

# 2024-2025 YAZILIM GELİŞTİRME I DERSİ PROJE I

## Proje Ödevi: Apriori Algoritması Kullanan Film Öneri Sistemi

### 1. Proje Tanımı

**Proje Adı:** Apriori Algoritması Kullanan Film Öneri Sistemi

**Amaç:** Bu projenin amacı, MovieLens 20M veri setini kullanarak kullanıcıların film izleme davranışlarını analiz etmek ve bu analizler doğrultusunda kişiselleştirilmiş bir film öneri sistemi geliştirmektir. Apriori algoritması aracılığıyla, kullanıcıların izledikleri ve beğendikleri filmler arasındaki ilişki kuralları çıkarılacak ve bu kurallar sayesinde kullanıcılara daha isabetli öneriler sunulacaktır.

### 2. Veri Seti ve Kullanılan Yöntemler

- MovieLens 20M Veri Seti:** Kullanıcıların film değerlendirmelerini, izleme alışkanlıklarını ve çeşitli meta verileri içerir.
- Apriori Algoritması:** Veri madenciliğinde kullanılan bu algoritma, belirli bir filmi beğenen kullanıcıların diğer beğenilerini tahmin etmek için kullanılır. Kullanıcılar arasındaki çoklu film ilişkilerini keşfetmek için sıklık ve ilişki analizi yapar. Bu projede Apriori algoritması kullanılarak sık öge kümeleri ve birliktelik kurallarının çıkarılması istenmektedir.

### 3. Proje Adımları

#### 3.1. Veri Hazırlığı

- Veri Temizleme:** Eksik veya gereksiz verilerin temizlenmesidir. Film derecelendirmeleri, zaman damgaları (timestamps) ve kullanıcı bilgileri incelenecektir. Aşağıda bazı izleyiciler ve onların izlemiş olduğu filmlerin yer aldığı bir tablo örnek olarak verilmiştir. Tablo ve içindeki veriler ihtiyaca göre oluşturulmalıdır.

İzleyiciler	İzlenen Filmler
İzleyici 1	Film 1, film 3, film 4, film 6
İzleyici 2	Film 1, film 2, film 4, film 7, film 9
İzleyici 3	Film 2, film 3, film 5, film 7, film 10
İzleyici 4	Film 1, film 5, film 6, film 9
İzleyici 5	Film 2, film 3, film 4, film 6, film 7, film 8

- **Veri Dönüşümü:** Kullanıcı-matris ilişkisi oluşturulacak ve kullanıcıların beğendiği filmlere göre veri seti yeniden düzenlenecektir. Aşağıda “Veri Temizleme” aşamasında örnek olarak verilen tablodaki bilgiler kullanılarak elde edilmiş izleyiciler-izlenen filmler matris ilişkisini gösteren örnek bir tablo verilmiştir. Tablodaki 1’ler ve 0’lar izleyiciler tarafından ilgili filmin izlenip izlenmediğini ifade etmektedir ve örnek olarak verilmiştir. İhtiyaca göre matris içeriği değiştirilmelidir.

İzleyiciler\Filmler	1	2	3	4	5	6	7	8	9	10
1	1	0	1	1	0	1	0	0	0	0
2	1	1	0	1	0	0	1	0	1	0
3	0	1	1	0	1	0	1	0	0	1
4	1	0	0	0	1	1	0	0	1	0
5	0	1	1	1	0	1	1	1	0	0

### 3.2. Film Türlerinin ve Popüler Filmlerin Belirlenmesi

- **Film Türleri:** Veri setindeki filmler, aksiyon, komedi, dram, bilim kurgu gibi toplam 10 farklı türde sınıflandırılacaktır.
- **Popüler Filmler:** Her tür için en çok izlenen ve beğenilen 50 film tespit edilecektir.

### 3.3. Apriori Algoritmasının Uygulanması

Bu projede Apriori algoritması kullanılarak sık öge kümeleri ve birliktelik kurallarının çıkarılması istenmektedir.

