

KECERDASAN BUATAN



DOSEN PENGAMPU : PROF.DR.IR. SITI NURMAINI, M.T

**KLASIFIKASI MARKETING BANK
DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES**

NAMA : ANISA WULANDARI

NIM : 09011282025051

KELAS : SK5A INDRALAYA

**PROGRAM STUDI SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
INDRALAYA**

2022

BAB 1

PENDAHULUAN

A. LATAR BELAKANG

Bank merupakan perusahaan yang memiliki data yang besar yang tersimpan di dalam database dan diolah menghasilkan sebuah informasi yang saling berkaitan tentang nasabah, data tersebut serta dapat digunakan untuk menjaga hubungan antar bank dengan nasabah yang valid, sehingga berguna untuk menentukan secara individual tentang penawaran produk bank .

Masalah utama yang dihadapi dalam bank marketing ini merupakan bagaimana untuk mencapai sebuah akurasi yang tinggi dalam proses klasifikasi berdasarkan kecepatan informasi tertentu yang diperoleh dari nasabah dan dianggap penting oleh pihak bank . Hal ini mempengaruhi pada pihak bank marketing dalam menentukan pengambilan keputusan, karena akurasi yang tinggi pada proses klasifikasi. Pihak bank harus memiliki cara bagaimana untuk mengetahui kendala pada nasabah bank marketing yang ingin melakukan deposito di bank. Target utama pada sebuah perusahaan bank marketing yaitu mencoba memprediksi harapan terhadap nasabah yang memiliki kemungkinan tertinggi dalam pelayanan dengan menggunakan teknik data mining . Pada bank harus selektif dalam memilih nasabah yang menerima pinjaman kredit .

Perusahaan Bank, harus memiliki ide dan terobosan baru guna mengetahui kendala pada nasabah marketing yang ingin melakukan deposito pada bank tersebut, yang bertujuan agar bank terhindar dari ancaman krisis keuangan. Dengan kata lain Bank harus berupaya meningkatkan modal perbankan, salah satunya dengan cara menawarkan produk, yaitu deposito baik jangka pendek ataupun jangka panjang.

Bank marketing harus memiliki dan membuat sebuah rencana untuk target nasabah mana yang berpotensi, agar dapat ditawarkan dan melakukan deposito dengan menganalisis data klien bank yang ada . Data mining adalah sebuah proses yang digunakan untuk memanipulasi data dengan mengekstraksi informasi yang sebelumnya tidak diketahui dari dataset yang berukuran besar . Saat ini, peran penggunaan data mining sering digunakan pada beberapa industri termasuk asuransi dan perbankan. Pada penggunaan teknik data mining dalam bank marketing bertujuan untuk menganalisa data pelanggan dan mengembangkan data pelanggan secara statistika berdasarkan produk dan pelayanan yang lebih diminati oleh pelanggan. Dalam menyelesaikan permasalahan klasifikasi, pada penggunaan metode atau teknik bertujuan untuk

mempermudah pada proses klasifikasi. Dengan semakin pesat perkembangan ilmu data mining banyak penelitian yang digunakan untuk memprediksi suatu kasus khususnya perbankan, seperti klasifikasi persetujuan kredit , nasabah kredit potensial dan nasabah deposito sesuai target.

B. Rumusan Masalah

1. Bagaimana memprediksi apakah seseorang akan melakukan deposit kedalam bank atau tidak?
2. bagaimana untuk mencapai sebuah akurasi yang tinggi dalam proses klasifikasi berdasarkan informasi tertentu yang diperoleh dari nasabah dan dianggap penting oleh pihak bank?

C. Batasan Masalah

1. Klasifikasi atau project ini hanya dilakukan untuk memprediksi apakah seseorang akan melakukan deposit kedalam bank atau tidak.
2. Metode yang digunakan dalam project ini adalah metode naive bayes

D. Tujuan

Gambaran atau tujuan pada klasifikasi ini adalah untuk menentukan atau memprediksi apakah seseorang akan melakukan deposit ke dalam bank, analisis kumpulan data pemasaran (marketing) bank yang berisi data atau informasi tentang nasabah dan bertujuan untuk mendapatkan wawasan yang berguna dari data tersebut lalu memprediksi apakah nasabah baru akan menerima tawaran deposit atau tidak.

BAB 2

DASAR TEORI

A. Metode Naïve Bayes

Naïve Bayes classifier merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik, Naive Bayes adalah metode yang cocok untuk klasifikasi biner dan multiclass yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Algoritma Naïve Bayes merupakan Algoritma data mining yang relatif sederhana yang memiliki kelebihan tingkat akurasi yang tinggi dan dapat menangani data dalam jumlah besar. Oleh karena itu, saya akan menerapkan metode Naïve Bayes Classifier dalam mengklasifikasi marketing bank. Ciri utama dari metode ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian.

