSO memiliki keuntungan yang signifikan dalam menangani *fitting* non-linier dan parameter multi-input [16]. PSO tidak memiliki operator evolusi seperti persilangan dan mutasi, sehingga mudah diimplementasikan dan hanya ada sedikit parameter yang harus disesuaikan [17]. Berdasarkan beberapa penelitian terdahulu menyebutkan bahwa kombinasi dari PSO dan *Naïve Bayes* mampu memberikan hasil performa klasifikasi data *imbalance* yang lebih baik dari pada menggunakan *Naïve Bayes* saja [18] [19], bahkan PSO mampu meningkatkan akurasi *Naïve Bayes* hingga lebih dari 10% [20].

GA merupakan metode optimasi yang dikembangkan berdasarkan mekanisme seleksi alam dengan cara meniru genetika makhluk hidup dalam memecahkan masalah [21]. Optimasi yang dilakukan oleh GA adalah dengan memprediksi jumlah iterasi yang tepat, sehingga tidak diperlukan lagi perhitungan dengan jumlah iterasi yang berbeda untuk mendapatkan kemunculan yang lengkap dari jalur bebas [22]. Keuntungan paling signifikan dari GA adalah kemampuannya dalam pencarian global serta kemampuan beradaptasi terhadap spektrum masalah yang luas [23]. Berdasarkan beberapa penelitian sebelumnya menyebutkan bahwa penggunaan

GA mampu meningkatkan performa klasifikasi dari *Naïve Bayes* [24] [25].

Berdasarkan penelitian sebelumnya menunjukkan bahwa baik PSO ataupun GA sama-sama dikatakan mampu untuk meningkatkan performa klasifikasi pada *Naïve Bayes*, hal tersebut menjadi gap penelitian karena dibutuhkan penelitian lebih lanjut untuk mengetahui optimasi manakah yang paling cocok dikombinasikan dengan *Naïve Bayes* untuk klasifikasi data *Airline Passenger Satisfaction*. Penelitian ini melakukan perbandingan optimasi PSO dan GA pada *Naïve Bayes* untuk untuk klasifikasi data *Airline Passenger Satisfaction*.

## 2. Metode Penelitian

### 2.1. Data yang digunakan

Penelitian ini menggunakan data yang diambil dari situs [www.kaggle.com](http://www.kaggle.com) yang diakses pada tanggal 24 maret 2021 berupa data *Airline Passenger Satisfaction* [7]. Data *Airline Passenger Satisfaction* mulai di *upload* pada situs [www.kaggle.com](http://www.kaggle.com) sejak bulan Mei 2020, sehingga masih tergolong *dataset* baru yang belum banyak digunakan untuk penelitian. Data ini terdiri dari 22 atribut, 25976 *instance* dan 1 label dengan *type* data *boolean*. Data *Airline Passenger Satisfaction* merupakan data yang berisi survei kepuasan penumpang maskapai penerbangan di dunia. Penggunaan data ini bertujuan untuk mengetahui faktor apa yang paling berkorelasi dengan kepuasan penumpang maskapai penerbangan sehingga cocok digunakan untuk membuat model klasifikasi. Label yang digunakan ada pada atribut “*Satisfaction*” dengan sebaran 44% pelanggan merasa puas dan sisanya pelanggan merasa tidak puas. Setiap atribut dan label yang digunakan pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1 Konten dari data *Airline Passenger Satisfaction*

Konten	Keterangan	Ket
Gender	Jenis kelamin penumpang (Wanita, Pria)	Atribut
Customer Type	Jenis pelanggan (Pelanggan setia, pelanggan tidak setia)	Atribut
Age	Usia penumpang sebenarnya	Atribut
Type of Travel	Tujuan penerbangan penumpang (Perjalanan Pribadi, Perjalanan Bisnis)	Atribut
Class	Kelas perjalanan di pesawat penumpang (Bisnis, Eco, Eco Plus)	Atribut
Flight distance	Jarak penerbangan dari perjalanan ini	Atribut
Inflight wifi service	Tingkat kepuasan layanan wifi dalam pesawat (1-5)	Atribut
Arrival time convenient	Tingkat kepuasan waktu Keberangkatan / Kedatangan nyaman (1-5)	Atribut
Ease of Online booking	Tingkat kepuasan pemesanan online (1-5)	Atribut
Gate location	Tingkat kepuasan lokasi Gerbang (1-5)	Atribut

