



Minggu ke-4


Praktikum Classification

Ali Ridho Barakbah

Knowledge Engineering Research Group
Department of Information and Computer Engineering
Politeknik Elektronika Negeri Surabaya

Classification using k-NN

```
from sklearn.neighbors import KNeighborsClassifier  
  
...  
kNN=KNeighborsClassifier(n_neighbors=5, weights='distance')  
kNN.fit(training_data, training_label)  
class_result=kNN.predict(testing_data)  
precision_ratio=kNN.score(testing_data, testing_label)  
error_ratio=1-precision_ratio
```



Assignment# - Klasifikasi dengan k-NN

1. `dataset` \leftarrow `titanic.csv`
2. `test_dataset` \leftarrow `titanic_test.csv`
3. `train_data` \leftarrow ambil dataset kolom fitur (Age, Fare). Hilangkan baris data yang terdapat missing values (catat posisi data yang hilang \rightarrow `pos_missing_train`)
4. `test_data` \leftarrow ambil `test_dataset` kolom fitur (Age, Fare). Hilangkan baris data yang terdapat missing values (catat posisi data yang hilang \rightarrow `pos_missing_test`)
5. `train_label` \leftarrow ambil dataset kolom kelas (Survived), yang bukan `pos_missing_train`
6. `test_label` \leftarrow `titanic_testlabel.csv`, yang bukan `pos_missing_test`
7. `train_data` \leftarrow lakukan normalisasi pada `train_data` dengan Min-Max 0-1 (catat nilai min dan max setiap atribut)
8. `test_data` \leftarrow lakukan normalisasi pada `test_data` dengan Min-Max 0-1 (dengan nilai min dan max setiap atribut pada Langkah 7)
9. `class_result` \leftarrow Lakukan klasifikasi `test_data` terhadap `train_data` dengan k-NN ($k=1..10$)
10. Bandingkan hasil klasifikasi `class_result` dengan `test_label`. Jika tidak sama berarti error. Berapakah error ratio-nya untuk masing-masing k ?

Pengumpulan Tugas

- Buatlah coding dengan Bahasa pemrograman/tools apapun untuk semua assignment
- Buatlah laporan dalam slide ppt. Laporan terdiri dari screenshot coding dan hasil running untuk setiap assignment.
- Simpan laporan dalam file pdf dengan format penamaan:
DM_M4_NRP_namadepan.pdf
- Upload file tersebut di alamat ini:
<http://ridho.lecturer.pens.ac.id/lecture.php>
(Pilih Mata Kuliah : Data Mining)
- Deadline upload: Minggu, 2 Mei 2021