

ÜSKÜDAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI
2023-2024
FİNAL PROJE RAPORU

ÖĞRENCİ NUMARASI: 214327027

ÖĞRENCİ ADI-SOYADI: HAMZA TALİP EKŞİ

PROJE BAŞLIĞI: VERİ MADENCİLİĞİNDE MARKET SEPET ANALİZİNİN FİLM SEKTÖRÜNDE ETKİSİNİN OLUP OLMAYACAĞININ İNCELENMESİ

-ÖZET-

Film endüstrisi ya da sinema endüstrisi (film sektörü veya sinema sektörü olarak da bilinir), film yapım şirketleri, film stüdyoları, sinematografi, animasyon, film yapımları, senaryo yazımı, ön yapım, post prodüksiyon, film festivalleri, dağıtım gibi film yapımının teknolojik ve ticari kurumlarını ve oyuncularını, film yönetmenleri ile diğer film ekibini kapsayan bir kavramdır.

Böyle büyük mali yüklerin bulunduğu film endüstrisinde her karar bir risktir. Bu riske oyuncu seçimleri ve hatta yönetmen-film türü seçimleri bile dahildir.

Bu çalışmadaki amaç, veri madenciliği tekniklerinden market sepet analizinin bu riskleri azaltıp azaltamayacağını görmek, verilen kararlarda bir kesinliklik sağlayarak zaman kazancı sağlayıp sağlamayacağını görmektir.

Giriş bölümünde yönetmen, tür ve oyuncu seçiminin nasıl yapıldığından bahsedilip endüstriye kazandırmak istediğimiz kazançlardan bahsedilecektir.

Daha sonra çalışmanın nasıl yapılacağından bahsedilip bir araç ve bir kodlama dili aracılığı ile uygulama yapılacaktır.

Sonuç kısmında çalışmanın çıktıları ortaya konup bu yöntemin film sektörüne yararlı olup olmayacağı tartışılacaktır.

Anahtar Kelimeler: veri madenciliği, birliktelik kuralı, market sepet analizi, fp-growth algorithm, film sektörü, film

İÇİNDEKİLER

-ÖZET

-1.GİRİŞ

1.1 AKTÖR-AKTRİS SEÇİMİ NASIL YAPILIR?

1.2 TÜR SEÇİMİ NASIL YAPILIR?

1.3 YÖNETMEN SEÇİMİ NASIL YAPILIR?

2.MATERYALLER VE YÖNTEMLER

2.1 FP-GROWTH ALGORİTMASI

2.2 UYGULAMA

3. ARAÇLAR

4. KODLAR

5. SONUÇLAR

6.ÖNERİ VE TARTIŞMALAR

KAYNAKLAR

KONTROL LİSTESİ

1. GİRİŞ

Hollywood dünyanın en eski film endüstrisidir. Gişe hasılatı ve ekran sayısı bakımından da en büyüğüdür.2016 yılında dünya genelindeki sinema pazarında 38,6 milyar dolarlık gişe hasılatı elde edildi.Gişe hasılatlarında; 14,9 milyar Amerikan doları ile Asya-Pasifik birinci sırada yer alırken, 11,4 milyar ABD doları ile ABD, Kanada ve Avrupa ikinci sırada yer aldı.9,5 milyar Amerikan doları ile Orta Doğu ve Kuzey Afrika'da üçüncü sırada yer aldı.2016'dan itibaren, gişelerin en büyük pazarları sırasıyla şöyle; Amerika, Çin, Japonya, Hindistan ve Birleşik Krallık'tır.

1.1 Aktör-Aktris Seçimi Nasıl Yapılır?

Herhangi bir sinema projesinde kimin hangi rolü oynayacağı konusundaki süreç oyuncu seçimi olarak adlandırılmaktadır. Oyuncu seçimi, bir yönetmenin ele aldığı hikâyenin özünü anlayabilecek, taşıyabilecek ve aktarabilecek fiziksel ve psikolojik donanıma sahip oyuncular bulma çabası olarak tanımlanabilir.Oyuncu seçimi, hikâyeyi anlatacak karakteri tanımlamakla ilgilidir. Bu tanımlamanın tinsel boyutu, fiziksel boyutundan daha önemlidir.

Bununla birlikte yönetmenin role uygun kişiyi seçebilmesi ve karar verebilmesi için oyuncu seçimi aşamasında ‘deneme çekimi’ (audition) yapılmaktadır. Oyuncunun senaryonun kısa bir parçasını oynadığı bu süreçte, mümkünse performansları kayıt altına alınmalıdır. Böylece oyuncuların ekranda nasıl göründüklerini izleme fırsatı elde edilir. Çünkü bazı tiyatro oyuncular, kabiliyetli olsalar da, kamera önü oyunculuk tecrübesi (veya eğitimi) yoksa ekranda görünen ifadeleri abartılı olabilir. Amatör oyuncular ise doğallıkları açısından cazip görünseler de, set sürecinde yönetmene bazı sorunlar çıkarabilir.

Deneme çekimi vermek ve performans iki ayrı beceridir.Çoğu aktör iki beceriye de sahiptir, ancak iyi deneme çekimi vermesine rağmen iyi oynayamayan aktörler de vardır. Aynı zamanda işinde ve performansta çok iyi olan ancak iyi deneme çekimi veremeyen aktörler de mevcuttur. Oyuncu seçimiyle ilgili anlaşılması gereken nokta, seçimlerde hataların olabileceği ve bunun oyunculuğun doğasından kaynaklandığı (stres faktörü) bilgisidir.

Oyuncu seçimindeki problemde buradan kaynaklanır.Deneme çekiminde iyi performans verse bile eldeki senaryoya uygun bir oyuncu olmayabilir.Tersi durumda da uygun bir oyuncu olmasına rağmen deneme çekimindeki performansı yüzünden yönetmen tarafından göz ardı edilebilir.Bu süreç hem para hem zaman kaybına hemde yönetmenin ve oyuncuların kariyerinde lekeye sebep olabilir.

1.2 Tür Seçimi Nasıl Yapılır?

Tür, Fransızca “genre” sözcüğünün karşılığıdır. Sanat alanındaki yapıtların gruplandırılması, sınıflandırılması amacıyla kullanılır.Sinema endüstrisini özel girişim ve kar temelinde örgütleyen bütün ülkelerde film türleri ortaya çıkmıştır. Ancak Hollywood diğer ülke sinemalarındaki endüstri beklentilerini de şekillendiren büyük bir endüstridir. Hollywood 1920’lerden itibaren rekabetçi yapı çerçevesinde neyin nasıl satılacağına önem vermiş; formüller aramıştır.Tür filmlerinin sınıflandırılmasında ilk adımı endüstri atmış; broşürler aracılığıyla günün yerleşik eğilimlerine göre filmleri gruplandırmıştır.

Bu belirli dönemlerde belirli türlerin popüler olduğu anlamına gelir.Yani 2.Dünya Savaşı sonrası Chaplin'in komedi filmleri popülerken 60lar ve 70lerde “Rambo” tarzı aksiyon filmleri popüler olmuştur.

Türlerin birbiriyle uyumuda burada önemli hale gelir.Suç-Drama iyi bir ikili olabilecekken Belgesel-Komedi o kadar iyi bir ikili olmayabilir.

Yanlış türde bir film yapmak seyircinin ilgisi az olacağı için para kaybına sebep olacaktır.

