

# Deteksi Penyakit Alzheimer Menggunakan Algoritma Naïve Bayes dan Correlation Based Feature Selection

Siti Khotimatul Wildah<sup>1</sup>, Sarifah Agustiani<sup>2</sup>, M. Rangga Ramadhan S<sup>3</sup>, Windu Gata<sup>4</sup>,  
Hendri Mahmud Nawawi<sup>5</sup>

<sup>1,2,3,4,5</sup> STMIK Nusa Mandiri

e-mail: <sup>1</sup>siti.ska@nusamandiri.ac.id, <sup>2</sup>sarifah.sgu@nusamandiri.ac.id,

<sup>3</sup>rangga.mgg@nusamandiri.ac.id, <sup>4</sup>windu@nusamandiri.ac.id, <sup>5</sup>hendri.hiw@nusamandiri.ac.id

## Abstrak

Alzheimer merupakan kelainan berupa penimbunan plak atau protein tidak normal dalam otak sehingga menyebabkan hilangnya sel neuron dan menjadi salah satu pemicu penyakit demensia yang dapat mengakibatkan terhambatnya aktivitas sehari-hari karena penurunan daya ingat, kesulitan dalam berkomunikasi, tidak dapat berpikir jernih, terjadinya perubahan sikap dan perilaku hingga menimbulkan hilangnya kemampuan untuk mengurus diri sendiri. Di negara berpenghasilan tinggi penyakit ini diakui berada pada peringkat ke 7 sebagai penyakit fatal yang berujung pada kematian. Akan tetapi belum adanya obat yang dapat digunakan untuk menyembuhkan penyakit Alzheimer hingga saat ini. Oleh sebab itu pentingnya deteksi dini agar dapat memulai untuk merencanakan perawatan dan kebutuhan medis yang memadai. Penelitian ini bertujuan untuk melakukan deteksi penyakit Alzheimer dengan menerapkan metode klasifikasi Naïve Bayes dan seleksi atribut menggunakan Correlation Based Feature Selection pada dataset OASIS Longitudinal. Tahapan analisa data menggunakan metode CRISP-DM. Hasil penelitian ini, menunjukkan bahwa pada pengujian algoritma Naïve Bayes nilai akurasi yang didapatkan sebesar 93,83%, dan kurva ROC yang terbentuk memiliki nilai AUC sebesar 0,937% sedangkan pada pengujian algoritma Naïve Bayes dan Correlation Based Feature Selection menghasilkan nilai akurasi sebesar 94,64% dan kurva ROC yang terbentuk memiliki nilai AUC sebesar 0,945%. Sehingga dapat disimpulkan bahwa penerapan algoritma Naïve Bayes dan metode Correlation Based Feature Selection dapat meningkatkan nilai akurasi.

**Kata Kunci:** Alzheimer, Naïve Bayes, Correlation Based Feature Selection.

## Abstract

*Abstract - Alzheimer's is a disorder in the form of accumulation of plaque or abnormal proteins in the brain that causes loss of neuron cells and is one of the triggers of dementia which can result in inhibition of daily activities due to decreased memory, difficulty in communication, unable to think clearly, changes in attitude and behavior that results in a loss of ability to take care of oneself. In high-income countries, this disease is ranked 7th as a fatal disease that leads to death. But until now there has not been found a drug that can cure Alzheimer's disease. Therefore the importance of early detection in order to begin to plan for treatment and adequate medical needs. This study aims to detect Alzheimer's disease by applying the Naïve Bayes classification method and attribute selection using Correlation Based Feature Selection on the Longitudinal OASIS dataset. Stages of data analysis using the CRISP-DM method. The results of this study, showed that the Naïve Bayes algorithm testing obtained an accuracy value of 93.83%, and the formed ROC curve had an AUC value of 0.937% while the Naïve Bayes algorithm testing and Correlation Based Feature Selection resulted in an accuracy value of 94.64 % and the formed ROC curve has an AUC value of 0.945%. So it can be concluded that the application of the Naïve Bayes algorithm and the Correlation Based Feature Selection method can increase the accuracy value.*

**Keywords:** Alzheimer's, Naïve Bayes, Correlation Based Feature Selection.

## Pendahuluan

Penyakit Alzheimer ialah pemicu paling umum demensia yang terjadi pada

orang tua (J.S., Birks, & Harvey, 2018), dimana penyakit tersebut menyerang sistem saraf otak yang secara *irreversible* dapat



memicu hilangnya sel neuron (Indra, Marhaendraputro, & Hidayat, 2017) dan berakibat terganggunya aktivitas sehari-hari karena kebingungan dalam mencerna pertanyaan, ingatan yang berantakan (Novitasari, Puspitasari, Wulandari, & Foady, 2018), dan berujung pada hilangnya kemampuan untuk mengingat, kesulitan dalam berkomunikasi, berpikir jernih, terjadinya perubahan perilaku dan kemampuan untuk mengurus diri sendiri, (J.S. et al., 2018). penurunan kemampuan pada penderita Alzheimer dikarenakan telah rusaknya sel-sel syaraf pusat yang mengakibatkan fungsi kognitif tidak berjalan dengan baik.

