

Klasifikasi Data Nasabah yang Berpotensi Membuka Simpanan Deposito Menggunakan Algoritme Naive Bayes

Paulina Mangande¹, Eko Hari Parmadi²

^{1,2}Informatika, Fakultas Sains dan Teknologi, Universitas Sanata Dharma

E-mail: ¹ Paulinamangande619@gmail.com, ²hari@usd.ac.id

ABSTRAK

Deposito merupakan salah satu instrumen investasi yang populer dipilih karena beragam keuntungan yang ditawarkan dibandingkan dengan tabungan biasa. Keuntungan yang ditawarkan dari deposito adalah suku bunga yang relatif tinggi dibandingkan dengan tabungan, serta minimnya risiko yang harus dihadapi dibandingkan dengan instrumen investasi lainnya seperti investasi saham. Namun memilih atau mencari nasabah yang berpotensi membuka simpanan deposito sering menjadi permasalahan tersendiri terutama bagi pihak bank yang menawarkan deposito. Penelitian ini menggunakan algoritme naïve bayes untuk klasifikasi serta data publik yang diperoleh dari situs UCI Repository Machine Learning berjudul “Bank Marketing Data Set” yang didapat dari program pemasaran langsung dari sebuah bank di Portugis dengan jumlah data 41188 dengan 20 atribut dan 1 output. Tujuan penelitian ini adalah menentukan atribut-atribut yang berpengaruh dalam klasifikasi data nasabah yang berpotensi membuka simpanan deposito serta akurasi dari klasifikasi tersebut menggunakan algoritme naïve bayes. Hasil pengujian yang dilakukan terhadap 18559 data menggunakan 3-fold cross validation menghasilkan tingkat keakuratan sebesar 86,6419%. Adapun atribut-atribut yang berpengaruh dalam proses klasifikasi adalah Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m.

Kata kunci : deposito, klasifikasi, naïve bayes

1. PENDAHULUAN

Menurut Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 tentang perbankan, dijelaskan bahwa bank adalah suatu badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkan kepada masyarakat dalam bentuk kredit atau bentuk lainnya dalam meningkatkan taraf hidup masyarakat. Umumnya bank sendiri membutuhkan dana tambahan dari berbagai pihak diantaranya dari para nasabah yang dapat digunakan sebagai sumber dana yang berupa nama rekening giro, tabungan dan deposito. Salah satu dana tambahan dari nasabah yang digunakan sebagai sumber dana adalah deposito. Para nasabah berasal dari latar belakang yang berbeda-beda antara lain umur, pekerjaan, pendidikan, status, dan lain sebagainya. Masalahnya adalah pihak bank sering mengalami kesulitan dalam mencari nasabah yang berpotensi membuka deposito. Melalui data nasabah yang ada serta proses klasifikasi menggunakan algoritme naïve bayes data nasabah akan diolah sehingga menghasilkan klasifikasi nasabah yang potensial membuka deposito. Adapun data nasabah yang digunakan adalah data publik yang diperoleh dari situs UCI Repository Machine Learning berjudul “Bank Marketing Data Set” yang didapat dari program pemasaran langsung dari sebuah bank di Portugis dengan jumlah data 41188 dengan 20 atribut dan 1 output.

Metode yang dapat digunakan untuk mengklasifikasi dalam kumpulan data dari website UCI Repository Machine Learning telah meneliti tentang klasifikasi data nasabah yang berpotensi membuka simpanan deposito dengan menggunakan algoritme Decision Tree dengan Penerapan Algoritme C4.5.. Hasil dari penelitian tersebut memiliki akurasi 82,19%. [1]. Penelitian tentang klasifikasi data nasabah berpotensi terkena kredit macet dalam pembayaran tagihan kartu kredit di bank swasta di Yogyakarta dengan menggunakan metode Naive Bayes. Hasil dari penelitian tersebut memiliki akurasi 77,28% [2]. Algoritme Naïve Bayes juga banyak digunakan dalam berbagai penelitian tentang klasifikasi diantaranya Implementasi Algoritma Naive Bayes pada Sistem Pendukung Keputusan Rekomendasi Wisata di Lamongan [3] dengan kriteria yang digunakan dalam sistem pendukung keputusan mencakup waktu buka, fasilitas, harga masuk, jarak, kebersihan, dan bintang wisata. Sistem ini memiliki akurasi sebesar 94%. Penelitian yang dilakukan oleh Preawestina dkk, yang berjudul Sistem Pakar Diagnosa Gangguan Pencernaan Balita Dengan Metode Naive Bayes [4] menghasilkan tingkat akurasi sebesar 87%. Algoritme Naïve Bayes merupakan salah satu algoritma klasifikasi teks atau dokumen yang