1.3 Yönetmen Seçimi Nasıl Yapılır?

Film ekibi, prodüksiyon şirketi tarafından bir film yapımı için işe alınan insan topluluğudur. Teknik ekip, Kast'tan yani oyuncu grubundan farklı bir ekiptir. Aynı zamanda film şirketinde ya da filmin fikir haklarında pay sahibi olan prodüktörlerden de ayrılır.

Yönetmen filmin yaratıcılık bölümünden sorumlu olmanın yanı sıra içeriğin kontrolü ve filmin akışı, oyuncuların yönetimi, filmin çekileceği mekânların seçilmesi ve organize edilmesi, teknik detaylardan da (kameraların yerleşimi, ışık kullanımı, zamanlama ve filmin müziği gibi) sorumludur. Yönetmenin büyük gücü olmasına rağmen filmin yapımcılarına bağlı çalışır. Bazı yönetmenler, kendini ispat etmiş olanlar, aynı zamanda yapımcının rolünü de üstlenirler ve bu iki rol arasındaki ayrım zaman zaman belirsizleşebilir.

Prodüksiyon şirketinin verdiği “Yönetmen kim olsun?” kararı bu sebeple çok önemli bir karardır.Bazı yönetmenler bazı türlerde iyi iş çıkardıkları gibi başka türlerde tamamen başarısız olabilirler.Örnek vermek gerekirse Gani Müjde iyi bir komedi yönetmeni olabilirken kötü bir drama yönetmeni olabilir.Bu kararı yanlış vermek kötü bir film ortaya koyacağından ciddi para kaybına sebep olur.

1.4 Problemin Tanımı

Yukarıda açıklandığı üzere belirlediğimiz üç problem; oyuncu seçimi, tür seçimi ve yönetmen seçimidir.

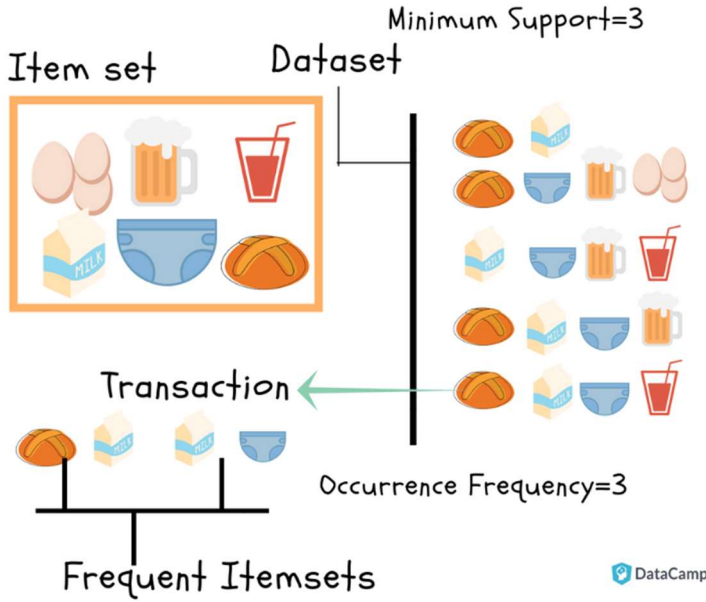
İlk problemde beraber projelerde bulunmuş oyuncuları belirleyerek oyuncu seçim sürecini kısaltmayı ve bunun getirebileceği hasarı azaltmak istiyoruz.

İkinci problemde senaristlerin uygun türleri belirleyip nasıl bir yol izlemesi gerektiğine yönelik bir fikir oluşturmak istiyoruz.

Üçüncü problemde ise eldeki senaryoya uygun yönetmen seçimini doğru yapmak adına katkıda bulunmak istiyoruz.

2. MATERYALLER VE YÖNTEMLER

Bir markette herkes sepetine farklı ürünler doldurur. Ancak her sepette genellikle bir arada bulunan ürünler vardır. Bu ürünlerden örneğin kola ile cipsin bir arada alınması beklenebilir ancak beklenmediği halde bir arada olan ürünler olabilir mi? Örneğin bira ile çocuk bezi bir arada sıkça görülür mü? İşte buna benzer kalıpları görebilmek için “market sepeti analizi” adı verilen bir analiz yapılır.



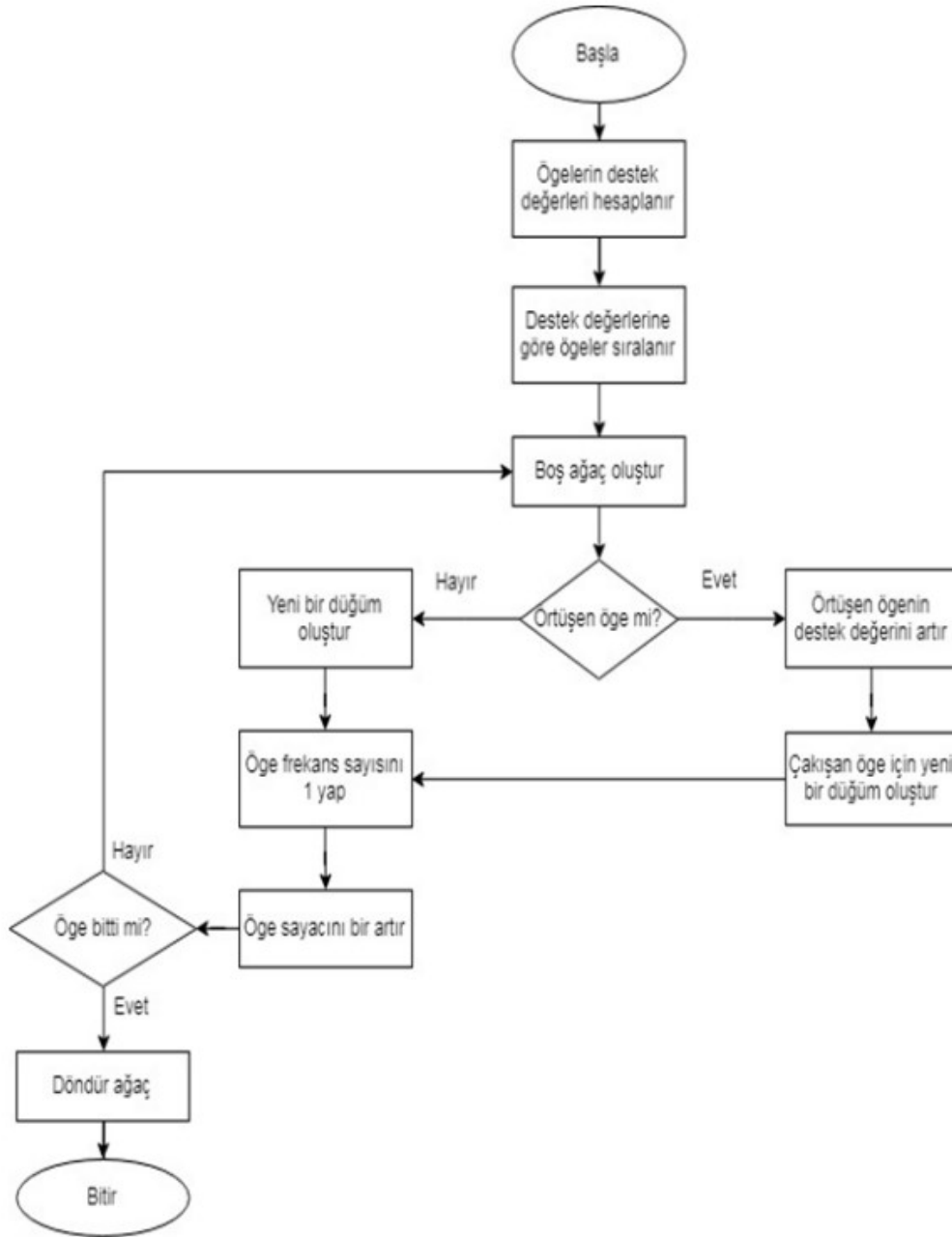
Bizde aslında kendi çalışmamızda, filmleri bir market sepeti gibi kabul edip bu sepette birlikte bulunan oyuncularını, film türlerini ve yönetmenleri inceliyoruz.