- **Destek, Güven ve Lift Değerleri:** Kullanıcıların izledikleri filmler arasındaki birliktelik kuralları çıkartılacaktır. Bu değerler, filmler arasındaki ilişkilerin tanımlanmasına yardımcı olacaktır.
  - Destek: Belirli bir filmi izleyen kullanıcıların oranı.
  - Güven: Belirli bir filmi izleyen bir kullanıcının diğer filmi izleme olasılığı.
  - Lift: İki filmin birlikte izlenme olasılığının, rastgele izlenme olasılığına göre artış oranı.
- **Sık Kullanılan Öge Kümelerinin Bulunması Örneği:** Minimum destek oranını %40 olarak seçilmiş olsun. Bu demektir ki bir film kombinasyonu, izleyicilerin %40'ında görülüyorsa sık bir öge kümesi olarak kabul edilecektir. İzleyiciler-izlenen filmler matris ilişkisi verileri kullanılarak aşağıdaki hesaplamalar yapılmıştır.
  - Film 1: 5 kişiden 3 kişi izledi (İzleyici 1, 2, 4) — %60 destek
  - Film 2: 5 kişiden 3 kişi izledi (İzleyici 2, 3, 5) — %60 destek
  - Film 3: 5 kişiden 3 kişi izledi (İzleyici 1, 3, 5) — %60 destek

- Film 4: 5 kişiden 3 kişi izledi (İzleyici 1, 2, 5) — %60 destek
- Diğer filmler %40'ın altında kaldığı için dikkate alınmaz.
- **Birliktelik Kurallarının Bulunması Örneği:** Sık kullanılan öge kümelerine dayalı olarak birliktelik kuralları çıkarılır. Şimdi, güven ve lift değerlerini bulalım.

#### **Örnek Kural:**

**Kural:** "Film 1 izleyen, Film 4'ü de izler."

- Destek: Bu kuralı destekleyen izleyicilerin oranı. Örneğin, Film 1 ve Film 4'ü birlikte izleyenler (İzleyici 1, İzleyici 2) toplam izleyicilere göre %40 oranında.
- Güven: Film 1 izleyenlerin Film 4'ü izleme olasılığı. Film 1'i izleyen 3 kişiden 2'si Film 4'ü izlemiş, yani %66,6 güven.
- Lift: Bu iki filmin rastgele izlenme olasılığına göre izlenme oranının artışıdır. Eğer lift değeri 1'den büyükse, iki öge arasında pozitif bir ilişki vardır.

### **3.4. Veri Yapıları ve Algoritma Analizi**

Kullanılan Apriori algoritmasının performansını artırmak için uygun veri yapılarının kullanımı büyük önem taşımaktadır. Algoritmanın zaman karmaşıklığını ve kullanılan bellek alanını önemli ölçüde düşürerek daha verimli çalışabilmesine olanak sağlamaktadır. Bu nedenle, bu projede sık öge kümelerinin depolanmasında uygun bir ağaç veri yapısı kullanılması istenmektedir. Tercih edilen ağaç veri yapısının neden tercih edildiği raporda açıklanmalıdır. Ardından sık öge kümelerinin depolanmasında ağaç veri yapısı kullanıldığı ve kullanılmadığı durumlar zaman karmaşıklığı ve kullanılan bellek alanı açısından raporda açıklanarak bir performans analizi yapılması istenmektedir.

### **3.5. Öneri Sistemi Tasarımı**

Öneri sistemi için bir arayüz (GUI) tasarlanmanız istenmektedir.

Öneri sisteminin toplamda 4 çeşit öneride bulunması istenmektedir. İstenilen öneriler aşağıda listelenmiştir.

