Nama : Triansyah Amarullah Ahmad Prayoga

NPM : 41155050210034

Kelas : TIF-A2 2021

Matkul : Machine Learning

TUGAS PERTEMUAN 4

1.0. Lakukan praktik dari https://youtu.be/Sj1ybuDDf9I?si=hCajHe1zasTQ9HGY , buat screenshot dengan nama kalian pada coding, kumpulkan dalam bentuk pdf, dari kegiatan ini:

1.1. Pengenalan Bayes Theorem | Teori Bayes | Conditional Probability

Bayes' Theorem

Bayes' theorem menawarkan suatu formula untuk menghitung nilai probability dari suatu event dengan memanfaatkan pengetahuan sebelumnya dari kondisi terkait; atau sering kali dikenal dengan istilah conditional probability.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

$$P(y|X) = \frac{P(X|y) \times P(y)}{P(X)}$$

$$Posterior = \frac{Likelihood \times Prior}{Evidence}$$

1.2. Pengenalan Naive Bayes Classification

Bayes' Theorem

Bayes' theorem menawarkan suatu formula untuk menghitung nilai probability dari suatu event dengan memanfaatkan pengetahuan sebelumnya dari kondisi terkait; atau sering kali dikenal dengan istilah conditional probability.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

$$P(y|X) = \frac{P(X|y) \times P(y)}{P(X)}$$

$$Posterior = \frac{Likelihood \times Prior}{Evidence}$$

1.3. Pengenalan Prior Probability

Asep	Joko
+ siomay:0.1 + bakso:0.8 + lumpia: 0.1	+ siomay: 0.5 + bakso: 0.2 + lumpia: 0.3

Misi: Lakukan prediksi siapa pelanggan yang melakukan pemesanan dengan diketahui pesanannya adalah **lumpia** dan **bakso**.

Prior Probability: $P(y)$

- **Referensi:** https://en.wikipedia.org/wiki/Prior_probability
- $P(Asep) = 0.5$
- $P(Joko) = 0.5$

1.4. Pengenalan Likelihood

Likelihood: $P(X|y)$

- **Referensi:** https://en.wikipedia.org/wiki/Likelihood_function
- Asep:

$$\begin{aligned} P(lumpia, bakso|Asep) &= (0.1 \times 0.8) \\ &= 0.08 \end{aligned}$$

- Joko:

$$\begin{aligned} P(lumpia, bakso|Joko) &= (0.3 \times 0.2) \\ &= 0.06 \end{aligned}$$

1.5. Pengenalan Evidence | Normalizer

Evidence atau Normalizer: $P(X)$

$$\begin{aligned} Evidence &= \sum (Likelihood \times Prior) \\ P(lumpia, bakso) &= (0.08 \times 0.5) + (0.06 \times 0.5) \\ &= 0.07 \end{aligned}$$

1.6. Pengenalan Posterior Probability

Posterior Probability: $P(y|X)$

- Referensi: https://en.wikipedia.org/wiki/Posterior_probability
- Formula:
  $Posterior = \frac{Likelihood \times Prior}{Evidence}$
- Asep:
$$P(Asep|lumpia, bakso) = \frac{0.08 \times 0.5}{0.07} = 0.57$$
- Joko:
$$P(Joko|lumpia, bakso) = \frac{0.06 \times 0.5}{0.07} = 0.43$$

1.7. Studi kasus dan implementasi Naive Bayes

Studi Kasus 2

Asep	Joko
+ siomay:0.1	+ siomay: 0.5
+ bakso:0.8	+ bakso: 0.2
+ lumpia: 0.1	+ lumpia: 0.3

Misi: Lakukan prediksi siapa pelanggan yang melakukan pemesanan dengan diketahui pesanannya adalah **siomay** dan **bakso**.

Posterior Probability: $P(y|X)$ ***(kasus 2)***

- pesanan: siomay, bakso

Posterior Probability: $P(y|X)$ ***(kasus 2)***

- pesanan: siomay, bakso
- Evidence: $P(X)$
$$P(siomay, bakso) = (0.1 \times 0.8 \times 0.5) + (0.5 \times 0.2 \times 0.5) = 0.09$$
- Asep:
$$P(Asep|siomay, bakso) = \frac{(0.1 \times 0.8) \times 0.5}{0.09} = 0.444$$
- Joko:
$$P(Joko|siomay, bakso) = \frac{(0.5 \times 0.2) \times 0.5}{0.09}$$

Mengapa disebut Naive?