Penurunan daya ingat pada penderita Alzheimer terjadi secara bertahap yang dialami selama kurun waktu tiga sampai sembilan tahun (Novitasari et al., 2018). Diperkirakan sekitar 44 juta orang diseluruh dunia pada tahun 2015 memiliki penyakit Alzheimer atau demensia (Cauwenberghe, Broeckhoven, & Sleegers, 2016). Pada tahap akhir penderita Alzheimer memerlukan perawatan yang intensif karena penyakit tersebut dapat berakibat fatal (Gaugler, James, Johnson, Scholz, & Weuve, 2016) yang berujung kematian dan merupakan penyebab utama kematian peringkat ke 7 di negara berpenghasilan tinggi (Dailey, 2016).

Faktor terbesar yang mempengaruhi penyakit Alzheimer adalah usia dan didiagnosis terjadi pada usia lebih dari 65 tahun akan tetapi tidak menutup kemungkinan orang yang berusia dibawah 65 tahun dapat memiliki penyakit tersebut. Faktor lain penyebab Alzheimer seperti Apolipoprotein E (APOE)  $\epsilon 4$  gene, riwayat keluarga, kerusakan kognitif ringan, faktor risiko penyakit kardiovaskular, tingkat pendidikan, hubungan sosial dan cedera otak traumatis (Gaugler, James, Johnson, Scholz, & Weuve, 2016).

Fakta lain mengenai penyakit ini ialah belum adanya obat yang dapat digunakan untuk menyembuhkan penyakit Alzheimer (Friskahaja, Ilhamsyah, & Barlian, 2018) serta belum banyak hal yang dapat terungkap mengenai penyakit tersebut, terutama perubahan biologis yang menyebabkan seseorang dapat memiliki penyakit ini, kemudian bagaimana tingkat percepatan penyakit ini dapat berlangsung secara berbeda pada setiap orang. Oleh sebab itu pentingnya deteksi dini agar dapat memulai untuk merencanakan perawatan

dan kebutuhan medis yang memadai serta dapat dilakukan pula untuk pengujian obat baru penyakit tersebut (Samper-González et al., 2018). Dalam memprediksi penyakit Alzheimer salah satu metode yang dapat digunakan adalah dengan menerapkan metode *Data Mining*. *Data Mining* digunakan untuk mendapatkan pola maupun informasi yang menonjol dalam suatu data dengan menggunakan metode maupun teknik tertentu (Yuli, 2017). Metode atau teknik yang digunakan dalam penelitian ini yaitu menerapkan metode klasifikasi, dimana metode tersebut menggambarkan sebuah proses pencarian model data untuk memprediksi suatu kelas target dari atribut kelas lain yang belum diketahui labelnya (Annur, 2018). Metode klasifikasi *Data Mining* diantaranya *Bayes*, *Decision Tree*, *Neural Network*, *SVM*, *Nearest Neighbour*, *Fuzzy Logic*, *Logistic*, *Classification via Regression*.

Penelitian terkait mengenai penyakit Alzheimer menggunakan teknik klasifikasi *Data Mining* telah banyak dilakukan, seperti penelitian yang dilakukan oleh C R Aditya dan M B Sanjay Pande mengenai Algoritma *Affiliation* yang digunakan untuk mendeteksi Penyakit Alzheimer (Aditya & Pande, 2016), penelitian yang dilakukan oleh Gopi Battinenia, Nalini Chintalapudi, dan Francesco Amentaa mengenai perhitungan performa deteksi demensia menggunakan algoritma *Support Vectore Machine* (SVM) (Battineni, Chintalapudi, & Amenta, 2019), dan penelitian yang dilakukan oleh El Mehdi Benyoussef, Abdeltif Elbyed, dan Hind El Hadiri mengenai pendekatan *Data Mining* untuk diagnosa penyakit Alzheimer, dalam penelitian ini dilakukan perbandingan algoritma *Decision Tree*, *Logistic Regression* dan *Discriminant Analysis* (Benyoussef, Elbyed, & Hadiri, 2017).