- **Birliktelik Kurallarına Dayalı Öneriler:** İzleyicilerin beğendikleri filmler arasındaki birliktelik kurallarına göre yeni filmler önerilmesi istenmektedir.
  - Film Türüne Göre Öneriler
  - Film İsmine Göre Öneriler
- **Kişiselleştirilmiş Öneriler:** Her izleyicinin kendi beğenisine uygun filmlerden yola çıkılarak öneriler sunan kişiselleştirilmiş bir deneyim sağlanması istenmektedir.
  - Film Türüne Göre Öneriler
  - Film İsmine Göre Öneriler

- **Arayüz (GUI) Tasarımı:** Arayüz üzerinde kullanıcıya bilgilendirme yapmak için bir çıktı ekranı, seçim yapabilmesi için 2 nesne alanı ve bir buton bulunmalıdır. (Arayüzü oluştururken nesneye yönelik programlama yöntemlerinin uygulanması istenmektedir.)
  - Çıktı ekranında kullanıcıya önerilen filmin adı gösterilecektir.
  - Alanlardan ilkinde, kullanıcının ne tür öneri istediğine dair kullanıcıya seçenek sunulmalıdır. Örneğin; birinci alandaki ilk seçenekte “Popüler Film Önerileri” yazmalı ve bu seçenek seçildiğinde birliktelik kurallarına dayalı öneriler yöntemini kullanarak sistemin önerilerde bulunulması istenmektedir. Birinci alandaki ikinci seçenekte “Kişiselleştirilmiş Film Önerileri” yazmalı ve bu seçenek seçildiğinde kişiselleştirilmiş öneriler yöntemini kullanarak sistemin önerilerde bulunulması istenmektedir.

**Not:** “Kişiselleştirilmiş Film Önerileri” yönteminde belirli bir kullanıcının beğenisine göre öneride bulunulacağı için “Kişiselleştirilmiş Film Önerileri” seçildiğinde veri setinde yer alan kullanıcılardan birinin seçilmesini ve seçilen kullanıcının verilerinin kullanılmasını sağlayan bir listbox eklenmelidir.
  - Alanlardan ikincisinde, önerinin hangi şartlara göre yapılacağı kullanıcıya seçenek olarak sunulmalıdır. Örneğin; ikinci alandaki ilk seçenekte “Film Türüne Göre Öneriler” yazmalı ve film önerisinde bulunurken eğer öncesinde “Popüler Film Önerileri” seçilmişse, birliktelik kurallarına dayalı öneriler yöntemini kullanarak ve filmin türünü dikkate alarak bir öneride bulunulmalıdır. Eğer öncesinde “Kişiselleştirilmiş Film Önerileri” seçilmişse, kişiselleştirilmiş öneriler yöntemini kullanarak ve filmin türünü de dikkate alarak bir öneride bulunulmalıdır.

**Not:** “Film Türüne Göre Öneriler” şartı seçildiğinde, öneri sisteminin dikkate alabilmesi ve kullanıcının film türünü seçebilmesi için film türlerinin yer aldığı bir listbox eklenmelidir.
  - Butonun işlevi ise, tıklandığında önerilen filmin çıktı ekranında gösterilmesidir.

#### 4. Sonuçlar ve Beklenen Çıktılar

- **Verimli Bir Film Öneri Sistemi:** Kullanıcıların izleme geçmişlerine dayanarak doğru ve isabetli film önerileri yapılacaktır.
- **Analitik Raporlar:** Film türleri arasındaki ilişkiler ve popüler filmlerin hangi kategorilerle ilişkilendirildiği üzerine detaylı raporlar sunulacaktır.

PROGRAMLAMA DİLİ OLARAK PYTHON VEYA JAVA KULLANILABİLİR, PYTHON KULLANILMASI ÖNERİLİR. SUNUM YAPMAYANLAR, RAPOR TESLİM ETMEYENLER SIFIR ALACAKTIR. PROJELER BİREYSELDİR. KODLAR YAPAY ZEKA KOPYALAMA TESTİNE TABİ TUTULACAKTIR. PROJEDE BAZI KODLAR SİLDİRİLİP, YENİDEN KODLAMANIZ İSTENEİLİR.