

# BLM442 Keşif aktivitesi-5: Makine Öğrenmesi

İsim/Soyisim: Murat Altınay

Öğrenci Numarası:170202109

E-mail:170202109@kocaeli.edu.tr

Kullanılan dataset adı: Metal Bands by Nation, Students Academics Performance

Dataset url: <https://www.kaggle.com/mrpantherson/metal-by-nation>,  
<https://archive.ics.uci.edu/ml/datasets/Student+Academics+Performance>

## İntihal:

Netten alınacak kısmi kod parçaları önceden kod içinde/raporda belirtilmek ve soru sorulduğunda cevaplanması durumunda sıkıntı çıkarmayacaktır. (i) İnternet kaynağını belirtmeyen/açıklayamayan/üzerinde geliştirme yapmayan veya (ii) birbirleriyle benzer/aynı çalışma teslim edenlerin aktiviteleri sıfır üzerinden değerlendirilecektir.

## KA-5 Gönderim şekli

1-pdf olarak "ogrenciNo.pdf" dokümanını gönderme -calisma dosyasında (File -> Print Preview) sonrası oluşan html dosyasını yazdır deyip pdf olarak kaydetme veya -(File -> Download as -> HTML) şeklinde indirip pdf'e çevirme <http://html2pdf.com/>

2-(File -> Download as -> Notebook) indirip "ogrenciNo.ipynb" dokümanını gönderme

suleyman.eken@kocaeli.edu.tr

**Kısım 1 (Regresyon Analizi):** Daha önce kullanmış olduğunuz (kesif aktivitesi 1 spreadsheet) veriseti üzerindeki kolonlar (Nümerik predictors ve nümerik cikti) arasındaki pozitif ve negatif korelasyonu bulunuz. r veya R<sup>2</sup> değerlerini belirleyiniz.

In [1]:

```
import pandas as pd
df = pd.read_csv("C:/Users/immur/Desktop/big_data/datasetler/metal_bands_2017.csv")
df.corr() #korelasyon analizi (R^2)
```

Out[1]:

	fans	formed	split	sahne_suresi
fans	1.000000	-0.287060	-0.107912	0.117575
formed	-0.287060	1.000000	0.419936	-0.364421
split	-0.107912	0.419936	1.000000	0.692112
sahne_suresi	0.117575	-0.364421	0.692112	1.000000

**Kısım 2 (Ekteki verisetlerinden biri üzerinde herhangi bir sınıflandırma algoritması gerçekleştirme)** Seçtiğiniz verisetini Google Drive'da belirtiniz. Her kişi farklı veriseti seçmelidir. Veriseti üzerinde gerekiyorsa birtakım ön işlemler yapılmalı, bir sınıflandırma modeli oluşturularak modelin performansı çeşitli metriklerle sunulmalıdır.

<https://docs.google.com/spreadsheets/d/19g8vDmC20FQGo2t-NQNb1MD6jwcJvflJXwXl7vQCIIA/edit?usp=sharing>

In [2]:

```
from sklearn.feature_selection import RFE
from sklearn.svm import SVR
import pandas as pd

dff2 = pd.read_csv("C:/Users/immur/Desktop/big_data/datasetler/studentinfo.csv")
```

```
X = dff2.iloc[:, :-1]
y = dff2.iloc[:, -1:]
X_change= pd.get_dummies(X)
```

In [13]:

```
from sklearn.preprocessing import LabelEncoder
y_encode = LabelEncoder().fit_transform(y)
```

In [4]:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_change, y_encode, test_size=0.2, random_state=42)
X_train
```

Out[4]:

	ge_F	ge_M	cst_G	cst_MOBC	cst_OBC	cst_SC	cst_ST	tnp_Best	tnp_Good	tnp_Pass	...	sh_Poor	ss_Govt	ss_Private	me
78	0	1	0	0	0	0	1	0	1	0	...	0	1	0	
47	0	1	1	0	0	0	0	0	0	0	...	0	1	0	
0	1	0	1	0	0	0	0	0	1	0	...	1	1	0	
12	0	1	0	0	1	0	0	0	1	0	...	1	0	1	
42	0	1	1	0	0	0	0	0	0	0	...	0	0	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
71	0	1	1	0	0	0	0	0	0	1	...	1	1	0	
106	0	1	0	1	0	0	0	0	0	1	...	0	0	1	
14	1	0	1	0	0	0	0	0	0	0	...	1	1	0	
92	1	0	0	0	0	0	1	0	1	0	...	1	1	0	
102	0	1	1	0	0	0	0	0	0	1	...	0	1	0	

104 rows × 75 columns

In [5]:

```
from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier(n_neighbors=3)
ass= model.fit(X_train, y_train)
pred= ass.predict(X_test)
```

In [6]:

```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,pred)
```

Out[6]:

```
array([[ 5,  4,  1],
       [ 0, 10,  0],
       [ 6,  0,  1]], dtype=int64)
```

In [7]:

```
from sklearn.metrics import accuracy_score
accuracy_score(pred,y_test)
```

Out[7]:

0.5925925925925926

**sunuimaidir.**

<https://www.kaggle.com/oliversinn/fifa-18-visualisation-clustering-ml/data>

In [12]:

```
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
kmeans = KMeans(n_clusters=3, random_state=1).fit(dfff)
labels = kmeans.labels_
davies_bouldin_score(dfff, labels)
```

Out[12]:

0.501253466076213

In [ ]: