

Klasifikasi Data "Drug" Metode SVM



course machine learning
Tri Hesti Wahyuningsih



EXPLANATION



Support Vektor Machine

metode Support Vektor Machine merupakan metode klasifikasi supervised learning metode SVM merupakan algoritma yang bekerja menggunakan pemetaan untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi.



Dataset

Kumpulan data berisi berbagai informasi yang mempengaruhi prediksi seperti Usia, Jenis Kelamin, BP, kadar Kolesterol, Rasio Na hingga Kalium dan akhirnya jenis obat.





Studi Kasus

Dalam kasus ini saya akan membuat model yang akan memprediksi dari Usia Jenis Kelamin Tingkat Tekanan Darah (BP) Tingkat Kolesterol (Na) hingga Potassium Ration adalah jenis obat terlarang apa.



DATA EKSPLORATION



Code cell [8]:

```
data.replace("Drug", "DrugA", inplace = True)
data.replace("Drug", "DrugA", inplace = True)
data.replace("DrugC", "DrugC", inplace = True)
data.replace("DrugB", "DrugB", inplace = True)
data
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	20	F	HIGH	HIGH	25.355	DrugY
1	47	M	LOW	HIGH	13.093	DrugC
2	47	M	LOW	HIGH	10.114	DrugC
3	20	F	NORMAL	HIGH	7.796	DrugX
4	51	F	LOW	HIGH	18.043	DrugY
...
195	56	F	LOW	HIGH	11.557	DrugC
196	16	M	LOW	HIGH	12.056	DrugC
197	52	M	NORMAL	HIGH	9.894	DrugX
198	23	M	NORMAL	NORMAL	14.020	DrugX
199	40	F	LOW	NORMAL	11.349	DrugX

200 rows x 6 columns



Bagian ini bertujuan untuk menentukan nilai yang hilang dalam data ini

Code cell [9]:

```
data.isnull().sum()
```

Age	0
Sex	0
BP	0
Cholesterol	0
Na_to_K	0
Drug	0

dtype: int64



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[ ] from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

IMPORT LIBRARY DAN DATASET

PREPROCESSING



```
data.replace("drugX", "DrugX", inplace = True)
data.replace("drugA", "DrugA", inplace = True)
data.replace("drugC", "DrugC", inplace = True)
data.replace("drugB", "DrugB", inplace = True)
data
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	DrugY
1	47	M	LOW	HIGH	13.893	DrugC
2	47	M	LOW	HIGH	10.114	DrugC
3	28	F	NORMAL	HIGH	7.798	DrugX
4	61	F	LOW	HIGH	18.043	DrugY
...
186	56	F	LOW	HIGH	11.587	DrugC
188	16	M	LOW	HIGH	12.006	DrugC
187	52	M	NORMAL	HIGH	9.894	DrugX
188	23	M	NORMAL	NORMAL	14.020	DrugX
189	40	F	LOW	NORMAL	11.340	DrugX

200 rows x 6 columns

Bagian ini bertujuan untuk menentukan nilai yang hilang dalam data ini

```
[ ] data.isnull().sum()
```

```
Age      0
Sex      0
BP       0
Cholesterol  0
Na_to_K  0
Drug     0
dtype: int64
```


VISUALISASI DATA



inisialisasi

```
▶ Name= ['Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K']  
  
X = data[Name]  
Y = data['Drug']
```

Memisahkan data menjadi pelatihan dan pengujian

```
[ ] from sklearn.model_selection import train_test_split  
    X_train, X_test, Y_train, Y_test = train_test_split(X, Y, random_state=0)
```

```
[ ] X_train = pd.get_dummies(X_train, columns=Name, drop_first=True)  
    X_test = pd.get_dummies(X_test, columns=Name, drop_first=True)
```

Support Vector Machine (SVM)

```
[ ] from sklearn.svm import SVC  
  
svm = SVC()  
svm.fit(X_train, Y_train)  
print('Akurasi pengklasifikasi SVM pada pelatihan ini adalah {:.2f}'  
      .format(svm.score(X_train, Y_train)))
```

Akurasi pengklasifikasi SVM pada pelatihan ini adalah 0.88



TRANNING DATA &
PENGECEKAN AKURASI DENGAN
MODEL SVM