B. Machine Learning

Teknologi Machine Learning adalah mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunanya. Pembelajaran mesin dikembangkan berdasarkan disiplin ilmu lainnya seperti statistika, matematika dan data mining sehingga mesin dapat belajar dengan menganalisa data tanpa perlu di program ulang atau diperintah.

Dalam hal ini machine learning memiliki kemampuan untuk memperoleh data yang ada dengan perintah ia sendiri. Machine Learning juga dapat mempelajari data yang ada dan data yang ia peroleh sehingga bisa melakukan tugas tertentu. Tugas yang dapat dilakukan oleh Machine Learning pun sangat beragam, tergantung dari apa yang ia pelajari.

Istilah machine learning pertama kali dikemukakan oleh beberapa ilmuwan matematika seperti Adrien Marie Legendre, Thomas Bayes dan Andrey Markov pada tahun 1920-an dengan mengemukakan dasar-dasar machine learning dan konsepnya. Sejak saat itu Machine Learning banyak yang mengembangkan. Salah satu contoh dari penerapan Machine Learning yang cukup terkenal adalah Deep Blue yang dibuat oleh IBM pada tahun 1996.

BAB 3

PROGRES PROJECT

Langkah-langkah Melakukan Project

A. Memasukan Dataset

Sebelum itu saya akan menjelaskan bahwa saya mengerjakan project ini menggunakan jupyter notebook dan menggunakan bahasa pemrograman python.

Langkah pertama yang kita lakukan adalah memasukan file data set yang saya dapatkan dari <https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset> , dataset ini adalah sumber data yang akan diolah.

B. Import Libraries

Mengimport libraries yang akan digunakan dalam project kali ini seperti numpy,pandas,matplotlib,seaborn dan tensorflow seperti pada gambar dibawah ini

Import Library

Berikut adalah library yang diperlukan selama penyelesaian UTS ini.

```
[ ] # Import Library
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import tensorflow as tf
```

C. Import Dataset

- Membaca Dataset

Import Dataset yang akan digunakan, dataset ini adalah sumber data yang akan diolah jika dijalankan akan menampilkan 5teratas dari keseluruhan isi dataset.

```
# Import Data
df_bm = pd.read_csv('https://raw.githubusercontent.com/hafizhfajar/Klasifikasi_Marketing_Bank_DS0119/master/Dataset/bank.csv')

# Show 5 first Row
df_bm.head()
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0	unknown	yes
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0	unknown	yes
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0	unknown	yes
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0	unknown	yes
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0	unknown	yes

- Menampilkan Informasi Data

Berdasarkan Informasi umum data terdapat 11162 data dengan 2 tipe data yaitu numerik (7 atribut) dan objek (10 atribut).

- Tipe Data Numerik: age, balance, day, duration, campaign, pdays, previous
- Tipe Data objek: job, marital, education, default, housing, loan, contact, month, poutcome, deposit.

```
[ ] # Prints the Dataset Information
df_bm.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11162 entries, 0 to 11161
Data columns (total 17 columns):
#   Column      Non-Null Count  Dtype
---  ---
0    age         11162 non-null   int64
1    job         11162 non-null   object
2    marital     11162 non-null   object
3    education   11162 non-null   object
4    default     11162 non-null   object
5    balance     11162 non-null   int64
6    housing     11162 non-null   object
7    loan        11162 non-null   object
8    contact     11162 non-null   object
9    day         11162 non-null   int64
10   month       11162 non-null   object
11   duration    11162 non-null   int64
12   campaign    11162 non-null   int64
13   pdays       11162 non-null   int64
14   previous    11162 non-null   int64
15   poutcome    11162 non-null   object
16   deposit     11162 non-null   object
dtypes: int64(7), object(10)
memory usage: 1.4+ MB
```

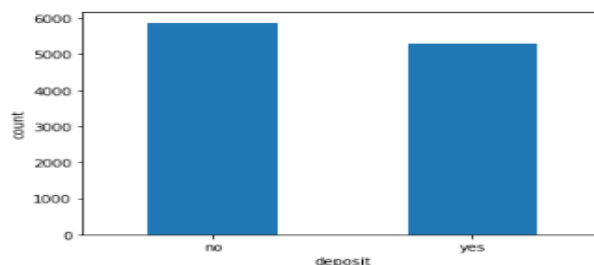
- Jumlah Kelas Deposit

Pada langkah ini saya melakukan visualisasi jumlah data pada setiap kelas untuk atribut deposit, dan jumlah nasabah bank yang melakukan deposit sebanyak 5289 dan yang tidak melakukan deposit 5873.

```
[ ] # Menampilkan jumlah kelas pada atribut deposit
cls = df_bm['deposit'].value_counts()
print(cls)

no      5873
yes     5289
Name: deposit, dtype: int64

[ ] df_bm['deposit'].value_counts().plot(kind='bar', rot=0);
plt.xlabel("deposit")
plt.ylabel("count");
```



D.Preprocessing Data

Tujuan dari tahapan ini adalah untuk membuat data menjadi ideal sebelum di proses.

- Mengecek Data Kosong

Mengecek apakah terdapat data yang kosong atau tidak. Jika terdapat data kosong maka data tersebut akan diisi dengan median berdasarkan atributnya. Dikarenakan pada data yang saya gunakan tidak terdapat data kosong maka tidak perlu dihitung median nya.