2.1 FP-Growth Algoritması

FP-Growth (Frequent-Pattern Growth) algoritması, sık rastlanan öğeleri tespit etmek için kullanılan, büyük veri kümelerinde verimli, hızlı kurallar oluşturabilen birliktelik kural algoritmasıdır. Birliktelik kural algoritmaları içerisindeki veri kümelerinde ilişkisi olan öğeleri tespit etme konusunda hızlı sonuçlar vererek, maliyeti azaltmıştır. Daha hızlı sonuçlar verebilmesinde en önemli etken, veri tabanını sık desen ağacı içinde tutuyor olmasıdır.

En çok kullanılan iki algoritmayı karşılaştıracak olursak; Apriori algoritmasına göre FP-Growth çok daha hızlı sonuçlar vermektedir. Çünkü Apriori algoritması veri tabanını defalarca kez tararken, FP-Growth algoritması veri tabanını büyüklüğü ne olursa olsun iki kez tarar.

İlk taramada veri tabanındaki tüm öğelerinin destek değeri hesaplanır. İkinci taramada ise ağaç veri yapısını oluşturur. Yeni nesne kümeler oluşturmadığı ve veri tabanını defalarca kez taramayıp sadece iki kez taradığından dolayı da çok hızlı sonuçlar vermesi yönüyle büyük veri tabanlarında kural ilişkisini tespit etmek hususunda büyük bir kazanımdır.



FP-Growth algoritması akış diyagramı

Çalışmada ilerlemeden önce bilmemiz gereken bazı terimler:

Support (Destek Değeri)

Bir ürünün/ilişkinin tüm olaylar içerisinde görülme olasılığıdır.

N: Toplam alışverişlerin sayısı

$$\text{Support } (X \rightarrow Y) = \text{Frequency}(X, Y) / N$$

Confidence (Güven Değeri)

X ürününü alan bir müşterinin Y ürününü alma olasılığıdır.

Confidence (X -> Y) = Frequency(X, Y) / Frequency(X)

2.2 Uygulama

Bu yöntemi film endüstrisine uygulamak için Kaggle'dan alınan "https://www.kaggle.com/datasets/PromptCloudHQ/imdb-data" veri setini kullanacağız. Bu veri seti 2006-2016 arasındaki 1000 filmin verisini içermekte.

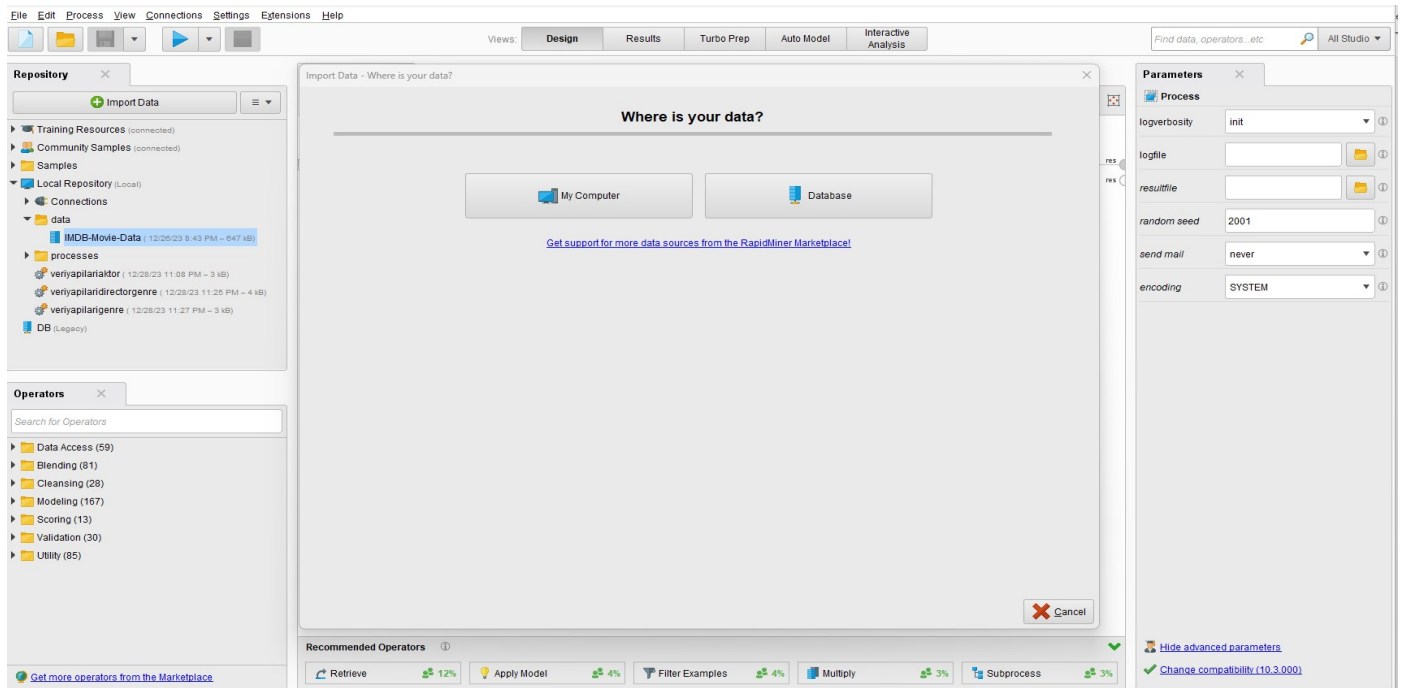
Rank		Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	757074	333.13	76.0
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	485820	126.46	65.0
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	7.3	157606	138.12	62.0
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey, Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32	59.0
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123	6.2	393727	325.02	40.0

