

## Optimasi Naive Bayes Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Deteksi Autisme Spectrum Disorder

Candra Agustina  
AMIK BSI Yogyakarta  
candra.caa@bsi.ac.id

*Abstract - The number of Autism and Spectrum Autism sufferers in Indonesia is increasing every year. By 2015 an estimated 1 in 250 children have autism, there are 12,800 autistic children or 134,000 people with autism spectrum. Will be able to detect Autistic Spectrum Disorder sufferers today is still difficult and takes a relatively long time. In the world of mental health, to perform their diagnosis on the Diagnostic and Stastistical Manual of Mental Disorder issued by the American Psychiatric Association. Several previous studies have made machine learning for ASD placement based on DSM-5. This study uses datasets from previous research results uploaded in the University of California Irvine machine learning data repository with the title Data Screening Spectrum Autism Spectrum for Adults. The data obtained is then processed using Naïve Bayes and Naïve Bayes algorithms with Particle Swarm Optimization. The data generated as much as 704 data which is then divided into 2 groups of data that is Data Training Total 563 and Data Testing tool 141. The data is then processed with the help of rapidminer software with a model that has been selected previously. From result of data processing result of data processing with naive bayes yielded data equal to 92,91% and data processing with naive bayes using Particle Swarm Optimization Algorithm give value equal to 98,51%.*

*Keyword : Autistic Spectrum Disorder, Naïve Bayes, Particle Swarm Optimization, Optimisation*

**Abstrak -** Jumlah penderita Autisme dan Spektrum Autis di Indonesia mengalami peningkatan setiap tahun. Pada tahun 2015 diperkirakan 1 dari 250 anak menderita autisme, artinya ada 12.800 anak penyandang autisme atau 134.000 penyandang spektrum Autis. Akan tetapi deteksi penderita *Autistic Spectrum Disorder* sekarang ini masih sulit untuk dilakukan dan memerlukan waktu yang *relative* lama. Dalam dunia kesehatan jiwa, untuk melakukan diagnosa mereka mengacu pada *Diagnostic and Stastistical Manual of Mental Disorder* yang dikeluarkan oleh *American Psychiatric Assosiasi*. Beberapa penelitian sebelumnya sudah membuat *machine learning* untuk mendeteksi ASD berdasarkan DSM-5. Penelitian ini menggunakan dataset dari hasil penelitian sebelumnya yang diunggah di *University of California Irvine machine learning data repository* dengan judul *Autistic Spectrum Disorder Screening Data for Adult*. Data yang didapatkan kemudian diolah menggunakan algoritma *Naïve Bayes* dan *Naïve Bayes* dengan *Particle Swarm Optimization*. Data yang diperoleh sebanyak 704 data yang kemudian dibagi kedalam 2 kelompok data yaitu *Data Training* sejumlah 563 dan *Data Testing* berjumlah 141. Data tersebut kemudian diolah dengan bantuan software *rapidminer* dengan model yang sudah ditentukan sebelumnya. Dari hasil pengolahan data didapatkan bahwa pengolahan data dengan *naïve bayes* menghasilkan akurasi sebesar 92,91% dan pengolahan data dengan *naïve bayes* yang digabungkan dengan Algoritma *Particle Swarm Optimization* memberikan akurasi sebesar 98,51%.

**Kata Kunci :** *Autistic Spectrum Disorder, Naïve Bayes, Particle Swarm Optimization, Optimasi*

### I. Pendahuluan

*Autisme Spectrum Disorder* adalah kondisi seseorang yang mengalami gangguan perkembangan otak yang mempengaruhi kemampuan bersosial, bahasa, komunikasi dan berperilaku. *Austistic Spectrum Disorder* ini dibagi ke dalam lima golongan: (Lubis & Suwandi, 2016).

1. Autistik
2. Asperger
3. Disintegratif Masa Kanak
4. Gangguan Rett
5. Perkembangan Pervasif

Diagnosa seseorang mengidap *Autisme Spectrum Disorder* harus dilakukan oleh ahli,

yang dalam hal ini adalah Psikolog berdasarkan aturan tertentu. Saat ini dalam menentukan diagnosa mengacu *Diagnostic and Stastistical Manual of Mental Disorder (DSM-5)* yang dikeluarkan oleh APA (*American Psychiatric Assosiasi*).

