**2016-2017 Makine Öğrenmesi Final Ödevi**

**Konu :** Bu ödevde, film izleyicilerine film önermek üzere kullanılmak amacıyla İşbirlikçi Filtreleme(Collaborative Filtering) algoritması tasarlanacak ve gerçeklenecektir.

**Veri Seti :** Ödevde NetFlix Prize’dan(<http://www.netflixprize.com/rules>) alınan film oylamalarının bir bölümünden oluşan **NetFlix.zip** altında bulunan eğitim ve test dosyaları kullanılacaktır. **TrainingRatings.txt** dosyası eğitim için, **TestingRatings.txt** dosyası test için kullanılacaktır. Eğitim dosyasında 3.25 milyon oy, test dosyasında 100 000 oy bulunmaktadır. Bu dosyalarda her satır bir izleyicinin bir film için verdiği oy değerini göstermektedir. Her satırın formatı MovieID, UserID, Ratingşeklindedir.

UserID 1-2649429 arasında, bazıları kullanılmış izleyici numaralarıdır. Rating 1-5 arası(1 kötü, 5 en iyi) oy değerleridir.

Örneğin test veri setindeki aşağıdaki satırın açıklaması yanındaki gibidir:

8,573364,1.0 // 8. film için 573364 numaralı kullanıcının verdiği oy 1.0

Veri seti ile ilgili detaylı açıklama için NetFlix Prize’dan alınan README dosyasını okuyunuz.

**Yapılacak işlemler:**

1. İşbirlikçi filtreleme yöntemi ile, iki izleyicinin film beğenileri arasındaki korelasyonu Pearson katsayısı ile bulan ve buna bağlı olarak bir izleyicinin bir film için vermesi beklenen oy değerini hesaplayan algoritmayı tasarlayarak kodunu yazınız.
2. Sistemin başarısını aşağıdaki şekilde ölçünüz:
   1. Eğitim dosyasında verilen oylama değerlerini kullanarak test setinde her kullanıcının oyladığı her film için tahmin edilen oylama değerini hesaplayarak eğitim ve test dosyalarındaki ile aynı formatta **PredictRatings.txt** dosyasına yazınız. Tahminleri tam sayıya çevirmeyiniz, kesirli sayı olarak kullanınız.
   2. Yaptığınız tahminlerin doğruluğunu değerlendirmek için, **TestingRatings.txt** dosyasındaki kullanıcıların film için verdiği **gerçek oy değeri** ile sizin hesapladığınız **tahmin edilen oy değeri** arasındaki farkı **Mean Absolute Error(MAE) (**[**http://en.wikipedia.org/wiki/Mean\_absolute\_error**](http://en.wikipedia.org/wiki/Mean_absolute_error)**)** metriği ile hesaplayınız.
   3. PredictRatings.txt dosyasındaki oy değeri 4.0’dan büyük her filmi izleyiciye önerebilirsiniz. Bu filmleri **RecommendMovie.txt** dosyasına MovieID, UserID formatı ile yazınız.

**Teslim Edilecekler:**

**1. Rapor :** Giriş, uygulama ve sonuç bölümlerinden oluşan makale düzeninde bir rapor hazırlayınız.

* Giriş bölümünde, yaptığınız çalışmayı 1 paragraf halinde tanıtınız.
* Uygulama bölümünde, yaptığınız başarılı 5 tahmini ve başarısız 5 tahmini veriniz. Bunların neden başarılı veya neden başarısız tahminler olduğunu yorumlayınız. Hesapladığınız MAE ve RMSE değerlerini veriniz.
* Sonuç bölümünde, oy tahmini sisteminizin başarısını MAE değerlerine göre yorumlaysraak, bu sistemin film önerisi için kullanılmasının uygunluğu hakkındaki görüşlerinizi yazınız. Sistemin iyileştirilmesi için önerileriniz varsa bunları belirtiniz.

**2. Kaynak Kod :** Yazdığınız programların kaynak kodlarınıaçıklamalı olarakhazırlayarak veriniz. Kaynak kodda İşbirlikçi Filtreleme ve hata hesabı için **hazır kütüphane kullanmayınız**, kodu kendiniz yazınız ☺

Teslim edilecekleri tek bir pdf dosya içinde birleştirip dosyanın ismini okul numaranız ve ödev numarası olarak veriniz. (10501002hw1.pdf gibi).

**Teslim Tarihi:** Ödevinizi **15 Ocak Pazar 23.59’a** kadar e-maille gönderiniz. Not girişi yapılabilmesi için geciken ödevler kabul edilmeyecektir.