- Karena sewaktu kita mendefinisikan Likelihood $P(lumpia, bakso|Asep)$,
- kita mengasumsikan $P(lumpia|Asep)$ conditionally independent terhadap $P(bakso|Asep)$; demikian sebaliknya.
- Sehingga dapat diformulasikan sebagai berikut:

$$P(lumpia, bakso|Asep) = P(lumpia|Asep) \times P(bakso|Asep)$$

Persiapan Dataset / Wisconsin Breast Cancer Dataset

```
[2]: from sklearn.datasets import load_breast_cancer

print("Triansyah Amarullah Ahmad Prayoga" , "41155050210034")

print(load_breast_cancer().DESCR)
```

```
Triansyah Amarullah Ahmad Prayoga 41155050210034
.. _breast_cancer_dataset:

Breast cancer wisconsin (diagnostic) dataset
--------------------------------------------

**Data Set Characteristics:**

    :Number of Instances: 569

    :Number of Attributes: 30 numeric, predictive attributes and the class

    :Attribute Information:
        - radius (mean of distances from center to points on the perimeter)
        - texture (standard deviation of gray-scale values)
        - perimeter
        - area
        - smoothness (local variation in radius lengths)
```

```
    The mean, standard error, and "worst" or largest (mean of the three
    worst/largest values) of these features were computed for each image,
    resulting in 30 features.  For instance, field 0 is Mean Radius, field
    10 is Radius SE, field 20 is Worst Radius.

    - class:
            - WDBC-Malignant
            - WDBC-Benign

:Summary Statistics:

===================================== ====== ======
                                       Min    Max
===================================== ====== ======
radius (mean):                        6.981  28.11
texture (mean):                       9.71   39.28
perimeter (mean):                     43.79  188.5
area (mean):                          143.5  2501.0
```

```
texture (standard error):            0.36   4.885
perimeter (standard error):          0.757  21.98
area (standard error):               6.802  542.2
smoothness (standard error):         0.002  0.031
compactness (standard error):        0.002  0.135
concavity (standard error):          0.0    0.396
concave points (standard error):     0.0    0.053
symmetry (standard error):           0.008  0.079
fractal dimension (standard error):  0.001  0.03
radius (worst):                      7.93   36.04
texture (worst):                     12.02  49.54
perimeter (worst):                   50.41  251.2
area (worst):                        185.2  4254.0
smoothness (worst):                  0.071  0.223
compactness (worst):                 0.027  1.058
concavity (worst):                   0.0    1.252
concave points (worst):              0.0    0.291
symmetry (worst):                    0.156  0.664
```

```
    symmetry (worst):                    0.156  0.664
    fractal dimension (worst):           0.055  0.208
    ===================================== ====== ======

    :Missing Attribute Values: None

    :Class Distribution: 212 - Malignant, 357 - Benign

    :Creator:  Dr. William H. Wolberg, W. Nick Street, Olvi L. Mangasarian

    :Donor: Nick Street

    :Date: November, 1995

This is a copy of UCI ML Breast Cancer Wisconsin (Diagnostic) datasets.
https://goo.gl/U2Uwz2

Features are computed from a digitized image of a fine needle
```

```
ftp ftp.cs.wisc.edu
cd math-prog/cpo-dataset/machine-learn/WDBC/

|details-start|
**References**
|details-split|

- W.N. Street, W.H. Wolberg and O.L. Mangasarian. Nuclear feature extraction
  for breast tumor diagnosis. IS&T/SPIE 1993 International Symposium on
  Electronic Imaging: Science and Technology, volume 1905, pages 861-870,
  San Jose, CA, 1993.
- O.L. Mangasarian, W.N. Street and W.H. Wolberg. Breast cancer diagnosis and
  prognosis via linear programming. Operations Research, 43(4), pages 570-577,
  July-August 1995.
- W.H. Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning techniques
  to diagnose breast cancer from fine-needle aspirates. Cancer Letters 77 (1994)
  163-171.
```

Training & Testing Set

```
•[4]: from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test =train_test_split(X,
                                                    y,
                                                    test_size=0.2,
                                                    random_state=0)

print(f'X_train shape: {X_train.shape}')
print(f'X_test shape: {X_test.shape}')
```

```
X_train shape (455, 30)
X_test shape (114, 30)
```

Implementasi Naive Bayes Classification dengan Scikit-Learn

```
[ ]: from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score


model = GaussianNB()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
accuracy_score(y_test, y_pred)
```

```
0.9298245614035088
```

```
model.score(X_test, y_test)
```

```
0.9298245614035088
```

Tidak ada tugas untuk Confusion Matrix.