Tabel 1 Konten dari data *Airline Passenger Satisfaction*

Konten	Keterangan	Ket
<i>Food and drink</i>	Tingkat kepuasan makanan dan minuman (1-5)	Atribut
<i>Online boarding</i>	Tingkat kepuasan boarding online (1-5)	Atribut
<i>Seat comfort</i>	Tingkat kepuasan kenyamanan kursi (1-5)	Atribut
<i>Inflight entertainment</i>	Tingkat kepuasan hiburan dalam pesawat (1-5)	Atribut
<i>On-board service</i>	Tingkat kepuasan layanan <i>On-board</i> (1-5)	Atribut
<i>Leg room service</i>	Tingkat kepuasan layanan kamar kaki (1-5)	Atribut
<i>Baggage handling</i>	Tingkat kepuasan penanganan bagasi (1-5)	Atribut
<i>Check-in service</i>	Tingkat kepuasan layanan <i>Check-in</i> (1-5)	Atribut
<i>Inflight service</i>	Tingkat kepuasan layanan dalam pesawat (1-5)	Atribut
<i>Cleanliness</i>	Tingkat kepuasan Kebersihan (1-5)	Atribut
<i>Departure Delay</i>	Menit ditunda saat keberangkatan	Atribut
<i>Arrival Delay</i>	Menit tertunda saat Kedatangan	Atribut
<i>Satisfaction</i>	Tingkat kepuasan maskapai penerbangan (Puas, Tidak Puas)	Label

Berdasarkan 25976 *instance* yang terdapat pada data *Airline Passenger Satisfaction*, seluruhnya tidak terdapat *missing value*, sehingga tidak dilakukan *pre-processing* data pada data ini.

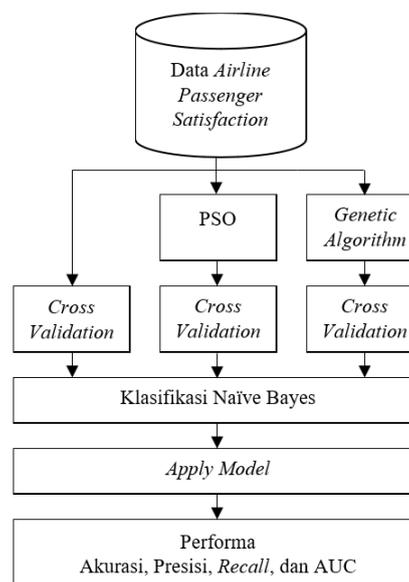
## 2.2. Model Penelitian

Penelitian ini dilakukan menggunakan teknik *data mining* dengan model klasifikasi. Model yang dibentuk digunakan untuk membandingkan penggunaan algoritma *Naïve Bayes*, *Naïve Bayes* dengan optimasi *particle Swarm Optimization* (PSO), dan *Naïve Bayes* dengan optimasi *Genetic Algorithm* (GA) untuk klasifikasi data *Airline Passenger Satisfaction*. Validasi yang digunakan pada model klasifikasi yang dibangun adalah menggunakan *cross validation*. *Cross Validation* dipilih karena *cross validasi* mampu memberikan hasil klasifikasi yang lebih baik dibandingkan dengan metode validasi yang lain [26].

Pada *cross validation*, pelatihan dan pengujian akan dilakukan sebanyak  $K$ . Penelitian ini menggunakan  $k=10$  untuk proses validasi. Penelitian ini tidak secara khusus membagi seluruh data kedalam data *training* dan data *testing*. Penelitian ini menggunakan teknik pengambilan sampel *stratified sampling* untuk digunakan pada *cross validation*. *Stratified sampling* adalah pengambilan sampel probabilitas dengan membagi seluruh populasi, dimana investigasi akan dilakukan pada subkelompok yang berbeda. Proses ini akan diulang sebanyak 10 kali sesuai nilai  $k$  yang digunakan pada penelitian ini. Hasil dari proses ini akan digunakan pada pembuatan *apply model*.

Ketika *apply model* yang dikehendaki sudah dibentuk, selanjutnya akan diukur performa klasifikasi

berdasarkan nilai akurasi, presisi, *recall*, dan *area under curve* (AUC). Secara lebih mudah, model penelitian yang dibangun pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Model Penelitian

Berdasarkan Gambar 1 dapat diketahui bahwa terdapat 3 model utama dalam penelitian ini yaitu (1) Klasifikasi data *Airline Passenger Satisfaction* menggunakan algoritma *Naïve Bayes*; (2) Klasifikasi data *Airline Passenger Satisfaction* menggunakan algoritma *Naïve Bayes* dengan optimasi PSO; dan (3) Klasifikasi data *Airline Passenger Satisfaction* menggunakan algoritma *Naïve Bayes* GA.

Setiap model dari ke-3 model tersebut akan menghasilkan performa berupa nilai akurasi, presisi, *recall*, dan *area under curve* (AUC) yang kemudian akan dibandingkan untuk dianalisa model mana yang memiliki performa yang paling baik. Nilai akurasi, presisi, dan *recall* akan dihitung menggunakan *binary confusion matrix*.

Pada *binary confusion matrix*, pengamatan yang diklasifikasikan dengan benar ke dalam kelas positif disebut *true positives* (TP) dan pengamatan yang diklasifikasikan dengan benar ke dalam kelas negatif disebut *true negatives* (TN). *Instances* dari kelas positif yang diklasifikasikan salah sebagai negatif disebut false negatives (FN) dan *instances* dari kelas negatif yang diklasifikasikan salah sebagai positif disebut false positives (FP). Berdasarkan nilai TP, FP, TN dan TP dapat dihitung indikator performa klasifikasi yang mencerminkan bagaimana kinerja pengklasifikasi dalam mendeteksi kelas yang diberikan. Indikator yang paling umum digunakan adalah akurasi, presisi, *recall* (sensitifitas) yang dapat dituliskan dengan persamaan berikut [27]:

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

$$\text{Presisi} = \frac{TP}{(TP+FP)}$$

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

Akurasi adalah metrik yang paling sederhana dan paling banyak digunakan untuk mengukur performa model klasifikasi. Selain menggunakan akurasi, pada penelitian ini juga mempertimbangkan ukuran performa klasifikasi dalam hal presisi dan *recall*. Menurut Brendan Juba dan Hai S. Le (2019), ukuran performa klasifikasi menggunakan akurasi, presisi dan *recall* lebih disarankan karena cocok untuk klasifikasi data *imbalance* [28]. Klasifikasi nilai AUC sendiri terbagi menjadi 5 kategori, yaitu: 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* [29].

### 3. Hasil dan Pembahasan

Data yang digunakan pada penelitian ini adalah data sekunder yang diambil dari situs [www.kaggle.com](http://www.kaggle.com) berupa data *Airline Passenger Satisfaction*, kemudian diklasifikasikan menggunakan algoritma *Naïve Bayes*. Penelitian ini juga menggunakan algoritma optimasi yaitu PSO dan GA untuk mengetahui optimasi mana yang paling tepat digunakan pada algoritma *Naïve Bayes* untuk klasifikasi data *Airline Passenger Satisfaction*. Model penelitian yang digunakan pada penelitian ini diuji menggunakan tools RapidMiner 9.9 yang kemudian diperoleh nilai akurasi, presisi, *recall*, dan AUC.

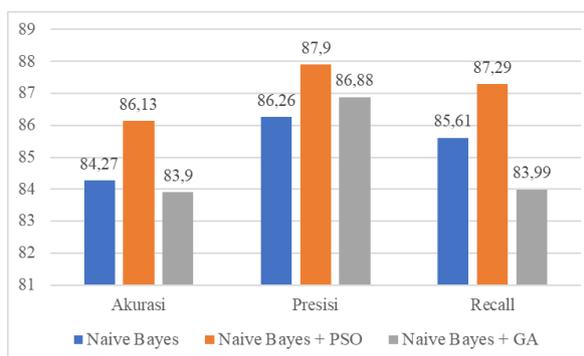
#### 3.1. Hasil Penelitian

Pada penelitian ini, pengujian dilakukan sebanyak 3 kali, yaitu klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naïve Bayes*, klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naïve Bayes* yang di optimasi menggunakan PSO, dan yang terakhir klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naïve Bayes* menggunakan optimasi GA. Pada setiap pengujian dilakukan menggunakan *cross validation* dengan teknik *stratified sampling*, dimana pengambilan sampel probabilitas dilakukan dengan membagi seluruh populasi dan investigasi akan dilakukan pada subkelompok yang berbeda. Dengan menggunakan cara ini, peneliti tidak membagi data kedalam data *training* dan data *testing* dengan ketentuan tertentu. Setelah dilakukan pengujian, diperoleh hasil akurasi, presisi, dan *recall* yang dapat dilihat pada Tabel 2.