3.ARAÇLAR

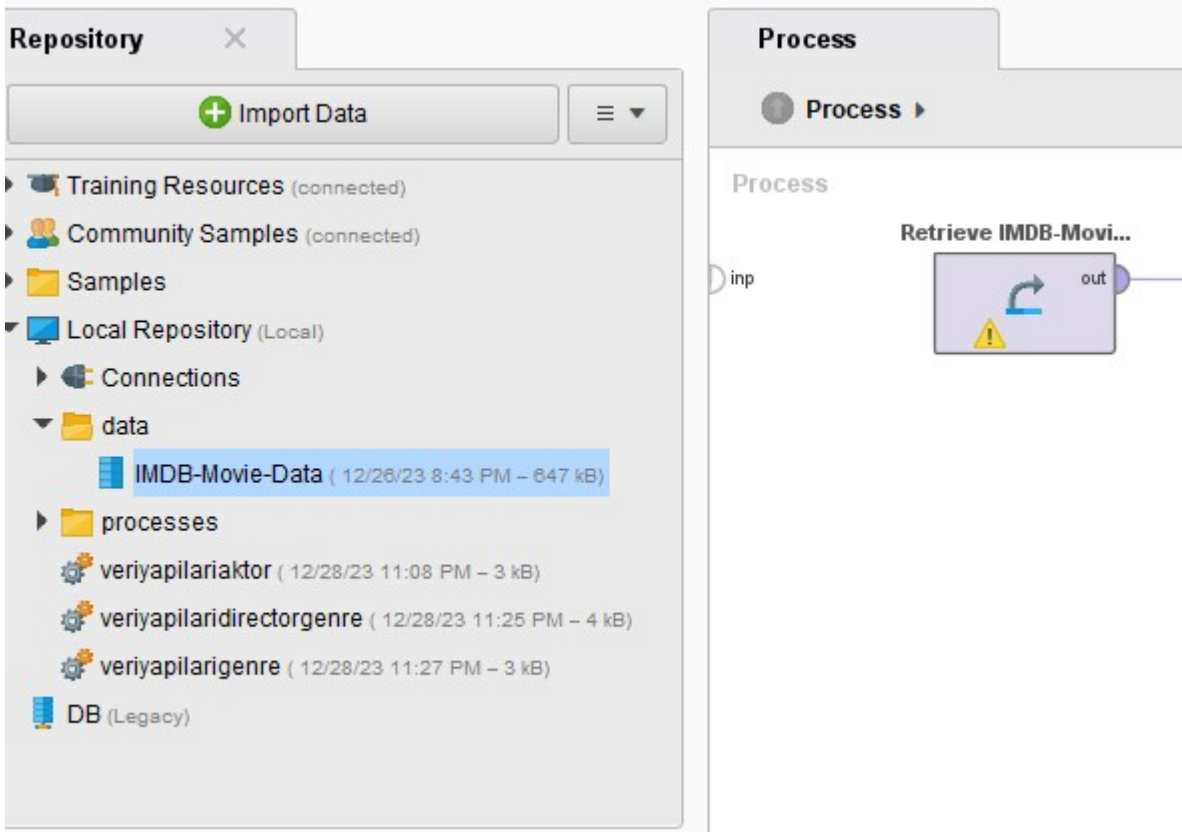
Yapmak istediğimiz çalışmayı hem kod ile hem de Rapidminer aracı ile yapmak istiyoruz. Bu bölümde Rapidminer ile yapılan çalışmalar bulunmaktadır. İlk deneme sonrasındaki çalışmalarda sadece uygulamada değişim olan bölgeler gösterilmiştir.

Öncelikle ilk çözüm odağımıza yani aktör seçimine yoğunlaşılacak.

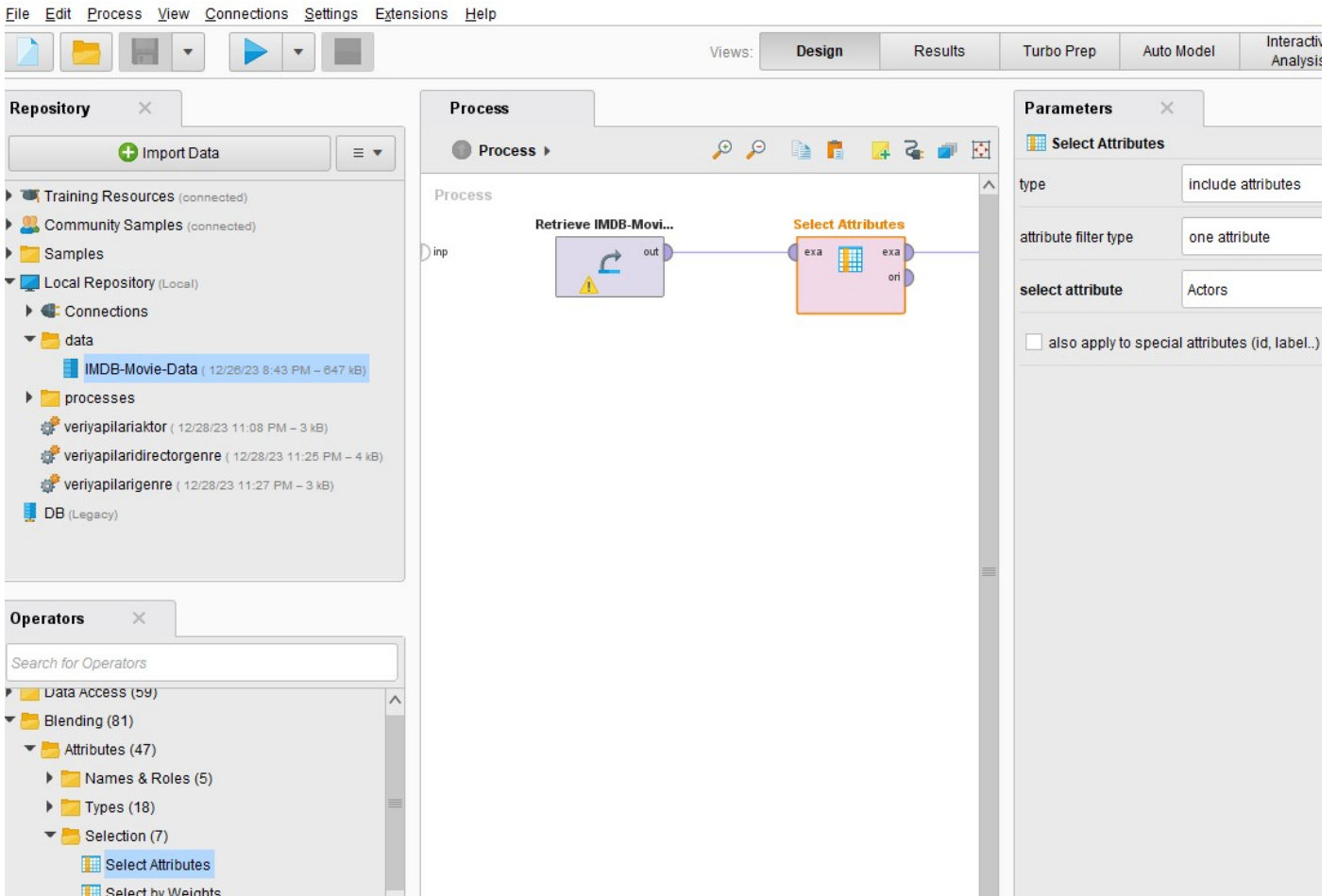
-İlk adım olarak "IMDB-Movie-Data.csv" dosyası Local Repository/data altına import edilerek yüklenmiştir.



-Veri seti ekran ortasına sürüklenerek Process içerisine alınmış olur.



-Birliktelik kuralı arayacağımız “Actors” özneliği “Select Attributes” operatörü ile seçilir.



-FP-Growth operatörünü ekleyip farklı minimum support değerleri denendi.0,004’ten fazla değerlerde sonuç elde edilemedi.Bu bir ürün 250 filmde bir görülüyor demek.

The screenshot shows the Orange3 software interface. On the left, the 'Repository' pane displays a tree view of data sources, including 'IMDB-Movie-Data'. Below it, the 'Operators' pane shows a list of operators, with 'FP-Growth' selected. The central 'Process' pane shows a workflow diagram with three operators: 'Retrieve IMDB-Movie-Data', 'Select Attributes', and 'FP-Growth'. The 'FP-Growth' operator is highlighted, and its 'Parameters' pane is open on the right. The parameters for 'FP-Growth' are as follows:

Parameter	Value
input format	item list in a column
item separators	,
use quotes	<input checked="" type="checkbox"/>
quotes character	"
escape character	\
trim item names	<input checked="" type="checkbox"/>
min requirement	support
min support	0.004
min items per itemset	1
max items per itemset	0
max number of itemsets	1000000
find min number of itemsets	<input type="checkbox"/>
must contain list	

- “Create Association Rules” operatörü eklenir. Parametre “confidence” değerini “0.7” olarak belirlemek, %70 güven ile kuralların kabul edileceğini belirtir.