Pada penelitian ini, dilakukan penerapan algoritma *Naïve Bayes* pada dataset Alzheimer OASIS *Longitudinal* dan melakukan seleksi atribut dengan menggunakan metode *Correlation Based Feature Selection*. Pemilihan atribut bertujuan untuk mengurasi dimensi data dengan menghilangkan atribut yang tidak relevan, karena atribut yang tidak relevan dapat mempengaruhi tingkat akurasi yang dihasilkan pada pengklasifikasian data (Utomo, 2020). Pemilihan algoritma *Naïve Bayes* pada penelitian ini, dikarenakan algoritma tersebut termasuk kedalam salah satu algoritma populer dalam *Machine*

*Learning* dan *Data Mining* yang mampu melakukan pembelajaran induktif secara optimal (Rosi, Fauzi, & Perdana, 2018) serta algoritma *Naïve Bayes* memiliki probabilitas yang sederhana dan tidak memerlukan parameter yang rumit (Rianto & Wahono, 2015).

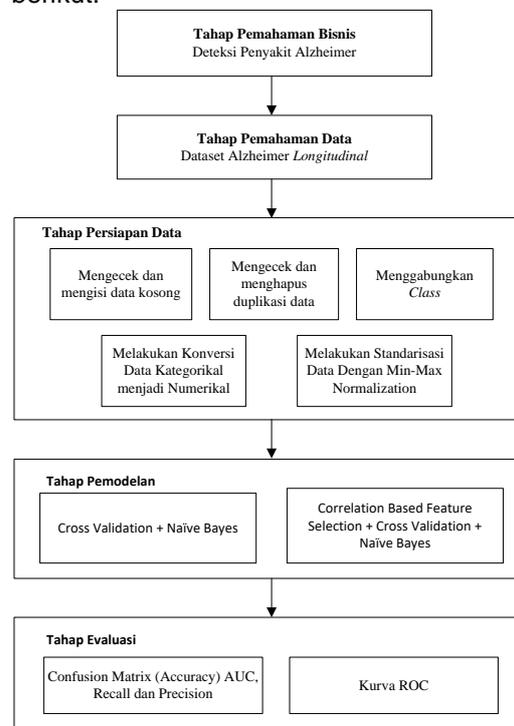
Metode *Naïve Bayes* terbukti dapat memberikan performa nilai akurasi yang tinggi dalam memprediksi data penyakit seperti pada penelitian yang dilakukan oleh Tutus Praningki dan Indra Budi mengenai Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, *Naive Bayes*, dan k-NN (2017) yang menyatakan bahwa algoritma *Naïve Bayes* memiliki performa nilai akurasi yang tinggi dibandingkan dengan metode CART maupun k-NN (Praningki & Budi, 2017) dan Penelitian yang dilakukan oleh Fatmawati mengenai Perbandingan Algoritma Klasifikasi *Data Mining* Model C4.5 Dan *Naïve Bayes* Untuk Prediksi Penyakit Diabetes, dari perbandingan kedua algoritma tersebut dinyatakan bahwa algoritma *Naïve Bayes* lebih akurat memprediksi penyakit diabetes dibandingkan dengan algoritma C.45 (Fatmawati, 2016).

Sedangkan pemilihan atribut menggunakan metode *Correlation Based Feature Selection* dikarenakan metode tersebut merupakan teknik penyeleksian atribut dalam sebuah dataset yang dilakukan dengan menghitung keterkaitan atribut prediksi dengan atribut target, dimana atribut yang terpilih merupakan atribut yang memiliki keterkaitan tinggi antara atribut prediksi dengan atribut target, tetapi memiliki nilai keterkaitan yang rendah dengan atribut lain (Anwar, Septian, & Septiana, 2019).

Berdasarkan latar belakang tersebut, maka untuk memprediksi penyakit Alzheimer pada penelitian ini mencoba menerapkan metode *Data Mining* dengan pemilihan algoritma *Naïve Bayes* yang terbukti memberikan performa akurasi yang tinggi pada prediksi data penyakit. Melakukan *data preprocessing* untuk dapat meningkatkan nilai akurasi pada algoritma tersebut dengan cara mengecek dan mengisi data yang kosong, mengecek dan menghapus duplikasi data, menggabungkan *class*, dan melakukan pemilihan atribut menggunakan metode *Correlation Based Feature Selection*.