penggunaannya sederhana dan mudah dalam implementasi. Algoritma Naïve Bayes memiliki kelebihan diantaranya mampu untuk menangani data kategorikal, tahan terhadap fitur yang tidak relevan, dan memiliki kinerja yang baik terhadap data yang tidak seimbang serta cocok untuk diaplikasikan pada masalah klasifikasi dengan lebih dari dua kelas (multi-class classification). Algoritma Naïve Bayes berasumsi bahwa fitur-fitur yang digunakan independen satu sama lain. [5]

2. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan antara lain pengumpulan data, *preprocessing*, *3 fold cross validation*, pemodelan dengan algoritme naïve bayes, penghitungan akurasi. Berikut ini akan dipaparkan setiap tahapan tersebut:

Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data publik yang diperoleh dari situs *UCI Repository Machine Learning* berjudul “*Bank Marketing Data Set*” yang didapat dari program pemasaran langsung dari sebuah bank di Portugis dengan jumlah data 41188 dengan 20 atribut dan 1 output. Data yang diperoleh terdiri dari : *age, job, material, education, default, housing, loan, contact, moth, day of week, duration, campaign, pdays, previous, poutcome, emp.var.rate, cons.price.idx, cons.idx, euribor3m, nr.employed, Y/N*. Adapun keterangan dari data atribut nasabah dapat dilihat pada tabel 1 berikut ini:

Tabel 1. Data Atribut Nasabah

No	Data	Keterangan dan Nilai
1	Age	Umur nasabah bank
2	Job	Jenis pekerjaan nasabah bank (admin/bluecollar,entrepreneur/housemaid /management /retired/selftempoyed/services/student/technician/unemployed/unknown)
3	Material	Status pernikahan (divorced/merried/single)
4	Education	Pendidikan nasabah (basic 4y/basic 6y/basic 9y/high school/illiterate/professional course/university degree/unknown)
5	Default	Apakah mempunyai kredit gagal/macet? (yes/no/unknown)
6	Housing	Apakah mempunyai kredit pinjaman rumah? (yes/no/unknown)
7	Loan	Apakah mempunyai pinjaman pribadi (yes/no/unknown)
8	Contact	Jenis komunikasi yang digunakan nasabah (cellular/telephone)
9	Month	Bulan terakhir menghubungi nasabah
10	Day_of_week	Hari terakhir kontak dengan nasabah
11	Duration	Durasi terakhir menghubungi nasabah dalam detik
12	Campaign	Jumlah kontak yang dilakukan selama promosi ini dan untuk nasabah ini
13	Pdays	Jumlah hari yang berlalu setelah nasabah terakhir dihubungi dan promosi sebelumnya
14	Previous	Jumlah kontak dilakukan sebelum promosi ini dan untuk klien ini
15	Poutcome	Hasil dari penawaran marketing sebelumnya (failure/nonexistent/success)
16	Emp.var.rate	Variasi tingkat pekerjaan
17	Cons.price.idx	Indeks harga konsumen
18	Cons.conf.idx	Indeks kepercayaan konsumen
19	Euribor3m	Tingkat eurbor 3 bulan
20	Nr.employed	Jumlah karyawan
21	Y	Apakah nasabah berlangganan deposito yang ditawarkan (yes/no)

Preprocessing

Tahap *prosprocessing* diawali dengan seleksi atribut, yaitu memilih atribut mana saja yang akan digunakan dalam sistem. Data akan dirangking atau diurutkan berdasarkan bobotnya yang terpenting kemudian data tersebut dipilih menjadi beberapa atribut. Proses seleksi atribut pada penelitian ini dilakukan menggunakan Tools Weka 3.8.3 menggunakan metode Information Gain. Perhitungan information gain dapat dilakukan dengan menggunakan rumus di bawah ini [6]:

$$\text{Info}(D) = -\sum_{i=1}^m p_i \log_2(p_i) \dots\dots\dots(1)$$

Keterangan rumus :

- D : Jumlah seluruh sampel data
- m : Jumlah nilai pada atribut target (jumlah kelas klasifikasi)
- i : Maksimal nilai pada atribut target
- p_i : Jumlah sampel untuk kelas i

$$\text{Info X}(D) = -\sum_{j=1}^v \frac{|D_j|}{D} \times \text{info}(D_i) \dots\dots\dots(2)$$

Keterangan rumus :

- x : Atribut
- v : Suatu nilai yang mungkin untuk atribut A
- j : Maksimal nilai yang mungkin untuk atribut A
- D : Jumlah seluruh sampel data
- $|D_j|$:Jumlah sampel untuk nilai j
- D_i :Jumlah sampel untuk kelas i

Nilai information gain ini digunakan untuk mengukur efektifitas suatu atribut dalam pengklasifikasian data dan dapat dihitung dengan rumus di bawah ini :

$$\text{Gain}(X) = |\text{Info}(D) - \text{Info X}(D)| \dots\dots\dots(3)$$

Keterangan :

- X : Atribut
- $\text{Info}(D)$: Entropi untuk kelas D
- $\text{Info X}(D)$: Entropi untuk kelas D pada atribut X

Gambar 1. Menunjukkan hasil proses seleksi atribut dan rumus (1), (2), dan (3)

```
Ranked attributes:
0.1094127    11 duration
0.0969668    18 cons.conf.idx
0.0896296    20 nr.employed
0.0687236    16 emp.var.rate
0.0580302    19 euribor3m
0.0444837    13 pdays
0.0438343    15 poutcome
0.0380966     9 month
0.0277329    14 previous
0.0214045    17 cons.price.idx
0.0184393     1 age
0.0168012     8 contact
0.0142231     2 job
0.0083306     5 default
0.0042851    12 campaign
0.0034476     4 education
0.0020686     3 marital
0.0004645    10 day_of_week
0.0000997     6 housing
0.0000193     7 loan

Selected attributes: 11,18,20,16,19,13,15,9,14,17,1,8,2,5,12,4,3,10,6,7 : 20
```

Gambar 1. Perangkingan Atribut

Setelah data melalui proses seleksi atribut, selanjutnya data melalui tahap transformasi dan normalisasi. Data set yang memiliki range data yang cukup jauh dan susah untuk diproses akan dilakukan proses transformasi menggunakan *min max*. Misalkan $\min X$ dan $\max X$ adalah nilai minimum dan maksimum atribut X , maka normalisasi Min-Max akan menghasilkan nilai baru yaitu N_{data} dalam kisaran $[\min X \text{ baru}, \max X \text{ baru}]$, seperti pada rumus [6]:

$$N_{data} = ((v - \min) * (\max - \min)) / (\max - \min) + \min \dots\dots\dots(4)$$

Keterangan :

v : data yang akan dinormalisasi
 N_{data} : Data hasil normalisasi
 \min : Nilai minimum dari data
 \max : Nilai maksimum dari data
 \min : Batas minimum yang diberikan
 \max : Batas maksimum yang diberikan

Nilai pada set tersebut diubah skalanya dengan batas nilai minimum yang diberikan yakni 0 sementara untuk batas maximum yang diberikan yakni 1.

Proses *transformasi* yang dilakukan pada penelitian ini sebagai berikut:

Transformasi data pada atribut *duration* dan *pdays* dilakukan transformasi dengan proses pendistribusian tabel frekuensi berkelompok, dengan memilih atribut yang bernilai data *numerik* dan akan dicari intervalnya. Kemudian pada masing-masing atribut tersebut dicari nilai minimum (X_{\min}) dan nilai Maksimum (X_{\max}).