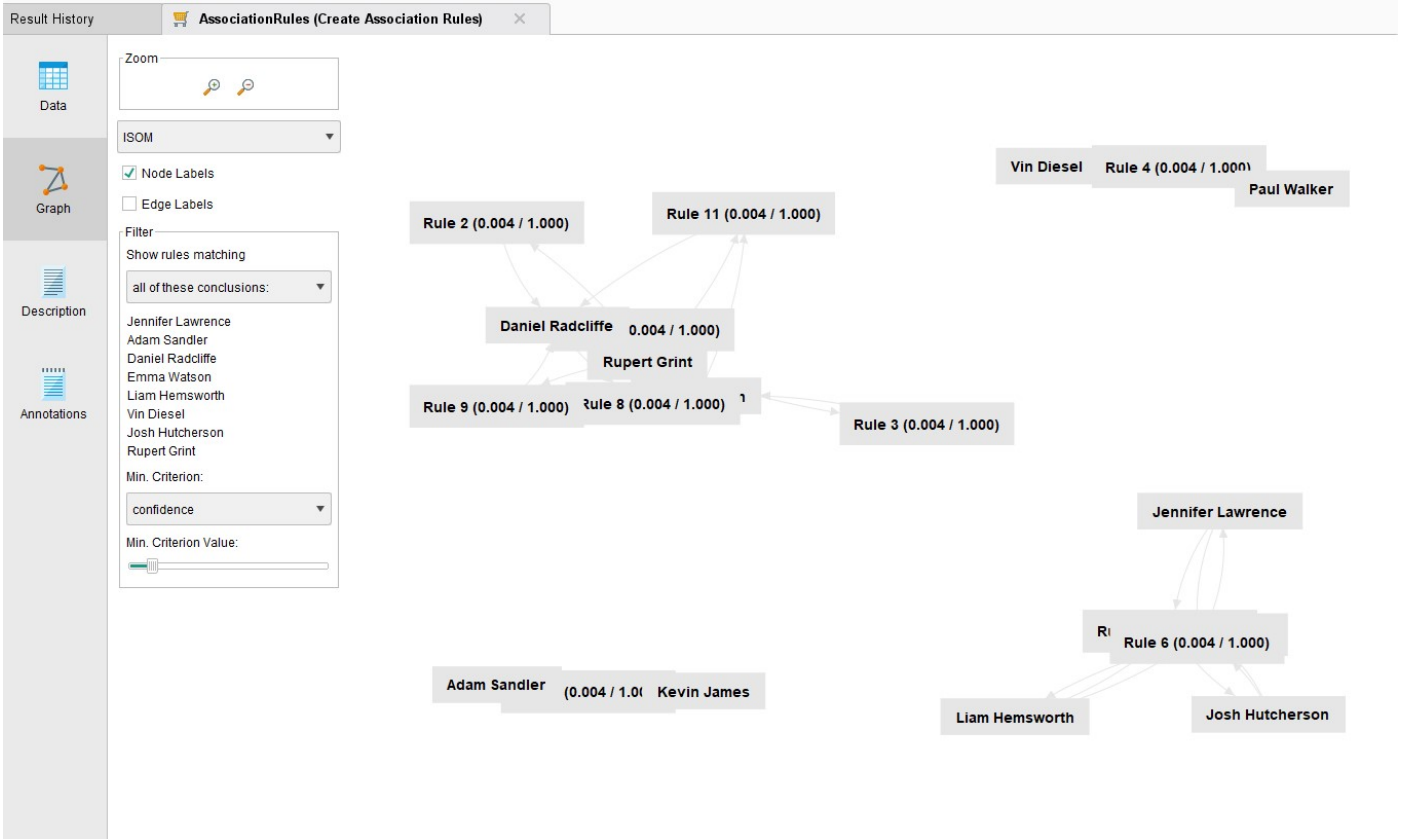
The screenshot shows the Orange3 software interface with the 'Create Association Rules' operator added to the workflow. The workflow diagram now includes four operators: 'Retrieve IMDB-Movie-Data', 'Select Attributes', 'FP-Growth', and 'Create Association Rules'. The 'Create Association Rules' operator is highlighted, and its 'Parameters' pane is open on the right. The parameters for 'Create Association Rules' are as follows:

Parameter	Value
criterion	confidence
min confidence	0.7
gain theta	2.0
laplace k	1.0

-Sonuçlardan görüldüğü üzere çok küçük bir support değerinde tam güven sağlanabildi.Daha yüksek support değerlerinde %50'nin üzerinde bile güven veren bir sonuç yok.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
1	Kevin James	Adam Sandler	0.004	1	1	-0.004	0.004
2	Rupert Grint	Daniel Radcliffe	0.004	1	1	-0.004	0.004
3	Rupert Grint	Emma Watson	0.004	1	1	-0.004	0.004
4	Paul Walker	Vin Diesel	0.004	1	1	-0.004	0.004
5	Jennifer Lawrence, Liam Hemsworth	Josh Hutcherson	0.004	1	1	-0.004	0.004
6	Jennifer Lawrence, Josh Hutcherson	Liam Hemsworth	0.004	1	1	-0.004	0.004
7	Liam Hemsworth, Josh Hutcherson	Jennifer Lawrence	0.004	1	1	-0.004	0.004
8	Daniel Radcliffe, Emma Watson	Rupert Grint	0.004	1	1	-0.004	0.004
9	Rupert Grint	Daniel Radcliffe, Emma Watson	0.004	1	1	-0.004	0.004
10	Daniel Radcliffe, Rupert Grint	Emma Watson	0.004	1	1	-0.004	0.004
11	Emma Watson, Rupert Grint	Daniel Radcliffe	0.004	1	1	-0.004	0.004

-Oluşan birliktelik kuralları şekildeki gibidir.



Bu çalışmadan elde edilen sonuç istenilen gibi değil. Öyleyse tür karşılaştırmalarına bakarak senaristlere yardımcı olmaya çalışalım.

-Öznitelik olarak “Genre” seçiyoruz.

Parameters

Select Attributes

type include attributes

attribute filter type one attribute

select attribute Genre

☐ also apply to special attributes (id, label..)

-Şekilde görüldüğü üzere support'u 0,03 kabul ettiğimizde sonuca ulaşabiliyoruz

Parameters

FP-Growth

input format

item list in a column

item separators

,

☒ use quotes

quotes character

"

escape character

\

☒ trim item names

min requirement

support

min support

0.03

min items per itemset

1

max items per itemset

0

max number of itemsets

1000000

☐ find min number of itemsets

must contain list

Edit Enumeration (0)...