## Metode Penelitian

Metode penelitian ini, menggunakan model eksperimen dengan desain eksperimen menggunakan proses CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahapan (T & Larose, 2015). Akan tetapi tahapan pada penelitian ini hanya menggunakan 5 tahapan yaitu tahap pemahaman bisnis, tahap pemahaman data, tahap persiapan data, tahap pemodelan dan tahap evaluasi model. Tahapan ini dapat digambarkan sebagai berikut:



Sumber: Hasil Penelitian (2020)

**Gambar 1. Tahapan Metode Penelitian**

### 1. Business Understanding

Tahap pemahaman bisnis bertujuan untuk menentukan capaian dari penelitian, ruang lingkup yang dibutuhkan pada penelitian secara lengkap, merumuskan batasan dan permasalahan serta strategi awal yang akan dilakukan agar dapat mencapai tujuan (Feblian & Daihani, 2017). Capaian pada penelitian ini yaitu dapat melakukan deteksi dini penyakit Alzheimer

### 2. Data Understanding

Tahap pemahaman data dimulai dengan pengumpulan data kemudian setelah data terkumpul dilanjutkan dengan proses pemahaman data itu sendiri seperti merumuskan masalah-masalah yang terdapat dalam data, tingkat kualitas sampai dengan menemukan gambaran mengenai

perkiraan yang akan terjadi dari informasi yang ada pada data tersebut (Hakim, 2020).

### 3. Data Preparation

Tahapan pengolahan data merupakan seluruh rangkaian kegiatan yang bertujuan untuk membangun sebuah dataset yang siap dipakai untuk pemodelan algoritma. (Setiawan, 2016). Tahap ini meliputi pembersihan data dari nilai data yang kosong, duplikasi data, normalisasi data, standarisasi data maupun pemilihan atribut.

### 4. Modelling

Tahap pemodelan dilakukan proses pemilihan algoritma yang sesuai dengan permasalahan yang ada. Pada penelitian ini akan dilakukan pemodelan menggunakan algoritma naive bayes dengan melakukan 2 model pemrosesan yaitu algoritma *Naive Bayes* menggunakan *k-Fold Cross Validation* dan algoritma *Naive Bayes* menggunakan *k-Fold Cross Validation* dan seleksi atribut menggunakan *Correlation Based Feature Selection*.

### 5. Evaluation

Tahap evaluasi bertujuan untuk menguji seberapa akurat sebuah model yang diterapkan. Pada penelitian ini model yang dipilih akan dievaluasi menggunakan metode *Confusion Matriks* dan kurva ROC.

## Hasil dan Pembahasan

### 1. Tahap Pemahaman Bisnis (*Business Understanding*)

Tujuan dari penelitian ini yaitu mendeteksi penyakit Alzheimer dengan menerapkan algoritma *Naive Bayes* dan melakukan seleksi atribut menggunakan metode *Correlation Based Feature Selection*. Dataset yang digunakan terdiri dari koleksi longitudinal yang berasal dari 150 subjek berusia 60 hingga 96 tahun. Dimana setiap subjek dilakukan pemeriksaan sebanyak dua atau lebih selama satu tahun sehingga total data yang diperoleh sebanyak 373 *record*. Seluruh subjek yang ada pada dataset ini bertangan kanan yang terdiri dari 72 subjek dinyatakan tidak memiliki demensia, 64 subjek dinyatakan demensia serta 14 subjek lainnya masuk kedalam kategori masih di konversikan. Akan tetapi pada tahap akhir pemeriksaan 14 subjek tersebut dinyatakan mengalami demensia (Marcus, Fotenos, Csernansky, Morris, & Buckner, 2010).

### 2. Tahap Pemahaman Data (*Data Understanding*)

Tahap pemahaman data pada penelitian ini yaitu menganalisa dan memahami dataset Alzheimer *Longitudinal* yang memiliki 15 atribut, 14 diantaranya sebagai atribut prediksi yaitu MRIID, SubjectID, Visit, MR Delay, CDR, Hand, Gender, Age, Educ, SES, MMSE, eTiv, nWBV, dan ASF. Sedangkan atribut target yaitu Group. Berikut daftar atribut yang ada pada dataset Alzheimer Longitudinal.

**Tabel 1. Atribut dan Nilai Kategori**

Atribut	Nilai
Group	Kelas <i>Demented</i> , <i>Non-demented</i> dan <i>Converted</i> .
MRIID	ID Tes 1 - 354.
SubjectID	ID pemeriksaan subjek 1-142
Visit	Jumlah pemeriksaan subjek
MR Delay	Nilai keterlambatan subjek sejak terakhir melakukan pemeriksaan.
CDR	<i>Clinical Dementia Rating</i> . (Tingkat demensia klinis) 0 = <i>No Dementia</i> , 0.5 = <i>Very Mild AD</i> , 1 = <i>Mild AD</i> , 2 = <i>Moderate AD</i> . (Morris, 1993)
Hand	<i>Right</i> (Semua subjek yang diteliti bertangan kanan)
Gender	Jenis kelamin <i>Male</i> (Laki-Laki) dan <i>Female</i> (Perempuan)
Age	Usia subjek pada saat dilakukan pemeriksaan
EDUC	<i>Years of education</i> (Tingkat Pendidikan)
SES	<i>Socioeconomic statue</i> (Status Sosial Ekonomi ) 1 ( <i>highest statue</i> ) - 5 ( <i>lowest statue</i> ). (Hollingshead, 1957)
MMSE	<i>Mini-Mental State Examination value</i> . (nilai pemeriksaan keadaan mental) 0 ( <i>worst value</i> ) - 30 ( <i>best value</i> ). (Folstein, Folstein, & McHugh, 1975)
eTIV	<i>Estimated total intracranial volume</i> (cm <sup>3</sup> ) / perkiraan total volume intrakranial (Buckner et al., 2004)
nWBV	<i>Normalized whole-brain volume, expressed as a percent of all voxels</i> (volume keseluruhan otak yang dinormalisasikan di ekspersika dengan persentase dari semua voxel) (Fotenos et al., 2005)
ASF	<i>Atlas Scale Factor; volume scaling factor for brain size</i> (Faktor Skala Atlas; faktor penskalaan volume untuk ukuran otak)

Sumber: G.Ertek, B.Tokdil & I.Gunaydin (2014)

Pengolahan data dilakukan dengan menggunakan bahasa pemrograman *Python*. *Python* memiliki *library* untuk

melakukan pengolahan data, hal pertama yang dilakukan adalah memanggil seluruh *library* yang dibutuhkan dan *file dataset* yang akan digunakan.

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
sns.set()
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler,MinMaxScaler
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import KFold, cross_val_score, cross_val_predict
from sklearn.metrics import confusion_matrix, roc_auc_score, auc
from sklearn.metrics import classification_report, roc_curve, accuracy_score
from sklearn.metrics import recall_score, precision_score
from sklearn.model_selection import train_test_split
df = pd.read_csv('oasis_longitudinal.csv')
```

Sumber: Hasil Penelitian (2020)  
**Gambar 2 Library dan Dataset**

### 3. Tahap Persiapan Data (*Data Preparation*)

Pada tahap ini dilakukan pengolahan data Alzheimer agar data siap digunakan pada saat pemodelan. Adapun pengolahan data pada penelitian ini, sebagai berikut:

#### a. Mengecek dan Mengisi data kosong

Tahap persiapan data untuk menangani data kosong dilakukan dengan mengisi nilai *mean*, setelah sebelumnya dilakukan pengecekan data kosong.

```
df.isna().sum()
Subject ID      0
MRI ID          0
Group           0
Visit           0
MR Delay        0
M/F             0
Hand            0
Age             0
EDUC            0
SES             19
MMSE            2
CDR             0
eTIV            0
nWBV            0
ASF             0
dtype: int64
```

Sumber: Hasil Penelitian (2020)  
**Gambar 3 Mengecek Data Kosong**

Data kosong pada dataset ini ditemukan pada kolom SES sebanyak 19 *record* dan pada kolom MMSE sebanyak 2 *record*, maka untuk mengatasi hal tersebut dilakukan pengisian dengan nilai *mean*.

```
df["SES"].fillna(df["SES"].mean(), inplace=True)
df["MMSE"].fillna(df["MMSE"].mean(), inplace=True)
```

Sumber: Hasil Penelitian (2020)  
**Gambar 4 Mengisi Data Kosong**

#### b. Mengecek dan Menghapus Duplikasi Data

Langkah *preprocessing* selanjutnya adalah mengecek duplikasi data pada *dataset* Alzheimer

```
df.duplicated().sum()
0
```

Sumber: Hasil Penelitian (2020)  
**Gambar 5 Mengecek Duplikasi Data**

#### c. Menggabungkan Class

Dalam *dataset* Alzheimer OASIS *Longitudinal* terdapat 3 *class* diantaranya *Demented*, *Non-demented* dan *Converted*. Akan tetapi *class* *Converted* dikategorikan kedalam kelas *Demented* (Marcus et al., 2010).

