

# **MAKİNE ÖĞRENMESİ VİZE ÖDEVİ**

# **KÜBRA GÖKÇE**

# **182802043**

1. **PROBLEM**

IMDB top500 web sitesinden 500 adetlik veri setinin çekilmesi.

1. **Crawler Yönteminin Açıklanması**

Ödevde verileri çekmek için ‘BeautifulSoup’ ve ‘request’ kütüphanesini import ettim. Daha sonra ‘url’ isimli bir değişkene sitemin urlsini atadım. Urlden çektiğim verileri ‘for’ döngüsüyle yazdırdım.Daha sonra ‘pandas’ kütüphanesini kullanarak çektiğim verileri Excel’e aktardım ve Excel tablosu oluştu.

1. **Veri setinin özellikleri**

Verilerim 4 sütundan,500 satır veri olacak şekilde çekilmiştir.

imdb.com sitesindeki top500 filmleri çekiyor.

1. **Ön işleme adımında yapılanlar**

* İlk olarak çektiğim ‘moviename’ verilerinin içerisindeki film yıllarını NLTK kütüphanesinin ‘word\_tokenize’ metoduyla kelimelere ayırdım ve farklı bir sütun oluşturarak verileri ön işlemeye hazır hale getirdim.
* Daha sonra ‘LabelEncoder’ metoduyla çektiğim verileri sayısal değerlere çevirdim.
* Daha sonra kategorik veri olan, film türlerini içeren ‘tur’ sütununu ‘Categorical’ ve ‘get\_dummies’ metotlarıyla ‘0’ ve ‘1’ değerlerine ayırdım.
* En son önişlemeli verileri de barındıran tablomu excele aktardım.