

Uygulama 5. Yapay Sinir Ağlarının Geliştirilmesi ve Kullanılması – Kümeleme

Bu uygulamada, SOM ile kümeleme işlemi gerçekleştirilmiştir. Çalışmada hangi uçak firmalarının güvenli uçuş gerçekleştirdiği belirlenmiştir. Bu verilere <https://www.kaggle.com/fivethirtyeight/fivethirtyeight-airline-safety-dataset> adresinden erişilebilir. Veri setinde 56 uçak firmasına ait veriler bulunmaktadır. Bu veriler, uçakların haftalık uçuğu kilometre sayısından, 1985-1999 yılları arasındaki kaza sayılarından, 1985-1999 yılları arasındaki ölümcül kaza sayılarından, 1985-1999 yılları arasındaki ölüm sayılarından, 2000-2014 yılları arasındaki kaza sayılarından, 2000-2014 yılları arasındaki ölümcül kaza sayılarından ve 2000-2014 yılları arasındaki ölüm sayılarından oluşmaktadır. Bu çalışma ile amacımız, uçakların bu performanslarına bakıp, onları güvenli olup olmadıklarına göre kümelemektir.

Problem Tanımı: En iyi hava yolunu belirlemek.

Problem Girdileri:

- 1985-1999 yılları arasındaki kaza sayıları,
- 1985-1999 yılları arasındaki ölümcül kaza sayıları,
- 1985-1999 yılları arasındaki ölüm sayıları,
- 2000-2014 yılları arasındaki kaza sayıları,
- 2000-2014 yılları arasındaki ölümcül kaza sayıları,
- 2000-2014 yılları arasındaki ölüm sayıları,

Problem Çıktısı ya da Çıktıları:

- Güvenli uçak firmaları,
- Orta güvenli uçak firmaları
- Riskli firmalar

Üç adet çıktı belirlediğimi için çalışmadaki küme sayısı 3 olarak tasarlanmıştır. Küme sayısının belirlenmesinden sonra, ilk olarak Python'a gerekli kütüphaneler eklenir.

```
#Gerekli kütüphanelerin eklenmesi
import pandas as pd
import SimpSOM as sps
from sklearn.cluster import KMeans
import numpy as np
```

SimpSOM kütüphanesinin eklenebilmesi için ilk olarak indirilmesi gerekmektedir. Aşağıda verilen komut ile kütüphane indirilebilir;

```
- pip install SimpSOM
```

Kütüphaneler yüklendikten sonra çalışmada kullanılacak olan veri seti yüklenir. Ardından gereksiz veriler, veri setinden temizlenir.

```
#Verisetinin yüklenmesi ve gerekli verier ile kullanılması
veri = pd.read_csv("C:/Users/burak/Desktop/Uygulama 5.1/airline-safety.csv")
X = veri.drop(["airline", "avail_seat_km_per_week"], axis=1)
```

Ardından, 20x20 boyutunda bir ağ oluşturulur. Bu ağın ağırlıkları ham veriden rastgele olacak şekilde elde edilir. Son olarak ağ oluşturulurken, periyodik sınır koşulları aktif edilir.

```
#Ağın oluşturulması
net = sps.somNet(20, 20, X.values, PBC=True)
```

Daha sonra ağ eğitilir. Aşağıdaki koda göre ağı eğitmek için öğrenme katsayısı 0.01 olarak belirlenmiştir. Kod içerisindeki 10000 değeri epok sayısını göstermektedir.

```
#Ağın eğitilmesi
net.train(0.01, 10000)
```

Eğitme işleminin ardından, her bir veri noktası 2 boyutlu bir harita üzerine yerleştirilir ve kümeleme işlemi yapılır.

```
#Veri noktalarının 2 boyutlu bir haritaya gömülmesi ve kümelemenin yapılması
hrt = np.array((net.project(X.values)))
kmeans = KMeans(n_clusters = 3, max_iter = 300, random_state = 0)
```

Ardından örneklerin hangi kümelere ait olduğu belirlenir ve kümeleme işlemi tamamlanmış olur. Bu işlem fit_predict() komutu ile gerçekleştirilir.

```
#Kümeleme sonuçlarının gösterilmesi
y_kmeans = kmeans.fit_predict(hrt)
```

56 adet firma olduğu için örnek sayımız 56'dır. Her bir örneğin, bulunduğu küme fit_predict() komutu ile belirlenmiştir.

```
In [41]: y_kmeans
Out[41]:
array([2, 1, 2, 0, 2, 1, 1, 1, 1, 0, 2, 1, 2, 0, 2, 2, 1, 0, 0, 0, 0, 2,
        0, 2, 0, 1, 2, 0, 1, 1, 2, 0, 0, 2, 1, 1, 0, 2, 0, 1, 0, 1, 0, 2,
        0, 0, 2, 1, 2, 0, 1, 1, 2, 0, 2, 0])
```

Sonuçlardan da görüldüğü üzere 3 adet kümemiz bulunmaktadır. Bu kümeler 0, 1 ve 2'dir. Sonuçlara göre ilk sıradaki havayolu firmasının bulunduğu küme 2, ikinci sıradaki havayolu firmasının bulunduğu küme 1'dir.

Ardından küme sonuçlarına göre, en iyi havayolu firması belirlenir.

```
#Kümelerin etiketlerinin belirlenmesi
veri["kümeler"] = kmeans.labels_
#1 numaralı kümenin değerlerine bakılması
print(veri[veri["kümeler"]==0].head(5))
#2 numaralı kümenin değerlerine bakılması
print(veri[veri["kümeler"]==1].head(5))
#3 numaralı kümenin değerlerine bakılması
print(veri[veri["kümeler"]==2].head(5))
```

Çıktı sonuçlarına göre en iyi firmaların 1 numaralı kümeye ait olduğu belirlenmiştir.

	airline	avail_seat_km_per_week	incidents_85_99	\
0	Aer Lingus	320906734	2	
1	Aeroflot*	1197672318	76	
2	Aerolineas Argentinas	385803648	6	
4	Air Canada	1865253802	2	
5	Air France	3004002661	14	
	fatal_accidents_85_99	fatalities_85_99	incidents_00_14	\
0	0	0	0	
1	14	128	6	
2	0	0	1	
4	0	0	2	
5	4	79	6	
	fatal_accidents_00_14	fatalities_00_14	kümeler	
0	0	0	0	
1	1	88	0	
2	0	0	0	
4	0	0	0	
5	2	337	0	

Uygulama 5. Değerlendirme Soruları

- 1) Benzer bir uygulamayı alış-veriş mağazası veri seti için geliştiriniz. Veri setine <https://www.kaggle.com/shwetabh123/mall-customers> adresinden erişilebilir.
- 2) Benzer bir uygulamayı kredi kartı verileri için geliştiriniz. Veri setine <https://www.kaggle.com/arjunbhasin2013/ccdata> adresinden erişilebilir.