



Fitur Penting

Oleh: M Mirza Fahmi

Definisi

Fitur penting mengacu kepada sekumpulan teknik untuk memberi penilaian terhadap fitur-fitur yang mengindikasikan adanya pengaruh relatif dari setiap fitur ketika membuat sebuah prediksi.



Manfaat Fitur Penting

Pentingnya kita menentukan fitur-fitur penting diantaranya:

1. Dapat memberikan insight terhadap dataset
2. Memberikan pemahaman lebih baik terhadap model
3. Mengurangi jumlah masukan fitur (Pengurangan Dimensi)





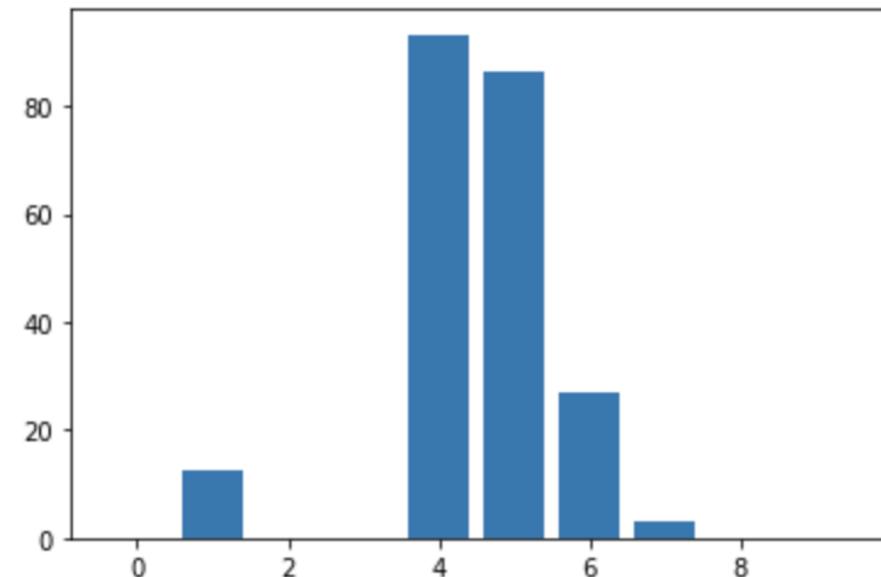
Teknik
Menghitung
Fitur Penting

Regresi Linier

```
# linear regression feature importance
from sklearn.datasets import make_regression
from sklearn.linear_model import LinearRegression
from matplotlib import pyplot
# define dataset
X, y = make_regression(n_samples=1000,
                       n_features=10,
                       n_informative=5,
                       random_state=1)
# define the model
model = LinearRegression()
# fit the model
model.fit(X, y)
# get importance
importance = model.coef_
# summarize feature importance
for i,v in enumerate(importance):
    print('Feature: %d, Score: %.5f' % (i,v))
# plot feature importance
pyplot.bar([x for x in range(len(importance))], importance)
pyplot.show()
```

Regresi Linier

```
Feature: 0, Score: -0.00000
Feature: 1, Score: 12.44483
Feature: 2, Score: -0.00000
Feature: 3, Score: -0.00000
Feature: 4, Score: 93.32225
Feature: 5, Score: 86.50811
Feature: 6, Score: 26.74607
Feature: 7, Score: 3.28535
Feature: 8, Score: -0.00000
Feature: 9, Score: 0.00000
```

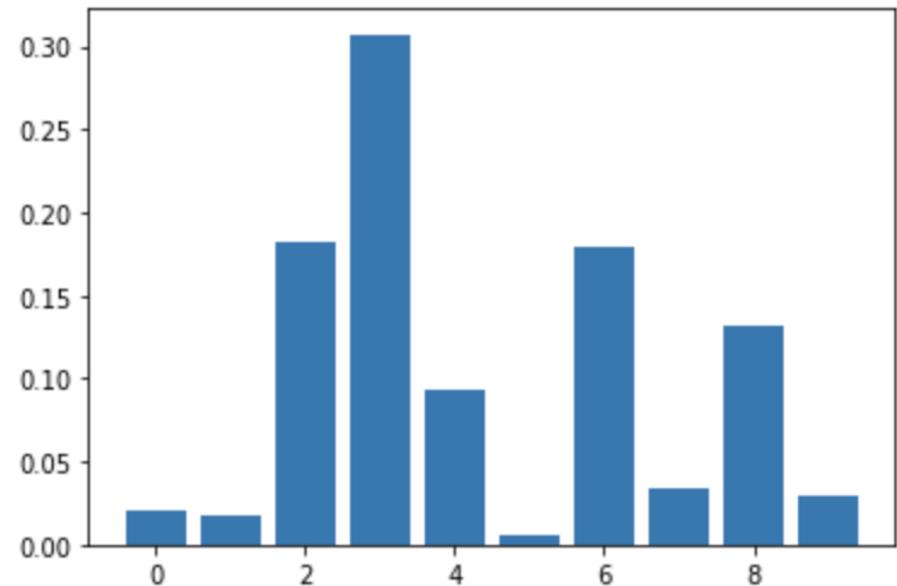


Decision Tree

```
# decision tree for feature importance on a classification problem
from sklearn.datasets import make_classification
from sklearn.tree import DecisionTreeClassifier
from matplotlib import pyplot
# define dataset
X, y = make_classification(n_samples=1000,
                           n_features=10,
                           n_informative=5,
                           n_redundant=5,
                           random_state=1)
# define the model
model = DecisionTreeClassifier()
# fit the model
model.fit(X, y)
# get importance
importance = model.feature_importances_
# summarize feature importance
for i,v in enumerate(importance):
    print('Feature: %d, Score: %.5f' % (i,v))
# plot feature importance
pyplot.bar([x for x in range(len(importance))], importance)
pyplot.show()
```

Decision Tree

Feature: 0, Score: 0.02006
Feature: 1, Score: 0.01817
Feature: 2, Score: 0.18160
Feature: 3, Score: 0.30723
Feature: 4, Score: 0.09251
Feature: 5, Score: 0.00600
Feature: 6, Score: 0.17933
Feature: 7, Score: 0.03347
Feature: 8, Score: 0.13194
Feature: 9, Score: 0.02969

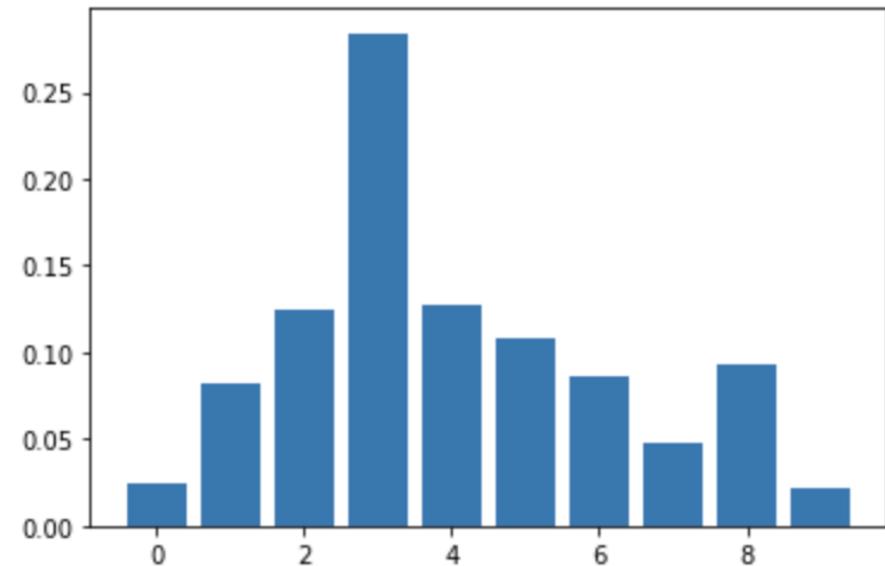


eXtreme Gradient Boosting

```
# xgboost for feature importance on a classification problem
from sklearn.datasets import make_classification
from xgboost import XGBClassifier
from matplotlib import pyplot
# define dataset
X, y = make_classification(n_samples=1000,
                            n_features=10,
                            n_informative=5,
                            n_redundant=5,
                            random_state=1)
# define the model
model = XGBClassifier()
# fit the model
model.fit(X, y)
# get importance
importance = model.feature_importances_
# summarize feature importance
for i,v in enumerate(importance):
    print('Feature: %0d, Score: %.5f' % (i,v))
# plot feature importance
pyplot.bar([x for x in range(len(importance))], importance)
pyplot.show()
```

eXtreme Gradient Boosting

```
Feature: 0, Score: 0.02464
Feature: 1, Score: 0.08153
Feature: 2, Score: 0.12516
Feature: 3, Score: 0.28400
Feature: 4, Score: 0.12694
Feature: 5, Score: 0.10752
Feature: 6, Score: 0.08624
Feature: 7, Score: 0.04820
Feature: 8, Score: 0.09357
Feature: 9, Score: 0.02220
```



Quiz

Apakah manfaat fitur penting dalam
menjabarkan insight bisnis?

Apakah pengertian dari model black-box?

Dalam regresi linier, bagaimana cara
untuk memperoleh fitur-fitur
pentingnya?