AssociationRules (Create Association Rules)

les matching

ese conclusions:

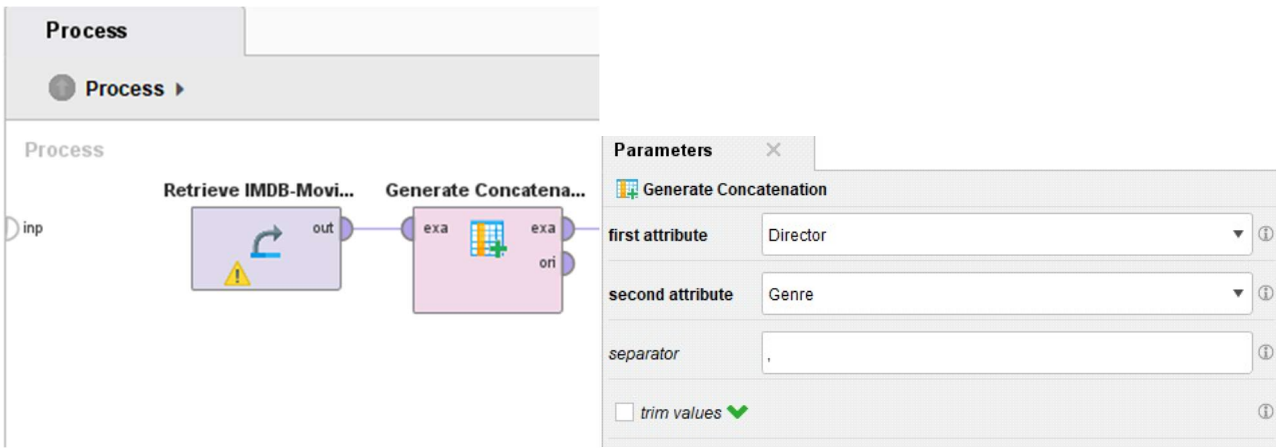
re

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain
2	Adventure, Sci-Fi	Action	0.050	0.847	0.991	-0.068
3	Biography	Drama	0.074	0.914	0.994	-0.088

Aktör seçimlerine göre daha başarılı bir deneme olsa da 100 filmde 3 kez görünme başarısı kesin bir kanı için yeterli sayılamaz.

Öyleyse son problemimize yani yönetmen seçimine bakalım.Bu seçimi belirlemek için yönetmen-tür başlıkları birleştirilip incelenmiştir.

-“Generate Concatenation” operatörü ile “Director” ve “Genre” özneliği birleştirilmiştir.



-Böylece “Select Attributes” operatöründe bu yeni özneliği seçebiliyoruz.

Parameters

Select Attributes

type include attributes

attribute filter type one attribute

select attribute Director,Genre

☒ also apply to special attributes (id, label..)

-Şekilde görüldüğü üzere support 0,005 seçildiğinde makul sonuçlar elde edilebilmiştir.

Parameters

FP-Growth

input format

item list in a column

item separators

,

use quotes

quotes character

"

escape character

\

trim item names

min requirement

support

min support

0.005

min items per itemset

1

max items per itemset

0

max number of itemsets

1000000

find min number of itemsets

must contain list

Edit Enumeration (0)...

AssociationRules (Create Association Rules)

les matching

ese conclusions:

/

re

erion:

nce

erion Value:

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain
3	Crime, Biography	Drama	0.009	0.750	0.997	-0.015
4	Romance, Mystery	Drama	0.006	0.750	0.998	-0.010
5	Adventure, Crime	Action	0.006	0.750	0.998	-0.010
6	War	Drama	0.010	0.769	0.997	-0.016
7	Animation	Adventure	0.038	0.776	0.989	-0.060
8	Crime, Mystery	Drama	0.023	0.793	0.994	-0.035
9	Comedy, Animation	Adventure	0.027	0.794	0.993	-0.041
10	Music	Drama	0.013	0.812	0.997	-0.019
11	Fantasy, Family	Adventure	0.014	0.824	0.997	-0.020
12	Michael Bay	Action	0.005	0.833	0.999	-0.007
13	David Yates	Fantasy	0.005	0.833	0.999	-0.007
14	Romance, Fantasy	Drama	0.005	0.833	0.999	-0.007
15	David Yates	Adventure, Fantasy	0.005	0.833	0.999	-0.007
16	Adventure, David Yates	Fantasy	0.005	0.833	0.999	-0.007
17	Adventure, Sci-Fi	Action	0.050	0.847	0.991	-0.068
18	Ridley Scott	Drama	0.007	0.875	0.999	-0.009
19	Comedy, Biography	Drama	0.007	0.875	0.999	-0.009
20	Sport	Drama	0.015	0.882	0.998	-0.019
21	Action, Animation	Adventure	0.009	0.900	0.999	-0.011
22	Biography	Drama	0.074	0.914	0.994	-0.088
23	Biography, History	Drama	0.014	0.933	0.999	-0.016
24	History	Drama	0.028	0.966	0.999	-0.030
25	David Fincher	Drama	0.005	1	1	-0.005

-Bu denemeden de görüldüğü üzere sonuç elde edilse de 0,005lik destek oranı istediğimiz kesinliği sağlamaz.

Bu çalışmalar kod ile yapıp sonucun aynı olup olmadığı kontrol edilecektir.

4.KODLAR

Kodlar Python dilinde Jupyter Notebook üzerinde yazılmıştır.

```
import numpy as np
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import fpgrowth
```

-Dizilerle çalışmak için numpy,veri analizi için pandas,birliktelik analizi için mlxtend kütüphaneleri kullanıldı.

```
In [2]: dataset = pd.read_csv('IMDB-Movie-Data.csv', header=0)
```

-Çalışacağımız veri setini çekiyoruz.

```
In [3]: dataset
```

```
Out[3]:
```

	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
0	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	757074	333.13	76.0
1	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	485820	126.46	65.0
2	3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	7.3	157606	138.12	62.0
3	4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey, Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32	59.0
4	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123	6.2	393727	325.02	40.0
...
995	996	Secret in Their Eyes	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015	111	6.2	27585	NaN	45.0
996	997	Hostel: Part II	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007	94	5.5	73152	17.54	46.0
997	998	Step Up 2: The Streets	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008	98	6.2	70699	58.01	50.0
998	999	Search Party	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2014	93	5.6	4881	NaN	22.0
999	1000	Nine Lives	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016	87	5.3	12435	19.64	11.0

1000 rows x 12 columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Rank                                  1000 non-null   int64
1   Title                                1000 non-null   object
2   Genre                                1000 non-null   object
3   Description                           1000 non-null   object
4   Director                             1000 non-null   object
5   Actors                               1000 non-null   object
6   Year                                  1000 non-null   int64
7   Runtime (Minutes)                    1000 non-null   int64
8   Rating                               1000 non-null   float64
9   Votes                                1000 non-null   int64
10  Revenue (Millions)                   872 non-null    float64
11  Metascore                            936 non-null    float64
dtypes: float64(3), int64(4), object(5)
memory usage: 93.9+ KB
```

```
df=dataset[ 'Actors' ]
```

```
df.info()
<class 'pandas.core.series.Series'>
RangeIndex: 1000 entries, 0 to 999
Series name: Actors
Non-Null Count  Dtype
-----  -----
1000 non-null   object
dtypes: object(1)
memory usage: 7.9+ KB
```

```
In [8]: df = list(df.apply(lambda clmn:clmn.split(',')))
df[0],df[-1]
```

```
Out[8]: (['Chris Pratt', 'Vin Diesel', 'Bradley Cooper', 'Zoe Saldana'],
        ['Kevin Spacey', 'Jennifer Garner', 'Robbie Amell', 'Cheryl Hines'])
```

```
In [9]: te=TransactionEncoder()  
te_ary=te.fit(df).transform(df)  
df=pd.DataFrame(te_ary,columns=te.columns_)  
df
```

[illegible]

- Dizi elemanlarına “one hot encoding” metodu uygulayarak, her birini bir öznitelik kolonu haline getirelim ve her bir özniteliğin projede olup olmamasına göre ikili karşılaştırma yapalım. Örnek olarak 50 Cent “0” nolu proje sepetinde varsa “True” yoksa “False” değerini alacaktır. Veri seti tabloya dönüşmüş oldu.