Tabel 2 Hasil Akurasi, Presisi, dan Recall dari Pengujian Model

Algoritma	Akurasi	Presisi	Recall
<i>Naïve Bayes</i>	84.27%	86.26%	85.61%
<i>Naïve Bayes</i> + PSO	86.13%	87.90%	87.29%
<i>Naïve Bayes</i> + GA	83.90%	86.88%	83.99%

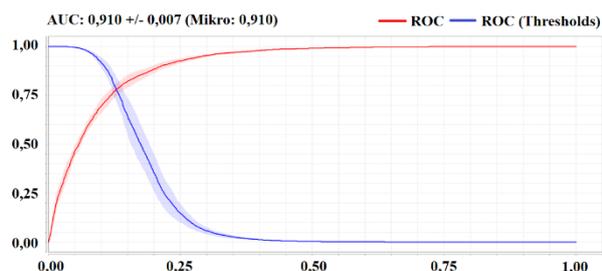
Tabel 2 menunjukkan bahwa model pengujian yang memiliki akurasi, presisi, *recall* paling baik adalah klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naïve Bayes* yang di optimasi menggunakan PSO dimana nilai akurasinya sebesar 86.13%, nilai presisi sebesar 87.90%, dan nilai *recall* sebesar 87.29%. Hasil klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naïve Bayes* memperoleh akurasi sebesar 84.27%, presisi sebesar 86.26%, dan *recall* sebesar 85.61%. Penggunaan optimasi GA pada algoritma *Naïve Bayes* untuk klasifikasi data *Airline Passenger Satisfaction* memperoleh performa paling kecil dibandingkan 2 model lain yang diuji dengan nilai akurasi sebesar 83.90%, nilai presisi sebesar 86.88%, dan nilai *recall* sebesar 83.99%.



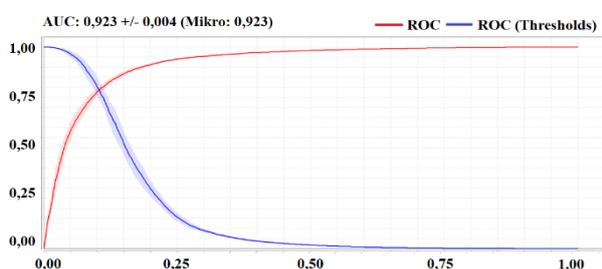
Gambar 2. Perbandingan Akurasi, Presisi, dan Recall dari Pengujian Model Penelitian

Meskipun memperoleh performa paling kecil, akan tetapi nilai presisi pada model klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naïve Bayes* menggunakan optimasi GA masih lebih besar dibandingkan dengan nilai presisi pada model klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naïve Bayes*. Secara lebih jelas perbandingan performa dari ketiga model digambarkan dalam bentuk grafik pada Gambar 2.

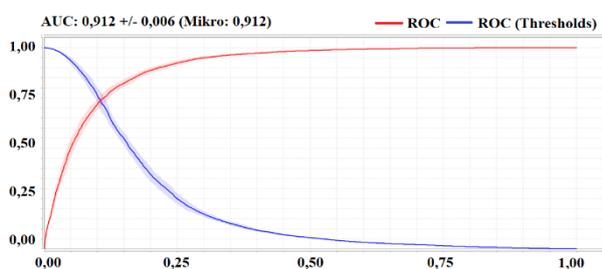
Penelitian ini juga melakukan pengujian untuk memperoleh nilai AUC (*area under the curve*). Nilai AUC diperoleh berdasarkan perhitungan antara hubungan nilai *false positive* dan *true positive*. Nilai AUC ini memiliki nilai maksimal 1 dan nilai minimal 0, apabila nilai yang diperoleh semakin tinggi, maka performa model klasifikasi yang dibangun akan dianggap semakin baik. Pemeriksaan “kecuraman” kurva penting bagi penelitian ini karena menggambarkan proses meminimalkan rasio *false positive* sambil memaksimalkan rasio *true positive*. Hasil nilai AUC dari model penelitian ini representasikan dengan grafik ROC (*Receiver Operating Characteristics*) dapat dilihat pada Gambar 3, Gambar 4 dan Gambar 5.

