

13.04.2018

BİL3000
2017-2018 Öğretim Yılı
Mühendislik Projesi 4. Proje

Arkadaş Öneri Sistemi

Cumhuriyet Üniversitesi
Bilgisayar Mühendisliği Bölümü

2017-2018 Öğretim Yılı

C.Ü. Bilgisayar Mühendisliği Mühendislik Projesi Dersi 4. Proje

i. Giriş

Bu proje kapsamında, bir arkadaş öneri sistemi geliştirilmesi beklenmektedir. Bu öneri sistemi kullanıcı tarafından seçilen bir kişi için arkadaş önerisinde bulunacaktır. Proje içerisinde kullanılacak olan veri, bu dersi alan öğrencilerden istenilen anketin doldurulmasıyla oluşmuştur.

Öğrenciler tarafından doldurulan anket içerisinde bu dersi alan her bir öğrenciden 10 adet arkadaşının okul numaralarının yazması ayrıca spor, sanat ve hobileri ile ilgili 15 adet aktiviteye olan ilgilerini 0-10 arası oylamaları istenmiştir. Anket içerisinde yapılan değerlendirmede 0 değeri öğrencinin sorulan aktiviteye hiç ilgi duymadığını; 10 değeri ise öğrencinin o aktiviteye en yüksek seviyede ilgi duyduğunu belirtmektedir.

Hazırlanan anket 90 adet öğrenci tarafından eksiksiz olarak doldurmuştur. Anketi doldurmayan diğer öğrenciler, Proje Anketinde daha önceden belirtildiği gibi bu projeden herhangi bir puan almayacaktır.

Toplanan veri bizler tarafından ön işleme tabi tutularak, gerekli düzeltmeler yapılmış olarak öğrencilere iki csv dosyası halinde sunulmuştur. Bu dosyalarının adları “ogrenciNetwork.csv” ve “ogrenciProfil.csv” olarak sunulmuştur.

“ogrenciNetwork.csv” dosyası içerisinde her bir satır bir öğrenciye denk gelmektedir ve her bir satır bir öğrencinin okul numarası ile başlamaktadır. Satırdaki diğer öğrenci numaraları bu öğrencinin arkadaş olduğu diğer öğrencilerin okul numaralarıdır.

“ogrenciProfil.csv” dosyasında her bir satır bir öğrenciye denk gelmektedir her bir satır bir öğrencinin okul numarasıyla başlar ve bu öğrencinin bahsi geçen aktivitelere olan puanlamaları ile devam eder. Bu puanlamalar öğrencinin profilini oluşturmaktadır. Her iki dosya dikkatlice incelendiğinde “ogrenciNetwork.csv” ve “ogrenciProfil.csv” dosyalarında öğrencilerin aynı sıra ile verildiği görülmektedir.

ii. Ödev Tanımı:

Bu proje kapsamında öğrenciden beklenen bu dersi alanlar kişiler arasından seçilen herhangi bir öğrenci için bu dersi alan diğer öğrenciler arasından 10 adet yeni arkadaş öneren ve kullanıcı grafik arayüzüne sahip yazılımın geliştirilmesidir. Geliştirilecek program üzerinden girilen bir öğrenci numarası veya kişi adı soyadı bilgisi için en yüksek puan alan kişiden başlamak üzere ve azalan sırada toplam 10 adet arkadaş önerisini Öğrenci numarası ve kişi bilgisi olacak şekilde ekrana yazdırmalıdır.

Geliştirilecek olan sistemde kullanılacak olan veriler sizlere vereceğimiz “ogrenciNetwork.csv” ve “ogrenciProfil.csv” isimli dosyalardan okunarak bu bilgiler tablolar halinde bir veri tabanı içerisinde saklanmalıdır. Ayrıca Öğrenci Numarası ile Kişilere ait Ad Soyadı bilgilerinin bulunduğu “ogrenciBilgi.csv” dosyası da sizlere verilecektir.

iii. Yöntem:

a) Birinci Aşama: Model Oluşturma

Geliştirilecek program, kullanıcı tarafından girilen bir okul numarası veya ad soyad bilgisi için öncelikle girilen kişinin arkadaşlarını “ogrenciNetwork.tbl” tablosundan çekecek, daha sonra bu arkadaşların profil bilgilerini “ogrenciProfil.tbl” tablosundan alacaktır. Her bir arkadaş için elde edilen bu profiller alt alta konularak bir matris oluşturulacaktır. En son olarak bu arkadaşlar “1” olarak etiketlenip, bu etiket (label) kolonu matrise eklenecektir. Örneğin 2015141029 no’lu öğrenci için bu matris aşağıdaki gibidir.