```
In [18]: dfs=fpgrowth(df, min_support=0.004, use_colnames=True, max_len=None, verbose=0)
dfs
```

```
In [19]: dfs.sort_values(by="support",ascending=False)
```

Out[19]:

	support	itemsets
74	0.011	(Mark Wahlberg)
61	0.011	(Christian Bale)
46	0.009	(Denzel Washington)
5	0.009	(Matthew McConaughey)
163	0.009	(Rose Byrne)
...
111	0.004	(Kate Mara)
110	0.004	(Natalie Portman)
108	0.004	(Mark Wahlberg)
106	0.004	(Russell Crowe)
239	0.004	(Adam Sandler, Kevin James)

240 rows x 2 columns

-“mlxtend” kütüphanesindeki “fpgrowth” algoritmasını uygulayarak destek değerlerini 0.004 alt limiti ile aldığımıza dikkat ederek, en yüksek değerli olanlardan başlayacak şekilde sıralayalım.

```
In [12]: association_rules(dfs, metric="confidence", min_threshold=0.7)
```

Out[12]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Robert Pattinson)	(Kristen Stewart)	0.005	0.006	0.004	0.8	133.333333	0.003970	4.97	0.997487
1	(Paul Walker)	(Vin Diesel)	0.004	0.006	0.004	1.0	166.666667	0.003976	inf	0.997992
2	(Rupert Grint)	(Emma Watson)	0.004	0.007	0.004	1.0	142.857143	0.003972	inf	0.996988
3	(Rupert Grint)	(Daniel Radcliffe)	0.004	0.008	0.004	1.0	125.000000	0.003968	inf	0.995984
4	(Rupert Grint, Emma Watson)	(Daniel Radcliffe)	0.004	0.008	0.004	1.0	125.000000	0.003968	inf	0.995984
5	(Rupert Grint, Daniel Radcliffe)	(Emma Watson)	0.004	0.007	0.004	1.0	142.857143	0.003972	inf	0.996988
6	(Emma Watson, Daniel Radcliffe)	(Rupert Grint)	0.004	0.004	0.004	1.0	250.000000	0.003984	inf	1.000000
7	(Rupert Grint)	(Emma Watson, Daniel Radcliffe)	0.004	0.004	0.004	1.0	250.000000	0.003984	inf	1.000000
8	(Josh Hutcherson)	(Liam Hemsworth)	0.005	0.006	0.004	0.8	133.333333	0.003970	4.97	0.997487
9	(Josh Hutcherson)	(Jennifer Lawrence)	0.005	0.008	0.004	0.8	100.000000	0.003960	4.96	0.994975
10	(Liam Hemsworth, Josh Hutcherson)	(Jennifer Lawrence)	0.004	0.008	0.004	1.0	125.000000	0.003968	inf	0.995984
11	(Liam Hemsworth, Jennifer Lawrence)	(Josh Hutcherson)	0.004	0.005	0.004	1.0	200.000000	0.003980	inf	0.998996
12	(Josh Hutcherson, Jennifer Lawrence)	(Liam Hemsworth)	0.004	0.006	0.004	1.0	166.666667	0.003976	inf	0.997992
13	(Josh Hutcherson)	(Liam Hemsworth, Jennifer Lawrence)	0.005	0.004	0.004	0.8	200.000000	0.003980	4.98	1.000000
14	(Kevin James)	(Adam Sandler)	0.004	0.009	0.004	1.0	111.111111	0.003964	inf	0.994980

-“association_rules” metodunu güven değerini 0,7 yaparak sonuç elde edildi.

İkinci çalışmada;

```
In [6]: df= dataset['Genre']
```

```
In [7]: df.info()
```

```
<class 'pandas.core.series.Series'>  
RangeIndex: 1000 entries, 0 to 999  
Series name: Genre  
Non-Null Count  Dtype  
-----  
1000 non-null   object  
dtypes: object(1)  
memory usage: 7.9+ KB
```

```
In [8]: df= list(df.apply(lambda clmn:clmn.split(',')))  
df[0],df[-1]
```

```
Out[8]: (['Action', 'Adventure', 'Sci-Fi'], ['Comedy', 'Family', 'Fantasy'])
```

-Çalıştığımız kolonu “Genre” olarak değiştiriyoruz.

```
In [10]: te=TransactionEncoder()  
te_ary=te.fit(df).transform(df)  
df=pd.DataFrame(te_ary,columns=te.columns_)  
df
```

```
Out[10]:
```

	Action	Adventure	Animation	Biography	Comedy	Crime	Drama	Family	Fantasy	History	Horror	Music	Musical	Mystery	Romance	Sci-Fi	Sport	Thriller
0	True	True	False	False	False	False	False	False	False	False	False	False	False	False	False	True	False	False
1	False	True	False	False	False	False	False	False	False	False	False	False	False	True	False	True	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False	False	False	False	False	False	False
3	False	False	True	False	True	False	False	True	False	False	False	False	False	False	False	False	False	False
4	True	True	False	False	False	False	False	False	True	False	False	False	False	False	False	False	False	False
...
995	False	False	False	False	False	True	True	False	False	False	False	False	False	True	False	False	False	False
996	False	False	False	False	False	False	False	False	False	False	True	False	False	False	False	False	False	False
997	False	False	False	False	False	False	True	False	False	False	False	True	False	False	True	False	False	False
998	False	True	False	False	True	False	False	False	False	False	False	False	False	False	False	False	False	False
999	False	False	False	False	True	False	False	True	True	False	False	False	False	False	False	False	False	False

1000 rows x 20 columns

-“one hot encoding” metodu uyguluyoruz.

```
In [11]: dfs=fpgrowth(df, min_support=0.03, use_colnames=True, max_len=None, verbose=0)  
dfs
```

```
Out[11]:
```

	support	itemsets
0	0.303	(Action)
1	0.259	(Adventure)
2	0.120	(Sci-Fi)
3	0.106	(Mystery)
4	0.195	(Thriller)
5	0.119	(Horror)
6	0.279	(Comedy)
7	0.051	(Family)
8	0.049	(Animation)
9	0.101	(Fantasy)
10	0.513	(Drama)
11	0.081	(Biography)
12	0.141	(Romance)
13	0.150	(Crime)
14	0.072	(Drama, Action)
15	0.155	(Action, Adventure)
16	0.052	(Drama, Adventure)
17	0.059	(Comedy, Adventure)
18	0.074	(Action, Sci-Fi)
19	0.059	(Adventure, Sci-Fi)
20	0.050	(Action, Adventure, Sci-Fi)
21	0.052	(Drama, Mystery)
22	0.035	(Mystery, Thriller)
23	0.054	(Thriller, Action)

-Destek değerini 0,03 alıyoruz.

```
In [13]: association_rules(dfs, metric="confidence", min_threshold=0.7)
```

Out[13]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Adventure, Sci-Fi)	(Action)	0.059	0.303	0.050	0.847458	2.796890	0.032123	4.569222	0.682742
1	(Animation)	(Adventure)	0.049	0.259	0.038	0.775510	2.994248	0.025309	3.300818	0.700343
2	(Biography)	(Drama)	0.081	0.513	0.074	0.913580	1.780858	0.032447	5.635286	0.477120

-Güven değerleri rapidminer ile elde ettiğimiz sonuçla aynı

Üçüncü çalışmada:

```
In [6]: df= dataset[['Director', 'Genre']].agg(', '.join, axis=1)
```

```
In [7]: df.info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 1000 entries, 0 to 999
Series name: None
Non-Null Count  Dtype
-----  -----
1000 non-null  object
dtypes: object(1)
memory usage: 7.9+ KB
```

-İki kolonu birleştirerek yeni bir öznelik elde etmiş oluyoruz.