Diagnosa yang dilakukan memerlukan waktu *relative* lama. Karena seorang penderita *Autisme Spectrum Disorder* mempunyai karakteristik berbeda dengan penderita lainnya.

Diagnosa awal sangat diperlukan untuk mengurangi dampak ekonomi yang ditimbulkan. Selain itu diagnosa yang tepat akan menghasilkan penanganan yang tepat sehingga dapat segera upaya penyembuhan. Oleh karena

itu diperlukan tools yang membantu pakar untuk mendiagnosa seseorang menderita *Autisme Spectrum Disorder* atau tidak.

Peneliti sebelumnya membuat dataset yang diolah menggunakan decision tree dan menghasilkan aplikasi android yang bisa didownload dengan nama ASD Test (Thabtah,2017)

Berdasarkan Diagnostic and Statistical Manual of Mental Disorder (DSM-5) yang dikeluarkan oleh APA (American Psychiatric Assosiasi) seseorang dikatakan menderita ASD jika memenuhi kriteria sebagai berikut :

- I. Kurangnya interaksi social dan komunikasi yang bersifat menetap.  
Dijabarkan lebih lanjut sebagai berikut:
  - a) Kesulitan dalam timbal balik sosial
  - b) Mengalami masalah komunikasi nonverbal
  - c) Kesulitan memahami dan mengembangkan sebuah hubungan
- II. Perilaku yang terbatas dan berulang dengan pola yang sama
  - a) Gerakan motorik yang selalu berulang
  - b) Tidak fleksibel terhadap rutinitas, ritual, ataupun pola verbal dan nonverbal
  - c) Sangat terbatas terbatas pada perilaku yang tidak normal.
  - d) Hyper atau hipo reaktifitas terhadap lingkungan

Tingkat keparahan *Autisme Spectrum Disorder* dibagi kedalam 3 level dengan penanganan yang berbeda.

No	Tingkat	Kebutuhan
1	Level 1 (Ringan)	Membutuhkan bantuan minimal
2	Level 2 (Menengah)	Membutuhkan bantuan yang cukup banyak untuk melakukan kegiatan sehari-hari.
3	Level 3 (Berat)	Sangat membutuhkan bantuan.

(DSM-V, 2013)

Deteksi *Autistic Spectrum Disorder* ini cenderung sulit dilakukan dengan cepat dan tepat. Oleh karena itu diperlukan *tools* untuk bisa membantu pendeteksian secara cepat agar penanganan dapat segera dilakukan. Dalam penelitian ini akan dicari algoritma dengan akurasi paling tinggi sehingga dapat dijadikan referensi untuk membuat *tools* tersebut.

## II. Landasan Teori

### Data Mining

Data mining adalah menemukan pola tertentu dari data yang berjumlah besar (Han & Kamber 2011).

Dalam Data Mining dikenal 5 Teknik yaitu (Hermawati 2013) :

#### a) Klasifikasi

Dalam teknik klasifikasi adalah membagi data yang ada ke dalam kelompok - kelompok tertentu atau kelas yang telah ditentukan sebelumnya.

Penerapan dalam aplikasi

- Penjualan langsung  
Digunakan untuk menentukan target pelanggan yang berpotensi.
- Fraud Detection  
Digunakan untuk mendeteksi transaksi yang tidak sah/curang.
- Customer Attrition  
Digunakan untuk memprediksi pelanggan yang berpotensi untuk berpindah ke perusahaan lain sejenis.

#### b) Regresi

Regresi digunakan untuk meramalkan nilai yang berkelanjutan berdasarkan variable lain yang mempengaruhi.

Contoh palikasi dalam teknik ini adalah :

- Memperkirakan jumlah penjualan.
- Memperkirakan kecepatan angin.

#### c) Klasterisasi

Klasterisasi dengan membagi sekelompok besar data kedalam sub bagian data sehingga setiap kelompok data tersebut mempunyai set properti. Dalam satu kelompok setiap data akan mempunyai tingkat similaritas yang tinggi sedangkan tingkat similaritas antar kelompok sangat rendah.

#### d) Kaidah Asosiasi

Kaidah Asosiasi mengenali sekumpulan atribut yang muncul secara bersama-sama.