Untuk atribut *duration*, didapat :
Nilai minimum : 0
Nilai maksimum : 4918

Untuk atribut *duration*, didapat :
Nilai minimum : 0
Nilai maksimum : 4918

Setelah mendapatkan nilai minimum (X_{\min}) dan nilai maksimum (X_{\max}) dari setiap atribut, maka langkah selanjutnya yaitu menghitung nilai jangkauan dari masing-masing atribut dengan rumus :

$$J = X_{\max} - X_{\min} \dots\dots\dots(5)$$

Selanjutnya menghitung jumlah kelas interval dengan rumus [7]

$$k = 1 + 3,3 \log n. \dots\dots\dots(6)$$

Keterangan :

k = Banyak kelas interval
 n = Banyak data/Jumlah data

Sehingga jumlah interval pada penelitian ini sesuai rumus (6) adalah :

$$k = 1 + 3,3 \log 41188$$

$$k = 1 + 15,228$$

$$k = 16,228$$

$k = 17$ dibulatkan menjadi 17, sehingga jumlah kelas interval adalah 17.

Kemudian menentukan panjang interval untuk masing-masing atribut, dengan rumus :

$$p = \frac{j}{k} \dots\dots\dots(7)$$

Keterangan :

p = Panjang interval kelas
 j = Jangkauan
 k = Banyak kelas interval

Panjang kelas interval untuk atribut *duration* sesuai rumus (7) adalah :

$$p = \frac{4918}{17} = 289,29$$

Panjang kelas interval untuk atribut *pdays* :

$$P = \frac{999}{17} = 58,76$$

Langkah terakhir yaitu menentukan batas interval dari masing-masing atribut yang terpilih :

Tabel 2. Transformation Batas Interval untuk atribut *pdays*

Batas Interval	Batas Bawah	Batas Atas	Transformasi
1	0	58,76	1
2	58.77	117.53	2
3	117.54	176.3	3
4	176.4	253.06	4
5	253.07	293.83	5
6	293.84	352.6	6
7	352.7	411.46	7
8	411.47	470.23	8
9	470.24	529	9
10	530	588.76	10
11	588.77	647.53	11
12	647.54	706.3	12
13	706.4	765.16	13
14	765.17	823.93	14
15	823.94	882.7	15
16	882.8	941.56	16
17	941.57		17

Tabel 3. Transformation Batas Interval untuk atribut *duration*

Batas Interval	Batas Bawah	Batas Atas	Transformasi
1	0	289,29	1
2	289,30	578,59	2
3	578.60	867.88	3
4	867,89	1,158	4
5	1.159	1.447	5
6	1.448	1.736	6
7	1.737	2.205	7
8	2.206	2.314	8
9	2.315	2.603	9
10	2.604	2.892	10
11	2.893	3.181	11
12	3.182	3.469	12
13	3.470	3.759	13
14	3.760	4.048	14
15	4.049	4.337	15
16	4.338	4.626	16
17	4.627	4.915	17

Setelah itu, dihitung *normalisasi* tiap data pada kolom *Duration* dan *Pdays*. Berikut hasil perhitungan data berdasarkan rumus (4) :

Normalisasi data atribut *Duration*

$$Ndata = \frac{(1-0)*(1-0)}{14-0} + 0 = 0.07142$$

Normalisasi data atribut *Pdays*

$$Ndata = \frac{(1-0)*(1-0)}{17-0} + 0 = 0.058824$$

Data kategorial ditransformasi ke 0, 1, 2, 3, dan seterusnya sedangkan nilai luaran akan bernilai 1 atau 0. Berikut representasi luaran dari sistem seperti yang terlihat pada tabel 4.

Tabel 4. Nilai Target Output

Status	Nilai Target Output
Yes	1
No	0

3-fold cross validation

Evaluasi sistem pada penelitian ini dilakukan dengan membagi data menjadi *data testing* dan *data training*. Pembagian data dilakukan dengan menerapkan model *3-fold cross validation* dengan membagi data menjadi 3 bagian seperti data Tabel 5.