```
df['Group'] = df['Group'].replace(['converted'], ['demented'])
df.Group.value_counts()
NonDemented    198
Demented       133
Name: Group, dtype: int64
```

Sumber: Hasil Penelitian (2020)  
**Gambar 6 Mengecek Duplikasi Data**

#### d. Mengubah Data Kategori Menjadi Numerik

Tahap persiapan data selanjutnya yaitu mengubah data kategori menjadi numerik. Karena dalam pemodelan menggunakan *python* seluruh data harus diubah menjadi numerik agar dapat diproses.

#### e. Melakukan Standarisasi Data

Standarisasi data dilakukan untuk menghindari salah satu atribut yang mendominasi. Pada penelitian ini metode standarisasi data menggunakan *Min-Max Normalization*.

#### f. Menghapus atribut

Terdapat 14 atribut prediksi, 3 diantaranya tidak akan digunakan yaitu MRIID, SubjectID dan Hand. Atribut MRIID dan SubjectID tidak memiliki pengaruh terhadap hasil analisa karena atribut tersebut hanya berisi Nomor ID dari pemeriksaan, sedangkan atribut *Hand* berisi nilai keseluruhan yang sama yaitu subjek bertangan kanan.

### 4. Tahap Pemodelan (*Modeling*)

Tahap pemodelan dilakukan perbandingan dengan menerapkan algoritma *Naïve Bayes* menggunakan *k-Fold Cross Validation* kemudian algoritma *Naïve Bayes* dan metode *Correlation Based Feature Selection* menggunakan *k-Fold Cross Validation*. Perbandingan ini bertujuan untuk melakukan pengujian

pengaruh metode seleksi atribut dalam meningkatkan nilai akurasi.

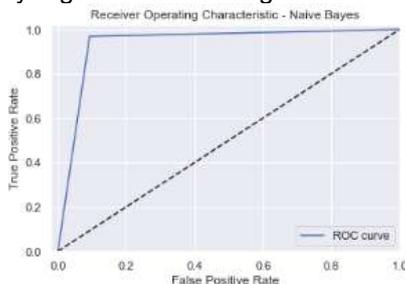
**a. Naïve Bayes Menggunakan k-Fold Cross Validation**

Langkah pertama yang dilakukan yaitu menentukan variabel target dan variabel prediksi. Variabel target yaitu variabel *Group* sedangkan variabel prediksi yaitu *Visit*, MR Delay, CDR, *Hand*, *Gender*, *Age*, *Educ*, *SES*, *MMSE*, *eTiv*, *nWBV*, dan *ASF*. Kemudian menerapkan algoritma Naïve Bayes menggunakan metode *k-Fold Cross Validation*. Sehingga hasil yang diperoleh pada dataset ini sebagai berikut:



Sumber: Hasil Penelitian (2020)  
**Gambar 7 Hasil klasifikasi NB**

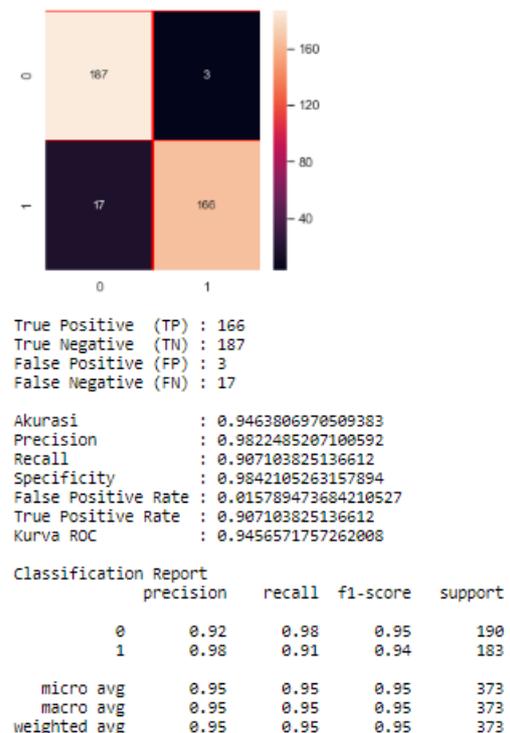
Pemodelan menggunakan algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 93,83%, nilai *precision* sebesar 0,915%, nilai *recall* sebesar 0,968%, nilai *specificity* sebesar 0,907%, nilai FPR sebesar 0,092%, nilai TFR sebesar 0,968% dan nilai ROC sebesar 0,937% serta kurva ROC yang terbentuk sebagai berikut:



Sumber: Hasil Penelitian (2020)  
**Gambar 8 Kurva ROC NB**

**b. Naïve Bayes dan Correlation Based Feature Selection Menggunakan k-Fold Cross Validation**

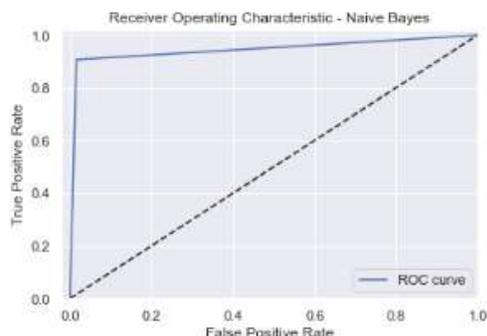
Langkah pertama yang dilakukan adalah menerapkan seleksi atribut menggunakan metode *Correlation Based Feature Selection*. Setelah dilakukan penerapan metode seleksi atribut, terdapat 2 atribut yang memiliki nilai keterkaitan yang tinggi dengan atribut target yaitu atribut MR Delay dan ASF. Maka kedua atribut tersebut harus dihilangkan, sehingga jumlah atribut yang tersisa yaitu atribut *Group*, *Visit*, *Gender*, *Age*, *EDUC*, *SES*, *MMSE*, *CDR*, *eTiv* dan *nWBV*. Langkah selanjutnya sama seperti yang dilakukan sebelumnya yaitu menentukan variabel target dan variabel prediksi, kemudian menerapkan algoritma Naïve Bayes dengan menggunakan metode *k-Fold Cross Validation*. Sehingga hasil yang diperoleh pada dataset ini sebagai berikut:



Sumber: Hasil Penelitian (2020)  