Teknik ini diterapkan dalam bidang :

- Pemasaran
- Mengelola Peletakan barang dalam supermarket.
- Digunakan untuk mengatur persediaan barang

#### e) Pencarian Pola Sequential

Digunakan untuk meenentukan kejadian yang terjadi secara bersamaan.

### Algoritma Naïve Bayes

Metode ini dikemukakan oleh Thomas Bayes, digunakan untuk memprediksi peluang dimasa

depan dengan masa lalu sebagai dasarnya (Saleh,2015)

Naïve Bayes didasarkan pada penyederhanaan bahwa nilai atribut secara kondisional akan bebas jika diberikan nilai output. Kelebihan ketika menggunakan Naïve Bayes adalah data training yang digunakan relatif kecil atau sedikit. Persamaan Theorema Bayes :

$$P(H|X) = \frac{P(X|H).P(H)}{P(H)}$$

Keterangan :

X : data yang kelasnya belum diketahui  
H : Hipotesis data dengan menggunakan sebuah kelas

P(H|X) : Probabilitas hipotesis H berdasarkan X

P(H) : Probabilitas H

P(X|H) : Probabilitas X Berdasarkan H

P(X) : Probabilitas X

Rumus diatas disesuaikan sebagai berikut

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1...Fn|C)}{P(F1...Fn)}$$

Rumus diatas disederhanakan lagi ke dalam bentuk seperti dibawah ini

$$posterior = \frac{prior \times likelihood}{evidence}$$

*Evidence* selalu bernilai sama/tetap pada kelas dalam sebuah sampel. Kemudian nilai *posterior* tersebut dibandingkan dengan nilai *posterior* kelas lain untuk menentukan kelas klasifikasi dari sebuah sampel.

### Particle Swarm Optimization

Particle Swarm Optimization pertama kali ditemukan karena terinspirasi oleh perilaku sekawanan burung dalam mencari makan. Teori tersebut didasarkan pada tiga prinsip utama yaitu : Evaluasi, Bandingkan dan Tiru. Oleh karena prinsip tersebut Particle Swarm Optimization dapat meningkatkan akurasi algoritma sebelumnya karena mampu membandingkan sebelum meng eksekusi.(Liu dan Miao, 2009).

### Penelitian Terkait

Dalam penelitian Saeful Mujab membuktikan bahwa *Particle Swarm Optimization* meningkatkan akurasi Algoritma *Decision Tree* sebesar 0,43%. Peneliti membandingkan antara pengolahan data dengan Algoritma C.4.5 dan Algoritma C.4.5 dengan Particle Swarm

Optimization. Data yang diolah adalah data nasabah sebuah bank, penelitian dilakukan untuk mendapatkan algoritma terbaik dalam pemilihan nasabah potensial untuk membuka rekening deposito.(Mujab, 2013)

### III. Metode Penelitian

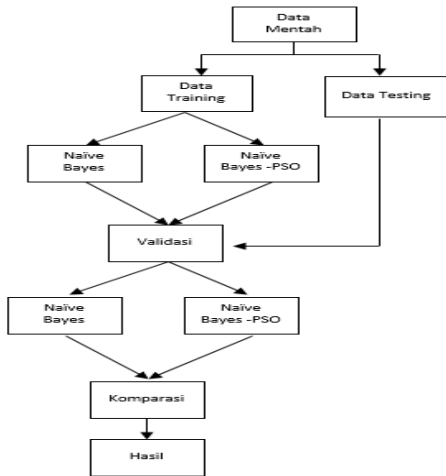
Penelitian yang dilakukan adalah jenis Penelitian Experimental dengan Data set yang peroleh dari UCI *Repository*. Data yang berhasil dikumpulkan berjumlah 704 record. Data mempunyai 20 atribut dengan 10 atribut data perilaku dan 10 karakteristik seseorang. Data dikumpulkan berdasarkan pertanyaan-pertanyaan yang sudah disesuaikan dengan DSM-V.(Thabtah,2017).

Tabel 3.1 Atribut Prediksi ASD

Attribute	Type
Umur	Number
Jenis Kelamin	String
Etnik	String
Kuning Waktu Lahir	Boolean (yes or no)
Keluarga Penyandang autis	Boolean (yes or no)
Membantu menggunakan aplikasi	String
Country of residence	String
Pernah melaksanakan tes sebelumnya	Boolean (yes or no)
Tipe Tes	Integer (0,1,2,3)
Jawaban No 1	Binary (0, 1)
Jawaban No 2	Binary (0, 1)
Jawaban No 3	Binary (0, 1)
Jawaban No 4	Binary (0, 1)
Jawaban No 5	Binary (0, 1)
Jawaban No 6	Binary (0, 1)
Jawaban No 7	Binary (0, 1)
Jawaban No 8	Binary (0, 1)
Jawaban No 9	Binary (0, 1)
Jawaban No 10	Binary (0, 1)

(Sumber : Respository UCI)

**Kerangka Penelitian**



Gambar 3.1. Kerangka Penelitian

Data mentah yang didapatkan berjumlah 704 kemudian dibagi kedalam dua kelompok data berupa data training dan data testing. Pembagiannya 80% data training dan 20% untuk data testing, berarti *data training* berjumlah 563 dan *data testing* sejumlah 141.

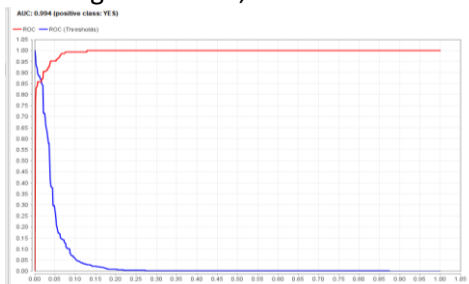
**IV. Hasil Dan Pembahasan Pengolahan Data Training Dengan Algoritma Naive Bayes**

Langkah pertama adalah mengolah data training menggunakan Algoritma Naive Bayes. Mendapatkan hasil sebagai berikut:

Tabel 4.1. Akurasi Data Training Menggunakan Naive Bayes

accuracy: 95,55%			
	true NO	true YES	class precision
pred. NO	400	11	97,32%
pred. YES	14	137	90,73%
class recall	96,62%	92,57%	

Dari tabel diatas terlihat akurasi pengolahan data training sebesar 95,55%



Gambar 4.1. Kurva ROC Data Training Naive Bayes

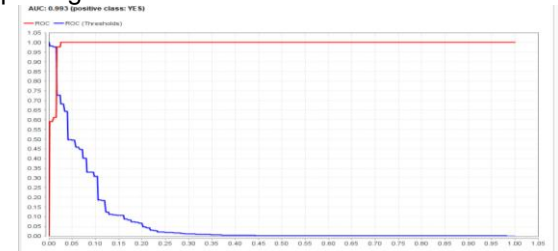
**Pengolahan Data Training Dengan Algoritma Naive Bayes berbasis Particle Swarm Optimization**

Data training sejumlah 563 kemudian diolah menggunakan Naive Bayes dipadukan dengan Algoritma optimasi yang sudah ditentukan sebelumnya, yaitu *Particle Swarm Optimization*. Hasil yang didapatkan dapat dilihat pada table 4.2

Tabel 4.2. Akurasi Data Training Dengan Naive Bayes PSO

accuracy: 97,62%			
	true NO	true YES	class precision
pred. NO	120	0	100,00%
pred. YES	4	44	91,67%
class recall	96,77%	100,00%	

Diperoleh akurasi sebesar 97,62%. Sedangkan pengolahan ROC mendapatkan hasil seperti pada gambar 4.2



Gambar 4.2. Kurva ROC Naive Bayes-PSO

Untuk mendapatkan hasil yang lebih akurat *Data Training* tersebut diolah lagi dengan memasukkan *Data Testing* kedalam model.

Tabel 4.3 Nilai Akurasi Pada Algoritma Naive Bayes

accuracy: 92,91%			
	true YES	true NO	class precision
pred. YES	39	8	82,98%
pred. NO	2	92	97,87%
class recall	95,12%	92,00%	

Akurasi yang diperoleh sebesar 92,91% Hasil pengolahan ROC untuk algoritma Naive Bayes



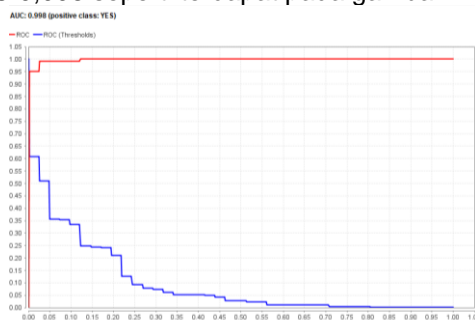
Gambar 4.3. Kurva ROC Algoritma Naive Bayes Hasil pengolahan ROC didapatkan angka AUC: 0.899 seperti yang terlihat dalam gambar 4.1.

## Pengolahan Data Dengan Algoritma Naïve Bayes Berbasis Particle Swarm Optimization

Tabel 4.4 Nilai Akurasi Pada Algoritma Naïve Bayes Dengan Particle Swarm Optimization

accuracy: 98.59%			
	true YES	true NO	class precision
pred. YES	40	1	97.56%
pred. NO	1	100	99.01%
class recall	97.56%	99.01%	

Mempunyai nilai akurasi sebesar 98,59% Hasil dari pengolahan ROC didapatkan angka AUC 0,998 seperti terdapat pada gambar 4.2



Gambar 4.4. Kurva ROC Naïve Bayes - PSO Hasil pengolahan data dengan Naïve Bayes dan Naïve Bayes dengan PSO dapat dilihat dalam tabel berikut:

Tabel 4.5. Komparasi Naïve Bayes dan Naïve Bayes PSO

Algoritma	Akurasi	ROC
Naïve Bayes	92,91%	0,989
Naïve Bayes- PSO	98,51%	0,998

## V. Kesimpulan Dan Saran

### a) Kesimpulan

1. Data set yang diolah menggunakan Algoritma Naïve Bayes mempunyai akurasi sebesar 92,91%.
2. Data set yang diolah menggunakan algoritma naïve bayes berbasis Particle Swarm Optimization mempunyai akurasi sebesar 98,59%.
3. Penerapan Particle Swarm Optimization dapat meningkatkan akurasi deteksi Autistic Spectrum Disorder sebesar 5,68%
4. Penerapan Particle Swarm Optimization terbukti meningkatkan akurasi pada Algoritma Naïve Bayes.

### b) Saran

1. Penelitian ini dilanjutkan dengan menerapkan Particle Swarm Optimization dalam berbagai algoritma.
2. Menerapkan Algoritma Optimasi lainnya.
3. Membantu untuk mendeteksi Autistic Spectrum Disorder pada orang dewasa.
4. Membuat aplikasi computer berdasarkan Algoritma Naïve Bayes- PSO.

### Daftar Pustaka

- [1] Ahmad Saleh, 2015, Klasifikasi Gejala Depresi Pada Manusia dengan Metode Naïve Bayes Menggunakan Java, Yogyakarta.
- [2] Al Fajar, Kemal, 2017. Bisakah Autisme Muncul Saat Usia Dewasa? Apa Saja Gejalanya? Diambil dari <https://hellosehat.com/pusat-kesehatan/autisme/penyebab-autisme-pada-orang-dewasa/> (10 Maret 2018)
- [3] American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C: American Psychiatric Association.
- [4] Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques (3rd upl.). USA: Elsevier.
- [5] Hermawati, Fajar Astuti. (2013). Data Mining. Yogyakarta: Andi
- [6] Lubis, Fauziah & Suwandi, Johns Fatriyadi, (2016). Paparan Prenatal Valproat dan Autism Spectrum Disorder (ASD) pada Anak. Lampung :Medical Jurnal Of Lampung University. Vol.5 No.3 : 85-90.
- [7] **Mujab, Saeful, (2013)**. Pencarian Model Terbaik Antara Algoritma C4.5 Dan C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Promosi Deposito. <http://eprints.dinus.ac.id/12380/> (10 Maret 2018)
- [8] S.-W. Fei, Y.-B. Miao, and C.-L. Liu, "Chinese Grain Production Forecasting Method Based on Particle Swarm Optimization-based Support Vector Machine," in Recent Patents on Engineering, Shanghai, 2009, pp. 8-12.
- [9] Thabtah, Fadi. 2017. Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. IMCHI'17.1-6