Tabel 5. Model 3-fold cross validation

Model	Data Training	Data Testing
1	1,2	3
2	1,3	2
3	2,3	1

Pemodelan Algoritme Naïve Bayes

Naive Bayes merupakan sebuah metode untuk klasifikasi yang diciptakan oleh Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [8]. *Naive bayes* merupakan metode yang membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [9].

Persamaan teorema *Bayes* adalah :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad \dots\dots\dots(8)$$

Keterangan :

- X : Data dengan *class* yang belum diketahui.
- H : Hipotesis data merupakan suatu *class* spesifik.
- $P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)
- $P(X|H)$: Probabilitas berdasarkan kondisi pada hipotesis.
- $P(H)$: Probabilitas hipotesis H (probabilitas prior).
- $P(X)$: Probabilitas X.

Proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Oleh karena itu, Metode *Naive Bayes* diatas disesuaikan sebagai berikut :

$$P(C|F1...Fn) = \frac{P(C)P(F1...Fn|C)}{P(F1...Fn)} \quad \dots\dots\dots(9)$$

Variabel C mempresentasikan kelas, sementara variabel $F1 \dots Fn$ mempresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus 2.2 tersebut menjelaskan bahwa peluang terjadinya kelas C (*posterior probability*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (*likelihood*) dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*).

Oleh karena itu rumus (9) di atas dapat ditulis secara sederhana sebagai berikut :

$$posteriority = likelihood \times \frac{prior}{evidence} \quad \dots\dots\dots(10)$$

Apabila data yang digunakan adalah data kontinu maka perhitungan klasifikasi menggunakan rumus *Densitas Gauss* :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \quad \dots\dots\dots(11)$$

Keterangan :

P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut i

Y : Kelas yang dicari

y_i : Sub kelas Y yang dicari

μ : Mean, menyatakan rata – rata seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut

Nilai Likelihood diperoleh dengan cara mengalikan peluang atribut x_i dengan nilai probabilitas kategori seperti pada rumus (11)

$$P(X_1) \times P(X_2) \times P(X_3) \dots P(X_n) \times \text{Nilai Probabilitas kategori} \dots \dots \dots (12)$$

Berikut ini adalah penyelesaian contoh kasus menggunakan algoritme Naive Bayes.

Terdapat dua kelas dari klasifikasi yang terbentuk :

1 : Yes

2 : No

Langkah-langkah perhitungan berikut ini :

- a. Menghitung mean dan standar deviasi untuk setiap data atribut.

Menghitung Mean atau nilai rata-rata menggunakan rumus sebagai berikut:

$$\text{Nilai rata – rata} = (\text{jumlah nilai})/(\text{Banyak data}) \dots \dots \dots (13)$$

Sementara untuk menghitung nilai standar deviasi setiap atribut dengan rumus :

$$S = \sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2)/(n - 1))} \dots \dots \dots (14)$$

- b. Menghitung probabilitas kategori kelas untuk atribut duration. Tabel berikut menunjukkan probabilitas atribut duration pada kategori kelas.
c. Menghitung probabilitas untuk setiap kategori pada kelas.
d. Menghitung probabilitas setiap kategori kelas dengan acuan nilai mean dan standar deviasi untuk setiap

$$\text{atribut sesuai rumus} . P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \dots \dots \dots (15)$$

- e. Menghitung likelihood ini digunakan hasil dari nilai perhitungan probabilitas tiap atribut. Nilai likelihood dibagi menjadi 2 yaitu likelihood Ya dan likelihood No.
f. Membandingkan nilai probabilitas setiap kelas. Dari probabilitas di atas, masing-masing nilai akan dibandingkan untuk dicari nilai terbesarnya. Jika salah satu label memiliki nilai terbesar maka label tersebut merupakan hasil klasifikasi untuk data testing yang diuji.

Akurasi dengan Matritks Confusion.