Gambar 3. Grafik AUC dengan Algoritma *Naive Bayes*

Berdasarkan Gambar 3 dapat diketahui bahwa nilai AUC yang diperoleh dari klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naive Bayes* adalah sebesar 0.910 dengan ketepatan +/- 0.007.

Gambar 4. Grafik AUC dengan Algoritma *Naive Bayes* + PSO

Gambar 4 menunjukkan bahwa nilai AUC yang diperoleh dari klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naive Bayes* menggunakan optimasi PSO adalah sebesar 0.923 dengan ketepatan +/- 0.004.

Gambar 5. Grafik AUC dengan Algoritma *Naive Bayes* + GA

Hasil yang ditunjukkan pada Gambar 5 menunjukkan bahwa klasifikasi data *Airline Passenger Satisfaction* dengan algoritma *Naive Bayes* menggunakan optimasi GA memperoleh nilai AUC sebesar 0.912 dengan ketepatan +/- 0.006. Secara keseluruhan dari nilai AUC yang diperoleh berdasarkan pengujian model penelitian ini dapat dilihat pada Tabel 3.

Tabel 3 Hasil AUC dari Pengujian Model

Algoritma	AUC
<i>Naive Bayes</i>	0.910 +/- 0.007
<i>Naive Bayes</i> + PSO	0.923 +/- 0.004
<i>Naive Bayes</i> + GA	0.912 +/- 0.006

Berdasarkan Tabel 3 dapat diketahui bahwa nilai AUC dari penggunaan algoritma *Naive Bayes* adalah sebesar

0.910, nilai AUC penggunaan algoritma *Naive Bayes* dengan optimasi PSO sebesar 0.923, dan nilai AUC penggunaan algoritma *Naive Bayes* dengan optimasi GA sebesar 0.912. Artinya seluruh nilai AUC dari model penelitian yang digunakan memperoleh nilai diatas 0.900, sehingga dapat dikatakan bahwa hasil kasifikasi pada penelitian ini telah *Excelent Classification*.

### 3.2. Hasil Pembobotan PSO dan GA

Beberapa penelitian terdahulu beranggapan bahwa algoritma PSO dan GA memiliki banyak kesamaan, hal ini dikarenakan baik PSO ataupun GA melakukan optimasi dengan menentukan populasi solusi acak dan pencarian nilai optimal dengan memperbarui generasi [30]. Namun demikian, PSO tidak sama dengan GA, PSO tidak mempunyai operator evolusi seperti crossover dan mutasi. Pendekatan PSO mengibaratkan sekumpulan partikel yang secara bersamaan menjelajahi ruang penelusuran masalah dengan tujuan menemukan konfigurasi global yang optimal [31]. Proses optimasi pada GA dilakukan berdasarkan pada populasi sampel dengan mengembangkan kandidat populasi solusi menuju solusi yang lebih baik [32].

Cara kerja yang berbeda tersebut menghasilkan pembobotan yang berbeda meskipun data yang digunakan sama. Pada penelitian ini hasil pembobotan PSO dan GA pada data *Airline Passenger Satisfaction* dapat dilihat pada Tabel 4.

Tabel 4 Hasil Pembobotan pada GA dan PSO

Atribut	Pembobotan	
	PSO	GA
<i>Gender</i>	0	0
<i>Customer Type</i>	0,77	1
<i>Age</i>	0	1
<i>Type of Travel</i>	1	0
<i>Class</i>	1	1
<i>Flight Distance</i>	0	0
<i>Inflight wifi service</i>	1	1
<i>Departure/Arrival time convenient</i>	1	0
<i>Ease of Online booking</i>	1	0
<i>Gate location</i>	0	0
<i>Food and drink</i>	0	1
<i>Online boarding</i>	0,93	1
<i>Seat comfort</i>	0	0
<i>Inflight entertainment</i>	0	1
<i>On-board service</i>	1	0
<i>Leg room service</i>	1	0
<i>Baggage handling</i>	0	1
<i>Checkin service</i>	0	0
<i>Inflight service</i>	0	0
<i>Cleanliness</i>	1	0
<i>Departure Delay in Minutes</i>	1	0
<i>Arrival Delay in Minutes</i>	1	0

Pada penggunaan PSO terdapat 12 atribut yang dilakukan pembobotan, sedangkan pada penggunaan

GA hanya terdapat 8 atribut yang dilakukan pembobotan. Berdasarkan Tabel 4 juga dapat diketahui bahwa terdapat 6 atribut yang tidak dilakukan pembobotan, baik menggunakan PSO ataupun GA, yaitu: *Gender*, *Flight Distance*, *Gate location*, *Seat comfort*, *Checkin service*, dan *Inflight service*. Selain itu terdapat pula 4 atribut yang diberikan pembobotan oleh PSO dan GA yaitu: *Customer Type*, *Class*, *Inflight wifi service*, dan *Online boarding*. Berdasarkan temuan tersebut dapat diketahui bahwa 4 atribut tersebutlah yang paling perlu diperhatikan untuk mengetahui kepuasan pelanggan dari maskapai penerbangan.