**Gambar 9 Klasifikasi NB dan CBFS**

Pemodelan menggunakan metode Naïve Bayes dan *Correlation Based Feature Selection* menghasilkan nilai akurasi sebesar 94,64%, nilai *precision* sebesar 0,982%, nilai *recall* sebesar

0,907%, nilai *specificity* sebesar 0,984%, nilai FPR sebesar 0,015%, nilai TFR sebesar 0,907% dan nilai ROC sebesar 0,945% serta kurva ROC yang terbentuk sebagai berikut:



Sumber: Hasil Penelitian (2020)

**Gambar 10 Kurva ROC NB dan CBFS**

#### 5. Tahap Evaluasi (*Evaluation*)

Setelah dilakukan pemodelan dengan algoritma *Naïve Bayes* menggunakan 2 metode pengujian, dimana hasil terbaik yaitu algoritma *Naïve Bayes* dengan seleksi atribut *Correlation Based Feature Selection*. Berikut tabel perbandingan pengujian kedua model tersebut.

**Tabel 2. Hasil Evaluasi Model**

	<i>Naïve Bayes</i>	<i>Naïve Bayes dan CBFS</i>
<b>Accuracy</b>	93,83%	94,64%
<b>AUC</b>	0,937%	0,945%
<b>Recall</b>	0,968%	0,907%
<b>Precision</b>	0,915%	0,982%
<b>FPR</b>	0,092%	0,015%
<b>TFR</b>	0,968%	0,907%

Sumber: Hasil Penelilitan (2020)

#### Kesimpulan

Berdasarkan hasil pengujian dan analisis data menggunakan algoritma *Naïve Bayes* untuk mendeteksi penyakit Alzheimer diperoleh nilai akurasi sebesar 93,83% dan nilai AUC sebesar 0,937% sedangkan peneran algoritma *Naïve Bayes* seleksi atribut menggunakan *Correlation Based Feature Selection* diperoleh nilai akurasi sebesar 94,64% dan nilai AUC sebesar 0,945%. Selisih dari perbandingan percobaan ini sebesar 0,81% untuk nilai akurasi dan 0,008% untuk nilai AUC. Performa nilai akurasi dan AUC yang dihasilkan dengan penerapan metode seleksi atribut dapat disimpulkan lebih baik dibandingkan dengan hanya menggunakan

algoritma *Naïve Bayes* akan tetapi pada penelitian ini tidak meningkatkan nilai akurasi secara signifikan.

Agar penelitian ini bisa ditingkatkan, terdapat beberapa saran yang dapat diterapkan untuk pengembangan penelitian lebih lanjut, yakni dengan menggunakan jumlah data yang lebih besar, sehingga hasil pengujian yang didapatkan akan lebih baik, melakukan pengujian terhadap metode klasifikasi *Data Mining* lainnya seperti *Logistic Regression*, *Random Forest*, *Fuzzy Logic*, *Certainty Factor* atau dengan mencoba menerapkan metode seleksi atribut lainnya seperti *Particle Swarm Optimization*, *Genetic Algorithm*, *Ant Colony Optimization* dengan harapan dapat menghasilkan nilai akurasi yang lebih tinggi. Serta melakukan pengembangan perangkat lunak kedalam bentuk *mobile*, *website* maupun *desktop*.

#### Referensi

- Aditya, C. R., & Pande, M. B. S. (2016). An Algorithmic Approach for Alzheimer ' s Disease detection from Non- Image Data. *International Journal of Current Engineering and Technology*, 6(3), 784–787.
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode, 10, 160–165.
- Anwar, S., Septian, F., & Septiana, R. D. (2019). Klasifikasi Anomali Intrusion Detection System ( IDS ) Menggunakan Algoritma *Naïve Bayes Classifier* dan *Correlation-Based Feature Selection*, 2(4), 135–140.
- Battineni, G., Chintalapudi, N., & Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, 16(May), 100200. <https://doi.org/10.1016/j.imu.2019.100200>
- Benyoussef, E. M., Elbyed, A., & Hadiri, H. El. (2017). Data Mining Approaches for Alzheimer ' s Disease Diagnosis State of Art, 1, 619–631. <https://doi.org/10.1007/978-3-319-68179-5>
- Cauwenberghe, C. Van, Broeckhoven, C. Van, & Sleegers, K. (2016). Open The genetic landscape of Alzheimer disease: clinical implications and perspectives, 18(5). <https://doi.org/10.1038/gim.2015.117>