Öğrenci Numaranızı giriniz:	1) Sinemada film izlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	2) Yabancı dizi izlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	3) Yabancı müzik dinlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	4) Türk Halk müziği dinlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	5) Hard Rock (Metal) müzik dinlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	6) Futbola olan ilginizi (0-10) arası oylayınız:	7) Basketbola olan ilginizi (0-10) arası oylayınız:	8) Roman okumaya olan ilginizi (0-10) arası oylayınız:	9) Şiir okumaya olan ilginizi (0-10) arası oylayınız:	10) Bilgisayar oyunlarına olan ilginizi (0-10) arası oylayınız:	11) Sosyal medya (Facebook, Instagram, Twitter vs.) olan ilginizi (0-10) arası oylayınız:	12) Şans oyunlarına (İddaa, Sayısal Loto, Ganyan v.s.) olan ilginizi (0-10) arası oylayınız:	13) Şiir, öykü, blog yazmaya olan ilginizi (0-10) arası oylayınız:	14) Satranç, bric gibi zeka oyunlarına olan ilginizi (0-10) arası oylayınız:	Herhangi bir enstrüman (piyano, gitar, bağlama vs.) çalmaya olan ilginizi (0-10) arası oylayınız:	labels
2015141028	8	8	7	7	0	10	10	4	3	10	10	7	1	7	1	1
2015141014	9	5	5	5	3	10	10	6	5	10	10	9	2	8	8	1
2015141010	10	10	10	10	0	10	7	5	0	10	10	8	0	8	3	1
2015141005	10	7	5	9	7	4	4	9	10	5	10	1	10	1	8	1
2015141016	5	8	8	9	0	6	2	5	3	8	6	8	2	8	8	1
2015141057	5	8	9	4	8	5	8	1	1	9	6	1	0	3	7	1

İkinci adımda, seçilen öğrencinin hali hazırda arkadaş olmadığı kişilere ait profilleri bilgilerinin yarısı “ogrenciProfil” den alınacak daha sonra bu profiller bir önceki aşamadaki gibi alt alta konulup bir matris oluşturulacaktır. En son olarak bu kişiler “0” olarak etiketlenip , bu etiket (label) kolonu matrise eklenecektir. Örnek olarak aşağıdaki resme bakınız.

Öğrenci Numaranızı giriniz:	1) Sinemada film izlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	2) Yabancı dizi izlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	3) Yabancı müzik dinlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	4) Türk Halk müziği dinlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	5) Hard Rock (Metal) müzik dinlemeyi ne kadar sevdiğinizi (0-10) arası oylayınız:	6) Futbola olan ilginizi (0-10) arası oylayınız:	7) Basketbola olan ilginizi (0-10) arası oylayınız:	8) Roman okumaya olan ilginizi (0-10) arası oylayınız:	9) Şiir okumaya olan ilginizi (0-10) arası oylayınız:	10) Bilgisayar oyunlarına olan ilginizi (0-10) arası oylayınız:	11) Sosyal medya (Facebook, Instagram, Twitter vs.) olan ilginizi (0-10) arası oylayınız:	12) Şans oyunlarına (İddaa, Sayısal Loto, Ganyan v.s.) olan ilginizi (0-10) arası oylayınız:	13) Şiir, öykü, blog yazmaya olan ilginizi (0-10) arası oylayınız:	14) Satranç, bric gibi zeka oyunlarına olan ilginizi (0-10) arası oylayınız:	15) Herhangi bir enstrüman (piyano, gitar, bağlama vs.) çalmaya olan ilginizi (0-10) arası oylayınız:	labels
2014123056	9	7	3	8	1	7	5	6	4	8	5	0	3	7	4	0
2015141003	6	7	8	5	2	2	2	8	4	8	6	2	5	3	6	0
2015123017	9	7	7	7	6	7	6	7	8	8	9	0	6	8	7	0
2015123003	6	9	4	6	0	10	7	6	3	5	7	0	0	7	0	0
2015123019	10	3	5	8	0	0	0	10	8	7	7	0	0	8	10	0
2015123035	8	9	10	10	4	7	7	7	9	5	5	3	5	7	4	0
2015123010	10	10	6	5	1	5	5	4	1	0	4	0	0	2	2	0
2015123020	9	6	10	7	10	2	2	6	5	10	6	4	3	8	8	0
2015123110	0	0	10	10	0	0	0	0	0	10	10	0	0	10	10	0
2015123061	8	8	3	10	0	10	0	2	0	5	6	1	1	0	7	0
2015123045	5	8	9	8	9	0	6	4	6	10	5	0	9	9	4	0
2015123038	10	8	0	7	6	5	8	10	8	4	7	0	6	10	0	0