```
[ ] # Check for Missing Values
df_bm.isnull().sum()

age      0
job      0
marital  0
education 0
default  0
balance  0
housing  0
loan     0
contact  0
day      0
month    0
duration 0
campaign 0
pdays   0
previous 0
poutcome 0
deposit  0
dtype: int64
```

- Encode Categorical Data

Seperti yang telah diketahui bahwa terdapat dua tipe data yaitu numerik dan objek.

Fungsi pembelajaran mesin dalam sklearn mengharapkan atribut yang digunakan adalah atribut dengan tipe data numerik. Untuk mengonversinya, perlu menggunakan encoding yang mengacu pada proses transformasi tipe data objek menjadi bentuk numerik.

```
# Encoder
encoder = OneHotEncoder(sparse=False)

# Encode Categorical Data
df_encoded = pd.DataFrame(encoder.fit_transform(df_bm[['job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', 'poutcome']]))
df_encoded.columns = encoder.get_feature_names(['job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', 'poutcome'])
df_bm['deposit'].replace('no',0, inplace=True)
df_bm['deposit'].replace('yes',1, inplace=True)

# Replace Categorical Data with Encoded Data
df_bm.drop(['job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', 'poutcome'],axis=1, inplace=True)
df_encoded = pd.concat([df_bm, df_encoded], axis=1)

# Show Encoded Dataframe
df_encoded
```

/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names is warnings.warn(msg, category=FutureWarning)

	age	balance	day	duration	campaign	pdays	previous	deposit	job_admin.	job_blue-collar	...	month_jun	month_mar	month_may	month_nov	mo
0	59	2343	5	1042	1	-1	0	1	1.0	0.0	...	0.0	0.0	1.0	0.0	
1	56	45	5	1467	1	-1	0	1	1.0	0.0	...	0.0	0.0	1.0	0.0	
2	41	1270	5	1389	1	-1	0	1	0.0	0.0	...	0.0	0.0	1.0	0.0	
3	55	2476	5	579	1	-1	0	1	0.0	0.0	...	0.0	0.0	1.0	0.0	
4	54	184	5	673	2	-1	0	1	1.0	0.0	...	0.0	0.0	1.0	0.0	
...
11157	33	1	20	257	1	-1	0	0	0.0	1.0	...	0.0	0.0	0.0	0.0	
11158	39	733	16	83	4	-1	0	0	0.0	0.0	...	1.0	0.0	0.0	0.0	
11159	32	29	19	156	2	-1	0	0	0.0	0.0	...	0.0	0.0	0.0	0.0	
11160	43	0	8	9	2	172	5	0	0.0	0.0	...	0.0	0.0	1.0	0.0	
11161	34	0	9	628	1	-1	0	0	0.0	0.0	...	0.0	0.0	0.0	0.0	

- Set Feature and Target

Langkah ini untuk menentukan atribut-atribut yang digunakan sebagai feature dan sebagai target. Dimana feature pada data kali ini adalah semua data kecuali deposit dan targetnya adalah deposit.

```
[ ] # Select Features
feature = df_encoded.drop(['deposit'], axis=1)
feature
```

```
[ ] # Select Target
target = df_encoded['deposit']
target
```

- Set Training and Testing Data

Langkah ini digunakan untuk membagi data ke dalam data train(x) dan data test(y) dengan perbandingan 70:30 sebelum dilakukan klasifikasi

```
[ ] # Set Training and Testing Data (70:30)
X_train, X_test, y_train, y_test = train_test_split(feature, target, shuffle = True, test_size=0.3, random_state=100)

# Show the Training and Testing Data
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)

(7813, 51)
(3349, 51)
(7813,)
(3349,)
```

```
[ ] # Print data test
X_test
```

	age	balance	day	duration	campaign	pdays	previous	job_admin.	job_blue-collar	job_entrepreneur	...	month_jun	mont
1402	50	414	3	707	1	302	3	0.0	0.0	0.0	...	0.0	
9326	25	-209	13	56	2	-1	0	0.0	1.0	0.0	...	0.0	
3380	84	81204	1	390	1	94	3	0.0	0.0	0.0	...	0.0	
7413	45	999	16	656	2	-1	0	0.0	0.0	0.0	...	0.0	
10309	30	19358	19	258	2	-1	0	0.0	0.0	0.0	...	0.0	

```
[ ] # Print label test
y_test
```

```
1402    1
9326    0
3380    1
7413    0
10309    0
..
9318    0
10290    0
6965    0
666     1
5235    1
Name: deposit, Length: 3349, dtype: int64
```


E. Sumber Referensi

<https://journal.universitasbumigora.ac.id/index.php/matrik/article/view/826>

<https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset>