

អំពី data set

# អំពី data set

Data set ជាទិន្នន័យដែលគេរៀបចំហើយស្រេច តាមទម្រង់មួយ សម្រាប់យកទៅបង្កើត model បង្រៀនម៉ាស៊ីន។

ក្នុងមេរៀននេះ យើងយក dataset របស់ផ្កា iris ដែលមានស្រាប់នៅក្នុងបណ្ណាល័យ sklearn

# ទិន្នន័យ iris



Species

setosa

versicolor

virginica

Label

0

1

2

# តារាងទិន្នន័យ iris

**Fisher's *Iris* Data**

Dataset Order	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	<i>I. setosa</i>
2	4.9	3.0	1.4	0.2	<i>I. setosa</i>
3	4.7	3.2	1.3	0.2	<i>I. setosa</i>
4	4.6	3.1	1.5	0.2	<i>I. setosa</i>
5	5.0	3.6	1.4	0.3	<i>I. setosa</i>
6	5.4	3.9	1.7	0.4	<i>I. setosa</i>
7	4.6	3.4	1.4	0.3	<i>I. setosa</i>
8	5.0	3.4	1.5	0.2	<i>I. setosa</i>

# ទាញយកទិន្នន័យ (load dataset)

```
from sklearn.datasets import load_iris
```

```
iris_dataset = load_iris()
```

- Dataset នេះ ជាប្រភេទ numpy array ដូច្នេះហើយ យើងអាចប្រើអនុគមន៍ទាក់ទងនឹង numpy array ដើម្បីមើលវា
- iris\_dataset.data #ទិន្នន័យបានមកពីការវាស់ធ្លា
- iris\_dataset.target #ជា label

# បំបែកទិន្នន័យ

អនុគមន៍មានស្រាប់ក្នុង sklearn គឺ `train_test_split` សម្រាប់បែងចែកទិន្នន័យ ៧០% សម្រាប់បង្រៀនម៉ាស៊ីន និង ៣០% សម្រាប់ព្យាករណ៍ ឬវាយតម្លៃ

```
from sklearn.model_selection import  
train_test_split
```

```
x_train, x_text, y_train, y_text =  
train_test_split(iris_dataset.data,  
iris_dataset.target, random_state=0)
```

# បង្កើត model បង្រៀនម៉ាស៊ីន

ប្រើ DecisionTreeClassifier

```
from sklearn.tree import  
DecisionTreeClassifier
```

```
model = DecisionTreeClassifier()
```

```
my_model = model.fit(x_train, y_train)
```

# ព្យាករណ៍

```
result = my_model.predict(x_test)
# ក្នុងករណីមានឆ្នាំ iris ថ្មី យើងវាស់បានទិន្នន័យថ្មី
import numpy as np
new_iris = np.array([[1.4, 2.0, 3.3, 5.0]])
new_result = my_model.predict(new_iris)
```



# វាយតម្លៃ

ប្រើអនុគមន៍ score()

```
rate = my_model.score(x_test, y_test)
```