

## TUGAS PERTEMUAN 3

NAMA : MAYA NURROHMAH

NPM : 41155050210019

KELAS : INF\_A1

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

- 1) Pengenalan Bayes Theorem | Teori Bayes | Conditional Probability  
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

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

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

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

- 2) Pengenalan Naive Bayes Classification

*Studi Kasus 1*

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

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

- 3) Pengenalan Prior Probability

Prior Probability:  $P(y)$  adalah nilai probability dari kemunculan suatu nilai target label tertentu tanpa memperhatikan nilai fiturnya.

Jadi untuk kasus ini kita bisa asumsikan sebagai berikut :

- $P(\text{Asep}) = 0,5$
- $P(\text{Joko}) = 0,5$

- 4) Pengenalan Likelihood

Terminologi Likelihood;  $P(X/y)$  Likelihood untuk konteks Naive Bayes bisa didefinisikan sebagai probability kemunculan nilai fitur tertentu bila diketahui kemunculan nilai target labelnya

- Asep  
 $P(\text{lumpia, bakso}|\text{Asep}) = (0.1 \times 0.8)$   
 $= 0.08$

- Joko  

$$P(\text{lumpia, bakso}|\text{Joko}) = (0.3 \times 0.2)$$

$$= 0.06$$

##### 5) Pengenalan Evidence | Normalizer

Terminologi Evidence atau Normalizer:  $P(X)$  Evidence untuk konteks Naive Bayes bisa didefinisikan sebagai total akumulasi dari hasil perkalian antara nilai Likelihood dengan Prior

$$Evidence = \sum (Likelihood \times Prior)$$

$$P(\text{lumpia, bakso}) = (0.08 \times 0.5) + (0.06 \times 0.5)$$

$$= 0.07$$

##### 6) Pengenalan Posterior Probability

- Formula

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

Disini Kita memiliki 2 class maka posterior nya bisa dieskpresikan sebagai berikut :

- Asep:

$$P(\text{Asep}|\text{lumpia, bakso}) = \frac{0.08 \times 0.5}{0.07}$$

$$= 0.57$$

- Joko:

$$P(\text{Joko}|\text{lumpia, bakso}) = \frac{0.06 \times 0.5}{0.07}$$

$$= 0.43$$

Seperti yang kita lihat untuk kasus kita disini nilai posterior untuk asef lebih tinggi dibandingkan nilai posterior joko, maka naive bayes akan mengklasifikasikan asef sebagai pemesannya.

##### 7) Studi kasus Naïve Bayes Classification

Asef	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 pesannya adalah siomay dan bakso.

Jawaban :

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

- pesanan: siomay, bakso

- Evidence:  $P(X)$

$$P(\text{siomay, bakso}) = (0.1 \times 0.8 \times 0.5) + (0.5 \times 0.2 \times 0.5)$$

$$= 0.09$$

- Asef:

$$P(\text{Asef}|\text{siomay, bakso}) = \frac{(0.1 \times 0.8) \times 0.5}{0.09}$$

$$= 0.444$$

- Joko:

$$P(\text{Joko}|\text{siomay, bakso}) = \frac{(0.5 \times 0.2) \times 0.5}{0.09}$$

## 8) Penjelasan Algoritma yang Naïve

Mengapa model ini diberi nama Naïve Bayes?

kata Bayes berasal dari nama seorang station yaitu Thomas Bayes, beliau adalah yang menggagas suatu theorem, yang akhirnya dikenal sebagai Bayes theorem

Mengapa disebut Naïve?

- 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:

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

## 9) Persiapan Dataset | Wisconsin Breast Cancer | UCI Machine Learning

- Load Dataset

```
[2]: from sklearn.datasets import load_breast_cancer
    print(load_breast_cancer().DESCR)

... _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)
  - compactness (perimeter^2 / area - 1.0)

[6]: # load_breast_cancer?
    X, y = load_breast_cancer(return_X_y=True)
    X.shape

[6]: (569, 30)

[4]: nama = "Maya Nurrohmah"
    print(f>Nama : {nama}")

Nama : Maya Nurrohmah
```

- Training & Testing Set

```
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)

nama = "Maya Nurrohmah"
print(f>Nama : {nama}")

Nama : Maya Nurrohmah
```

## 10) Implementasi Naïve 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

: nama = "Maya Nurrohmah"
: print(f>Nama : {nama}")

Nama : Maya Nurrohmah
```