### 3.3. Pembahasan Hasil

Berdasarkan pengujian yang telah dilakukan, dapat diketahui bahwa dengan menggunakan optimasi PSO pada algoritma *Naïve Bayes* dapat meningkatkan performa klasifikasi data *Airline Passenger Satisfaction* dari sisi akurasi, presisi, *recall*, dan AUC. Penggunaan optimasi GA pada algoritma *Naïve Bayes* hanya mampu meningkatkan performa klasifikasi data *Airline Passenger Satisfaction* dalam hal presisi dan nilai AUC saja. Penggunaan PSO mampu memberikan peningkatan akurasi penggunaan *Naïve Bayes* sebesar 1.86%, sedangkan penggunaan GA belum mampu meningkatkan akurasi penggunaan *Naïve Bayes*.

Pada penelitian ini seluruh model yang dibangun memiliki nilai AUC diatas 0.900 yang berarti bahwa model klasifikasi yang dibangun telah masuk dalam kategori *Excellent Classification* sehingga layak untuk digunakan dalam menyelesaikan kasus lain dengan jenis data yang sama. Berdasarkan pembobotan atribut menggunakan PSO dan GA dapat diketahui bahwa atribut *Customer Type*, *Class*, *Inflight wifi service*, dan *Online boarding* adalah atribut yang paling perlu diperhatikan untuk mengetahui kepuasan pelanggan dari maskapai penerbangan.

Hasil akurasi dan *recall* dari penggunaan optimasi GA pada algoritma *Naïve Bayes* untuk klasifikasi data *Airline Passenger Satisfaction* lebih kecil dari penggunaan algoritma *Naïve Bayes* saja diduga karena pemberian bobot atribut yang ada hanya terdapat pada 8 atribut dari 22 atribut. Hal tersebut mengindikasikan bahwa atribut yang diberikan pembobotan oleh GA tidak sampai 50% dari seluruh atribut yang ada, sehingga perhitungan probabilitas yang ada pada *Naïve Bayes* menjadi tidak optimal yang berdampak pada lebih rendahnya nilai akurasi dan *recall* yang dihasilkan.

## 4. Kesimpulan

Penelitian ini dilakukan untuk mengklasifikasikan data *Airline Passenger Satisfaction* menggunakan algoritma *Naïve Bayes*, selain itu pada penelitian ini juga digunakan algoritma optimasi berupa PSO dan GA. Pengujian model penelitian yang dibangun dilakukan

dengan menggunakan *tools* RapidMiner 9.9. Hasil pengujian menunjukkan bahwa penggunaan GA belum mampu meningkatkan performa *Naïve Bayes* untuk mengklasifikasikan data *Airline Passenger Satisfaction*, sedangkan penggunaan PSO mampu meningkatkan performa akurasi algoritma *Naïve Bayes* untuk klasifikasi data *Airline Passenger Satisfaction* sebesar 1.86%. Adapun performa yang paling bagus dari model yang dibentuk adalah klasifikasi data *Airline Passenger Satisfaction* menggunakan algoritma *Naïve Bayes* dengan optimasi PSO, dimana diperoleh nilai akurasi sebesar 86.13%, nilai presisi sebesar 87.90%, nilai *recall* sebesar 87.29%, dan nilai AUC sebesar 0.923. Meskipun penggunaan PSO dapat memberikan peningkatan performa klasifikasi, namun dengan peningkatan akurasi sebesar 1.86% dirasa masih belum cukup maksimal, sehingga dibutuhkan penelitian lebih lanjut untuk menyelesaikan permasalahan tersebut.

Pada penelitian selanjutnya perlu mengeksplorasi optimasi dengan konsep *ensemble method*. *Ensemble method* menggabungkan beberapa model menjadi satu dan biasanya memberikan hasil yang lebih akurat daripada menggunakan algoritma klasifikasi *basic* saja. Diharapkan dengan menggunakan *ensemble method* dapat memberikan performa yang lebih baik dari penggunaan algoritma *Naïve Bayes* dalam mengklasifikasikan data *Airline Passenger Satisfaction*.

## Daftar Rujukan

- [1] E. L. Widjaja, A. Aprilia and A. Harianto, "Analisa Pengaruh Kualitas Layanan Terhadap Kepuasan Penumpang Maskapai Penerbangan Batik Air," *Jurnal Hospitality dan Manajemen Jasa*, vol. 5, no. 2, pp. 118-132, 2017.
- [2] W. Ardhia, "Tingkat Kepuasan Penumpang Terhadap Layanan Maskapai Penerbangan PT. Lion Air Rute Menuju Jakarta," *Jurnal Perhubungan Udara*, vol. 41, no. 1, pp. 19-28, 2015.
- [3] M. D. Darus and K. Mahalli, "Analisis Tingkat Kepuasan Penumpang Terhadap Kualitas Pelayanan di Bandar Udara Internasional Kualanamu," *Jurnal Ekonomi dan Keuangan*, vol. 3, no. 6, pp. 408-420, 2015.
- [4] M. S. Garver, "Using Data Mining for Customer Satisfaction Research," *Marketing Research*, vol. 14, no. 1, pp. 8-17, 2002.
- [5] S. Moro, J. Esmerado, P. Ramos and B. Alturas, "Evaluating A Guest Satisfaction Model Through Data Mining," *International Journal of Contemporary Hospitality Management*, vol. 32, no. 4, pp. 1523-1538, 2019.
- [6] V. Gopalakrishnan and C. Ramaswamy, "Patient Opinion mining to Analyze Drugs Satisfaction Using Supervised Learning," *Journal of Applied Research and Technology*, vol. 15, no. 1, pp. 311-319, 2017.
- [7] Kaggle, "Kaggle.com," Mei 2020. [Online]. Available: <https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction>. [Accessed 24 Maret 2021].
- [8] I. A. A. Amra and A. Y. A. Maghari, "Students Performance Prediction Using KNN and Naïve Bayesian," in *The 8th International Conference on Information Technology (ICIT)*, Al-Zaytoonah University of Jordan, Jordan, 2017.
- [9] F. Osisanwo, J. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi and J. Akinjobi, "Supervised Machine Learning

- Algorithms: Classification and Comparison," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, no. 3, pp. 128-138, 2017.
- [10] E. N. Azizah, U. Pujianto, E. Nugraha and Darusalam, "Comparative Performance Between C4.5 and Naive Bayes Classifiers in Predicting Student Academic Performance in A Virtual Learning Environment," in *The 4th International Conference on Education and Technology (ICET)*, Malang, Indonesia, 2018.
- [11] K. Madasamy and M. Ramaswami, "Data Imbalance and Classifiers: Impact and Solutions from A Big Data Perspective," *International Journal of Computational Intelligence Research*, vol. 13, no. 9, pp. 2267-2281, 2017.
- [12] E. M. Hassib, A. I. El-Desouky, E.-S. M. El-Kenawy and S. M. El-Ghamrawy, "An Imbalanced Big Data Mining Framework for Improving Optimization Algorithms Performance," *Journal & Magazines*, vol. 7, no. 1, pp. 170774-170795, 2019.
- [13] S. Chen, G. I. Webb, L. Liu and X. Ma, "A Novel Selective Naive Bayes Algorithm," *Knowledge-Based Systems*, vol. 192, pp. 1-15, 2020.
- [14] L. Jiang, L. Zhang, L. Yu and D. Wang, "Class-Specific Attribute Weighted Naive Bayes," *Pattern Recognition*, vol. 88, no. 1, pp. 321-330, 2019.
- [15] S. Ernawati, R. Wati, N. Nuris, L. S. Marita and E. R. Yulia, "Comparison of Naive Bayes Algorithm with Genetic Algorithm and Particle Swarm Optimization as Feature Selection for Sentiment Analysis Review of Digital Learning Application," *Journal of Physics: Conference Series*, vol. 1641, pp. 1-7, 2020.
- [16] X. Liu, Z. Liu, Z. Liang, S.-P. Zu, J. A. F. O. Correia and A. M. P. D. Jesus, "PSO-BP Neural Network-Based Strain Prediction of Wind Turbine Blades," *Materials*, vol. 12, no. 12, pp. 2-15, 2019.
- [17] S. Srivastava, J. Gupta and M. Gupta, "PSO & Neural-Network Based Signature Recognition for Harmonic Source Identification," in *IEEE Region 10 International Conference TENCON*, Singapore, 2009.
- [18] M. Misdram, E. Noersasongko, A. Syukur, Purwanto, M. Muljono, H. A. Santoso and D. R. I. M. Setiadi, "Analysis of Imputation Methods of Small and Unbalanced Datasets in Classifications using Naive Bayes and Particle Swarm Optimization," in *International Seminar on Application for Technology of Information and Communication (ISemantic)*, Semarang, Indonesia, 2020.
- [19] I. Romli, T. Pardamean, S. Butsianto, T. N. Wiyatno and E. B. Mohamad, "Naive Bayes Algorithm Implementation Based on Particle Swarm Optimization in Analyzing the Defect Product," *Journal of Physics: Conference Series*, vol. 1845, no. 1, pp. 1-6, 2021.
- [20] J. Li, L. Ding and B. Li, "A Novel Naive Bayes Classification Algorithm Based on Particle Swarm Optimization," *The Open Automation and Control Systems Journal*, vol. 6, no. 1, pp. 747-753, 2014.
- [21] Y. Religia, A. Nugroho and W. Hadikristanto, "Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 5, no. 1, pp. 187-192, 2021.
- [22] A. Arwan and D. S. Rusdianto, "Optimization of Genetic Algorithm Performance Using Naive Bayes for Basis Path Generation," *Kinetik*, vol. 2, no. 4, pp. 273-282, 2017.
- [23] E. Stripling, S. v. Broucke, K. Antonio, B. Baesens and M. Snoeck, "Profit Maximizing Logistic Model for Customer Churn Prediction Using Genetic Algorithms," *Swarm and Evolutionary Computation*, vol. 40, no. 1, pp. 116-130, 2018.
- [24] D. K. Choubey, S. Paul, S. Kumar and S. Kumar, "Classification of Pima Indian Diabetes Dataset Using Naive Bayes With Genetic Algorithm As An Attribute Selection," in *The International Conference on Communication and Computing Systems (ICCCS)*, Ranchi, India, 2016.
- [25] L. G. P. Suardani, I. M. A. Bhaskara and M. Sudarma, "Optimization of Feature Selection Using Genetic Algorithm with Naive Bayes Classification for Home Improvement Recipients," *International Journal of Engineering and Emerging Technology*, vol. 3, no. 1, pp. 66-70, 2018.
- [26] T. Horvat, L. Havaš and D. Srpak, "The Impact of Selecting a Validation Method in Machine Learning on Predicting Basketball Game Outcomes," *Symmetry*, vol. 12, no. 3, pp. 1-15, 2020.
- [27] S. Ruuskaa, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen and J. Mononen, "Evaluation of The Confusion Matrix Method in The Validation of An Automated System for Measuring Feeding Behaviour of Cattle," *Behavioural Processes*, vol. 148, no. 1, pp. 56-62, 2018.
- [28] B. Juba and H. S. Le, "Precision-Recall versus Accuracy and the Role of Large Data Sets," in *The 3th Conference on Artificial Intelligence (AAAI)*, Washington, United States of America, 2019.
- [29] I. Romli, E. Pusnawati and A. Siswandi, "Comparison of NB and NB-PSO to Determine Level of Vehicles Sales," *Journal of Physics: Conference Series*, vol. 1764, no. 1, pp. 1-6, 2021.
- [30] N. A. Maori, "Perbandingan Metode ANN-PSO dan ANN-GA untuk Peningkatan Akurasi Prediksi Harga Emas Antam," *Jurnal Disprotek*, vol. 10, no. 2, pp. 101-106, 2019.
- [31] B. Chopard and M. Tomassini, "Particle Swarm Optimization," in *An Introduction to Metaheuristics for Optimization*, Springer, Cham, Natural Computing Series, 2018, p. 97-102.
- [32] E. Habibi, M. Salehi, G. Yadegarfar and A. Taheri, "Optimization of ANFIS Using A Genetic Algorithm for Physical Work Rate Classification," *International Journal of Occupational Safety and Ergonomics*, vol. 26, no. 3, pp. 436-443, 2020.