



MODUL DATA MINING

Classification



Pada modul ini dijelaskan mengenai tahapan klasifikasi dengan menggunakan bahasa pemrograman python.
Diharapkan setelah mempelajari modul ini, mahasiswa mampu memahami alur proses klasifikasi.

EPS
4

KLASIFIKASI

Masih menggunakan data pada modul sebelumnya (data titanic) dan melanjutkan langkah pada modul sebelumnya juga.

1. Apa saja kolom yang akan dipergunakan ? skenario apa yang diambil ketika ada kolom yang mengandung outlier atau missing value ? tuliskan alasannya !
2. Tambahkan algoritma decision tree, metric dan juga split data pada bagian import.

```
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn.model_selection import train_test_split
```

3. Memisahkan kolom,

Kolom apa saja yang merupakan atribut dan kolom mana yang merupakan kolom target/label. Kolom yang bersifat sebagai atribut akan disimpan pada xdata, sedangkan ydata dipergunakan untuk menyimpan kolom yang bersifat sebagai label/target/class.

```
xdata = data.drop('Survived', axis=1)
ydata = data.Survived
```

4. Membagi data untuk training dan testing secara otomatis, mempergunakan fungsi split dengan prosentase 70% dan 30%

```
xtrain,xtest,ytrain,ytest=train_test_split(xdata,ydata,test_size=0.3, random_state=1)
```

5. Membangun model

Algoritma klasifikasi ada banyak, antara lain Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis, Logistic Regression, etc. Kali ini, algoritma yang dicoba adalah decision tree

```
clf=DecisionTreeClassifier()
clf=clf.fit(xtrain,ytrain)
```

Fungsi fit dipergunakan untuk mentraining model.

6. Memprediksi data uji

Model yang dibangun dapat dipergunakan untuk melakukan prediksi terhadap data tes/data uji

```
ypred=clf.predict(xtest)
```

7. Evaluasi model

Akurasi dapat diukur dengan berbagai macam cara, salah satunya dengan memakai fungsi `accuracy_score`

```
print("Akurasi : ",metrics.accuracy_score(ytest,ypred)*100)
```

Akurasi : 78.35820895522389

Didapatkan akurasi sebesar 78%

TUGAS MAHASISWA

1. Silahkan lakukan klasifikasi dengan menggunakan 2 jenis algoritma klasifikasi lain. Bagaimana hasil akurasi dari kedua algoritma tersebut? Bandingkan antar algoritma tersebut (termasuk decision tree yang digunakan pada contoh modul)