- Dailey, C. (2016). The Impact of Alzheimer's Disease -The Silent Killer. *JCCC Honors Journal*, 7(2), 1–16. Retrieved from [http://scholarspace.jccc.edu/honors\\_journal%5Cnhttp://scholarspace.jccc.edu/honors\\_journal/vol7/iss2/1](http://scholarspace.jccc.edu/honors_journal%5Cnhttp://scholarspace.jccc.edu/honors_journal/vol7/iss2/1)
- Ertek, G., Tokdil, B., & Günaydin, İ. (2014). Risk Factors and Identifiers for Alzheimer's Disease ;, 1–11.
- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data Mining Model C4 . 5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes. *Jurnal Techno Nusa Mandiri*, XIII(1), 50–59.
- Feblian, D., & Daihani, D. U. (2017). Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada PT X, 1–12.
- Friskahaja, H. H., Ilhamsyah, & Barlian, Y. A. (2018). Perancangan Kampanye Permainan Teka Teki Silang Sebagai Pencegahan Penyakit Alzheimer Di Usia Dewasa, 5(3), 1827–1831.
- Gaugler, J., James, B., Johnson, T., Scholz, K., & Weuve, J. (2016a). 2016 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*, 12(4), 459–509. <https://doi.org/10.1016/j.jalz.2016.03.001>
- Gaugler, J., James, B., Johnson, T., Scholz, K., & Weuve, J. (2016b). Alzheimer's Association Report 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 12(4), 459–509. <https://doi.org/10.1016/j.jalz.2016.03.001>
- Hakim, N. N. (2020). Implementasi Machine Learning pada Sistem Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia, 03(01), 25–35.
- Indra, M. R., Marhaendraputro, E. A., & Hidayat, R. R. (2017). Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata, 1–4.
- J.S., B., Birks, J. S., & Harvey, R. (2018). Donepezil for dementia due to Alzheimer's disease. *The Cochrane Database of Systematic Reviews*, (3), CD001190. <https://doi.org/10.1002/14651858.CD001190.pub3>. [www.cochranelibrary.com](http://www.cochranelibrary.com)
- Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2010). Open Access Series of Imaging Studies : Longitudinal MRI Data in Nondemented and Demented Older Adults, 2677–2684.
- Novitasari, D. C. R., Puspitasari, W. T., Wulandari, P., & Foeady, A. Z. (2018). Klasifikasi Alzheimer Dan Non Alzheimer Menggunakan Fuzzy C-Mean , Gray Level Co- Occurrence Matrix Dan Support Vector Machine, 04(02), 83–89.
- Praningki, T., & Budi, I. (2017). Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART , Naive Bayes , dan k-NN. *Citec Journal*, 4(2).
- Rianto, H., & Wahono, R. S. (2015). Resampling Logistic Regression untuk Penanganan Ketidakseimbangan Class pada Prediksi Cacat Software. *Journal of Software Engineering*, 1(1), 46–53.
- Rosi, F., Fauzi, M. A., & Perdana, R. S. (2018). Prediksi Rating Pada Review Produk Kecantikan Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(5), 1991–1997.
- Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., ... Colliot, O. (2018). Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. *NeuroImage*, 183(March), 504–521. <https://doi.org/10.1016/j.neuroimage.2018.08.042>
- Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru ( Studi Kasus : Politeknik LP3I Jakarta ), 3(1), 76–92.
- T, L. D., & Larose, C. D. (2015). *Data Mining and Predictive Analytics* (Second Edi). Canada: Simultaneously.
- Utomo, D. P. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung, 4(April), 437–444. <https://doi.org/10.30865/mib.v4i2.2080>
- Yuli, M. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.