Arkadaşlardan elde edilen matris ile arkadaş olmayanlardan elde edilen matris birleştirilerek (alt alta konularak) bir eğitim seti oluşturulacaktır.

Oluşturulan bu eğitim seti kullanılarak lojistik regresyon (*logistic regression*) yapılacaktır.

(Ekstra bilgi: Ortaya çıkan lojistik regresyon modelinin kolonlar (aktiviteler) için olan katsayıları bu aktivitelerin kişinin arkadaş seçimindeki önemini gösterir. Örneğin bir aktiviteye denk gelen katsayı pozitif ise kişinin bu aktiviteye ilgi gösterenlerle arkadaşlık yapma olasılığının yüksek olduğu ortaya çıkmaktadır. Diğer taraftan, negatif bir katsayıya sahip bir aktivite için, kişinin bu aktiviteye ilgi gösterenlerle arkadaş olma olasılığının düşük olduğu ortaya çıkmaktadır.)

b) İkinci Aşama: Modelin Uygulanması

Birinci aşamada seçilen kişi ile arkadaş olmayanların kullanmadığımız diğer yarısındaki (eğitim seti oluşturmak için kullanılan satırlar dışında olanlar) kişilerin profilleri, “ogrenciProfil.tbl” tablosundan alınıp alt alta konularak bir test seti elde edilecektir.

Birinci aşamada elde edilen lojistik regresyon modeli oluşan test setine uygulanacaktır. Bu uygulama sonucunda bu test setindeki her bir kişinin, seçilen kişi ile arkadaş olma olasılığı hesaplanacaktır.

Daha açık bir ifade ile, matematiksel olarak örneğin x kisisi bu test setindeki bir öğrenci ve, y seçilen öğrenci olsun, bu durumda lojistik regresyon bize $P(x \text{ ile } y \text{ arkadas}|x)$ olasılığını verecektir.

c) Üçüncü Aşama: Kişilerin sıralanması ve Arkadaş Önerilerinin Yapılması

Test setindeki öğrenciler için, ikinci aşamada elde edilen arkadaş olma olasılıkları büyükten küçüğe doğru sıralanacaktır. Bu sıralama sonucunda en yüksek olasılığa sahip 10 öğrenciye ait bilgiler arkadaş önerisi olarak program çıktısını oluşturacaktır.

iv. Değerlendirme Kriterleri

Görevler	Açıklamalar	Oran
Görev 1	<ul style="list-style-type: none">Öğrenci profili oluşturma (Anketin Doldurulması)Anket doldurmayan öğrenciler değerlendirmeye tabi değildir.	5%
Görev 2	<ul style="list-style-type: none">Kullanıcı Grafik Arayüzünün tasarımıCSV uzantılı Öğrenci Profilleri dosyasının okunarak veri tabanına kaydedilmesi	10%
Görev 3	<ul style="list-style-type: none">Seçilen bir öğrenci için Matrisin oluşturulmasıMatrisler üzerinde Etiketleme (Label) işleminin gerçekleştirilmesiArkadaş ve Arkadaş olmayan kişilere ait profil bilgilerinin birleştirilerek Eğitim Setinin oluşturulmasıOluşturulan matris üzerinde Lojistik Regresyon uygulanması	40%
Görev 4	<ul style="list-style-type: none">Bu Regresyon sonucu elde edilen katsayıları kullanarak seçilen öğrenci ile test seti üzerinde arkadaşı olmayan kişiler arasındaki benzerliklerin hesaplanması.Test seti içerisinde hesaplanan benzerliklerin sıralama sonucunda en yüksek olasılığa sahip 10 öğrenciye ait bilgiler arkadaş önerisi olarak sunulması.Proje raporunun hazırlanarak ciltli kapaklı çıktısının ilgili kişiye teslim edilmesi.	45%
Toplam:		100%

Hazırlayanlar:

Dr. Fırat İsmailoğlu

Dr. Emre Ünsal