

## Emlakjet Data Science - Case Study

Bir online iş bulma platformu için **ilan öneri sistemi** kurulacaktır. Bu sistem için 2 ana veriseti kullanılacaktır:

- Kullanıcı Hareket Verileri:** Veriye [buradan](#) ulaşabilirsiniz. Bu veride kullanıcıların sitede tıkladığı ve başvurduğu ilanların logları yer alıyor. Veri önizlemesi (Fig 1) ve kolon açıklamaları aşağıdaki gibidir.
- İlan Verileri:** Veriye [buradan](#) ulaşabilirsiniz. Bu veride ilanların id'si (item\_id), başlık (pozisyon\_adi) ve açıklama ( item\_id\_aciklama ) bilgilerini bulabilirsiniz.

event_type	client_id	item_id	ds_search_id	timestamp
click	777883739.2	3985048	250101255584	2025-01-01 20:29:09
click	1964663098	4027375	250101143218	2025-01-01 15:06:00
click	1442470832	4028900	25010165122	2025-01-01 10:33:58
click	1103680114	4019003	25010115642	2025-01-01 09:38:54

**Fig 1. Kullanıcı Veriseti Önizlemesi**

**event\_type:** Kullanıcı hareketinin tipini belirtir. click tıklama, purchase başvurma anlamına gelir.

**client\_id:** Anonimleştirilmiş kullanıcı id'si

**item\_id :** İlan id'si

**ds\_search\_id :** Kullanıcının site içinde yaptığı arama id'si. Bu arama sonrasında o satırdaki ilanı görmüş anlamına gelir.

**timestamp :** İlana başvurulma zamanı

Kurulacak öneri sistemi için beklenen adımlar aşağıdaki gibidir:

### Part 1 - Graph Kurulumu ve Analizi

- Kullanıcı verilerinden yola çıkarak bir Graph kurmanızı istiyoruz. Kurduğunuz graph'ta kenar bağlantılarını nasıl oluşturduğunuzu sebepleriyle birlikte açıklayınız.
- Kurduğunuz graph ı analiz etmenizi ve yorumlamanızı istiyoruz.

## Part 2 - Kullanıcı Verisi ile İlan Vektörü Üretme

- Part 1’de kurduğunuz graph’ı kullanarak her bir **ilan** için ilan vektörleri oluşturulmalı. Bunun için için uygun bulduğunuz bir **node embedding** metodunu kullanabilirsiniz. Seçtiğiniz embedding metodunu sebepleriyle birlikte açıklayınız.
- Oluşturulan embeddinglerin kalitesini anlayabilmek için en az 1 metrik belirleyip node embedding adımını optimize etmeye çalışınız. *(Bunun için veriden benzer olmasını ve olmamasını beklediğiniz ilanlar ile bir test kümesi oluşturabilirsiniz)*

## Part 3 - İlan Bilgileri ile İlan Vektörü Üretme

- Verilen ilan bilgilerini (başlık, açıklama) kullanarak her ilan için vektör oluşturmanızı bekliyoruz. Bunun için, önce gerekli temizleme işlemlerini yapmanızı daha sonra bir **pre-trained model** kullanarak vektörler oluşturulmalı. Kullandığınız yöntem ve modelleri seçme sebebinizi açıklamanızı bekliyoruz.

## Part 4 - Öneri Sistemi

- Bu adımda öneri sistemi geliştirmenizi istiyoruz. Bunun için 2 ilan vektörünü ((Part 2 ve Part 3) de kullanan bir sistem kurmanız gerekmektedir. Uyguladığınız sistemi sebepleriyle birlikte açıklayınız.
- Ürettiğiniz ilan vektörlerini bir **vector database** ’e kaydetmeniz i bekliyoruz (istediğiniz bir vector database seçebilirsiniz).
- Son olarak, input olarak ilan id’si alıp, en yakın N tane ilan önerisi yapan bir fonksiyon yazmanızı bekliyoruz. Bu fonksiyon vector database’e bağlanıp önerileri getirebilir olmalı.