Matriks confusion merupakan tabel yang mencatat hasil kerja klasifikasi. Contoh matriks confusion sebagai berikut :

Tabel 6. *Confution Matrix*

	Positif	Negatif
Positif	TP(True positif)	FN(False Negatif)
Negatif	False(False Positif)	TN(True Negatif)

Nilai Akurasi dapat dihitung dengan persamaan :

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN) \times 100\% \dots \dots \dots (16)$$

Dimana :

TP : Jumlah positif yang diklasifikasi sebagai positif

TN : Jumlah negatif yang diklasifikasi sebagai negatif

FP : Jumlah negatif yang diklasifikasi sebagai positif

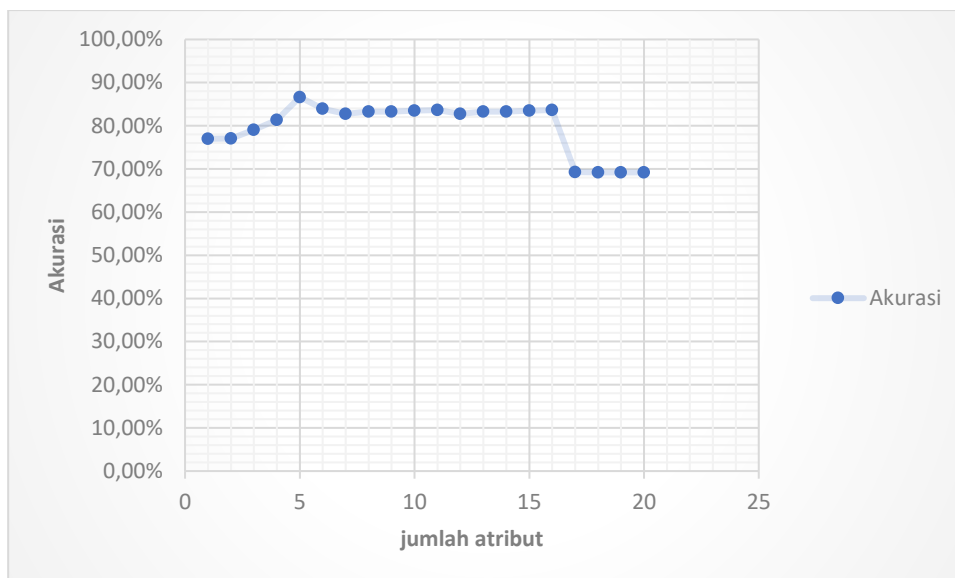
FN : Jumlah positif yang diklasifikasi sebagai negatif.

3. HASIL DAN PEMBAHASAN

Hasil pengujian akurasi yang dilakukan dengan menghitung tingkat keakuratan hasil perhitungan sistem dengan menggunakan Algoritme *Naive Bayes*, menggunakan 18553 data nasabah yang berpotensi membuka simpanan deposito diperoleh hasil pada tabel 4.1. berikut ini:

Tabel 7. Hasil Uji Atribut dan Akurasi

Attribut	Jumlah Attribut	Akurasi
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default, Campaign, Education, Marital, Dat_Of_Week, Housing, Loan</i>	20	69,1939%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default, Campaign, Education, Marital, Dat_Of_Week, Housing</i>	19	69,21%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default, Campaign, Education, Marital, Dat_Of_Week</i>	18	69,2047%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default, Campaign, Education, Marital</i>	17	69,2532%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default, Campaign, Education</i>	16	83,6405%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default, Campaign</i>	15	83,5165%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default</i>	14	83,2525%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job</i>	13	83,3064%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact</i>	12	82,7352%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx, Age, Contact, Job, Default, Campaign, Education</i>	11	83,6405%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous, Cons.Price.Idx,</i>	10	83,5165%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month, Previous</i>	9	83,2525%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome, Month</i>	8	83,3064%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays, Poutcome</i>	7	82,7352%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m, Pdays</i>	6	83,9099%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m</i>	5	86,6419%
<i>Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate</i>	4	81,3504%
<i>Duration, Cons.Conf.Idx, Nr.employed</i>	3	79,0441%
<i>Duration, Cons.Conf.Idx</i>	2	77,018%
<i>Duration</i>	1	76,9695%



Gambar 2. Akurasi vs Jumlah atribut

Berdasarkan gambar 1. diperoleh hasil bahwa dari 20 atribut yang digunakan ternyata apabila dipilih atribut dengan 5 nilai peringkat tertinggi diperoleh atribut-atribut Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m. Sedangkan berdasarkan gambar 2 dapat dilihat bahwa jumlah atribut optimal adalah 5 (lima), dengan demikian didapatkan akurasi sebesar 86,6419%. Penggunaan atribut lebih dari 5 tidak menghasilkan klasifikasi yang optimal. Adapun 5 atribut yang dipakai adalah Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m. Hal tersebut menunjukkan bahwa untuk klasifikasi data nasabah bank yang berpotensi membuka deposito cukup mempertimbangkan 5 atribut di atas yaitu: Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m

4. SIMPULAN

Pengujian yang dilakukan terhadap 18559 data menggunakan 3-fold cross validation pada klasifikasi data nasabah bank yang berpotensi membuka deposito menghasilkan tingkat keakuratan sebesar 86,6419% . Selain itu atribut yang berpengaruh dalam proses klasifikasi data nasabah bank yang berpotensi membuka deposito menggunakan algoritme naïve bayes adalah atribut-atribut: Duration, Cons.Conf.Idx, Nr.employed, Emp.Var.Rate, Euribor3m.

5. SARAN

Saran-saran untuk penelitian lebih lanjut antara lain: Sistem ini dapat juga dikembangkan dengan menggunakan metode yang lain misalnya correlated naïve bayes, *backpropagation* dan sistem dapat menerima masukan file tidak hanya bertipe file xlsx.

DAFTAR PUSTAKA

- [1] Febriani, F. 2019. Prediksi Nasabah Yang Berpotensi Membuka Simpanan Deposito Menggunakan Metode Decision Tree Dengan Penerapan Algoritme C4.5. Yogyakarta: Universitas Sanata Dharma.
- [2] Putri, R. M. Y. 2019. Klasifikasi Data Nasabah Berpotensi Terkena Kredit Macet Dengan Menggunakan Metode Naive Bayes. Yogyakarta: Universitas Sanata Dharma.
- [3] Rizqi I D, Wardhani R, Zamroni M H.. 2024. Implementasi Algoritma Naive Bayes pada Sistem Pendukung Keputusan Rekomendasi Wisata di Lamongan. PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS TAHUN 2024, Vol. 3. e-ISSN: 2828–299X , <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/4078/2995> diakses pada tanggal 30 Desember 2024
- [4] Prawestina E M , Rohman M G, Zamroni M R. 2024. Sistem Pakar Diagnosa Gangguan Pencernaan Balita Dengan Metode Naive Bayes. PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS TAHUN 2024, Vol. 3. e-ISSN: 2828–299X. <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/4079/2996> diakses pada tanggal 30 Desember 2024
- [5] Rahayu R. Algoritma Naïve Bayes. https://www.researchgate.net/publication/376713713_Algoritma_Naive_Bayes di akses pada tanggal 30 Desember 2024.

- [6] Han, J, Kamber M, Pei J . 2012 Data mining : Concepts and Techniques 3rd Edition. San Fransisco : Morgan Kaufmann Publishers.
- [7] Winanda, O.I., Zega, S.A. and Hilmawan, R., 2019. Prediksi Rating Film Animasi Berdasarkan Elemen Mise En Scene Menggunakan Neural Network. JOURNAL OF APPLIED MULTIMEDIA AND NETWORKING, 3(1), pp.15-26.
- [8] Bustami., 2013, *Penerapan Algoritme Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, TECHSI* : Jurnal Penelitian Teknik Informatika, Vol. 3, No.2, Hal. 127-146.
- [9] Pattekari, S. A., Parveen, A., 2012, Prediction System for Heart Disease Using Naive Bayes, International Journal of Advanced Computer and Mathematical Sciences, ISSN 2230-9624, Vol. 3, No 3, Hal 290-294.