```
In [8]: df= list(df.apply(lambda clmn:clmn.split(',')))
df[0],df[-1]
```

```
Out[8]: ([ 'James Gunn', 'Action', 'Adventure', 'Sci-Fi'],
[ 'Barry Sonnenfeld', 'Comedy', 'Family', 'Fantasy'])
```

```
In [10]: te=TransactionEncoder()
te_ary=te.fit(df).transform(df)
df=pd.DataFrame(te_ary,columns=te.columns_)
df
```

Out[10]:

	Aamir Khan	Abdellatif Kechiche	Action	Adam Leon	Adam McKay	Adam Shankman	Adam Wingard	Adventure	Afonso Poyart	Aisling Walsh	...	Will Gluck	Will Slocombe	William Brent Bell	William Oldroyd	Woody Allen	Xavier Dolan	Yimou Zhang
0	False	False	True	False	False	False	False	True	False	False	...	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	True	False	False	...	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
4	False	False	True	False	False	False	False	True	False	False	...	False	False	False	False	False	False	False
...
995	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
996	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
997	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False
998	False	False	False	False	False	False	False	True	False	False	...	False	False	False	False	False	False	False
999	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False

1000 rows × 664 columns

```
In [11]: dfs=fpgrowth(df, min_support=0.005, use_colnames=True, max_len=None, verbose=0)
dfs
```

Out[11]:

	support	itemsets
0	0.303	(Action)
1	0.259	(Adventure)
2	0.120	(Sci-Fi)
3	0.106	(Mystery)
4	0.008	(Ridley Scott)
...

-Destek değerini 0,005 alıyoruz.

```
In [13]: association_rules(dfs, metric="confidence", min_threshold=0.7)
```

Out[13]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Sci-Fi, Adventure)	(Action)	0.059	0.303	0.050	0.847458	2.796890	0.032123	4.569222	0.682742
1	(Mystery, Crime)	(Drama)	0.029	0.513	0.023	0.793103	1.546011	0.008123	2.353833	0.363722
2	(Mystery, Romance)	(Drama)	0.008	0.513	0.006	0.750000	1.461988	0.001896	1.948000	0.318548
3	(Ridley Scott)	(Drama)	0.008	0.513	0.007	0.875000	1.705653	0.002896	3.896000	0.417051
4	(Family, Fantasy)	(Adventure)	0.017	0.259	0.014	0.823529	3.179650	0.009597	4.199000	0.697355
5	(Animation)	(Adventure)	0.049	0.259	0.038	0.775510	2.994248	0.025309	3.300818	0.700343
6	(Animation, Adventure)	(Comedy)	0.038	0.279	0.027	0.710526	2.546689	0.016398	2.490727	0.631324
7	(Comedy, Animation)	(Adventure)	0.034	0.259	0.027	0.794118	3.066091	0.018194	3.599143	0.697569
8	(Animation, Action)	(Adventure)	0.010	0.259	0.009	0.900000	3.474903	0.006410	7.410000	0.719416
9	(Fantasy, Romance)	(Drama)	0.006	0.513	0.005	0.833333	1.624431	0.001922	2.922000	0.386720

-Güven değerleri rapidminer'dan elde edilen değerler ile uyuyor.

5.SONUÇLAR

İlk çalışma sonucu:

AssociationRules

```
Association Rules
[Kevin James] --> [Adam Sandler] (confidence: 1.000)
[Rupert Grint] --> [Daniel Radcliffe] (confidence: 1.000)
[Rupert Grint] --> [Emma Watson] (confidence: 1.000)
[Paul Walker] --> [Vin Diesel] (confidence: 1.000)
[Jennifer Lawrence, Liam Hemsworth] --> [Josh Hutcherson] (confidence: 1.000)
[Jennifer Lawrence, Josh Hutcherson] --> [Liam Hemsworth] (confidence: 1.000)
[Liam Hemsworth, Josh Hutcherson] --> [Jennifer Lawrence] (confidence: 1.000)
[Daniel Radcliffe, Emma Watson] --> [Rupert Grint] (confidence: 1.000)
[Rupert Grint] --> [Daniel Radcliffe, Emma Watson] (confidence: 1.000)
[Daniel Radcliffe, Rupert Grint] --> [Emma Watson] (confidence: 1.000)
[Emma Watson, Rupert Grint] --> [Daniel Radcliffe] (confidence: 1.000)
```

İkinci çalışma sonucu:

AssociationRules

```
Association Rules
[Animation] --> [Adventure] (confidence: 0.776)
[Adventure, Sci-Fi] --> [Action] (confidence: 0.847)
[Biography] --> [Drama] (confidence: 0.914)
```


Üçüncü çalışma sonucu:

AssociationRules

```
Association Rules
[Adventure, Animation] --> [Comedy] (confidence: 0.711)
[Action, Biography] --> [Drama] (confidence: 0.727)
[Crime, Biography] --> [Drama] (confidence: 0.750)
[Romance, Mystery] --> [Drama] (confidence: 0.750)
[Adventure, Crime] --> [Action] (confidence: 0.750)
[War] --> [Drama] (confidence: 0.769)
[Animation] --> [Adventure] (confidence: 0.776)
[Crime, Mystery] --> [Drama] (confidence: 0.793)
[Comedy, Animation] --> [Adventure] (confidence: 0.794)
[Music] --> [Drama] (confidence: 0.812)
[Fantasy, Family] --> [Adventure] (confidence: 0.824)
[Michael Bay] --> [Action] (confidence: 0.833)
[David Yates] --> [Fantasy] (confidence: 0.833)
[Romance, Fantasy] --> [Drama] (confidence: 0.833)
[David Yates] --> [Adventure, Fantasy] (confidence: 0.833)
[Adventure, David Yates] --> [Fantasy] (confidence: 0.833)
[Adventure, Sci-Fi] --> [Action] (confidence: 0.847)
[Ridley Scott] --> [Drama] (confidence: 0.875)
[Comedy, Biography] --> [Drama] (confidence: 0.875)
[Sport] --> [Drama] (confidence: 0.882)
[Action, Animation] --> [Adventure] (confidence: 0.900)
[Biography] --> [Drama] (confidence: 0.914)
[Biography, History] --> [Drama] (confidence: 0.933)
[History] --> [Drama] (confidence: 0.966)
[David Fincher] --> [Drama] (confidence: 1.000)
[Paul W.S. Anderson] --> [Action] (confidence: 1.000)
[Justin Lin] --> [Action] (confidence: 1.000)
[Zack Snyder] --> [Action] (confidence: 1.000)
[David Yates] --> [Adventure] (confidence: 1.000)
[Romance, Biography] --> [Drama] (confidence: 1.000)
[Biography, Sport] --> [Drama] (confidence: 1.000)
[Fantasy, David Yates] --> [Adventure] (confidence: 1.000)
```

Bu çalışmada market sepet analizinin, yönetmenlerin oyuncu seçiminde, senaristlerin tür seçiminde veya prodüksiyonun yönetmeni seçmesinde kullanılmasının yararlı olup olmayacağı incelenmiştir. **Market sepet analizi sonucunda bazı güven değerleri elde edilsede sonuç elde etmek için, çalışmalarda alınan destek değerleri çok küçük olduğundan market sepet analizinin film endüstrisinde kullanılamayacağı söylenebilir.**

6.ÖNERİ VE TARTIŞMALAR

Bir oyuncunun çalışma hayatının ortalama 50-60 yıl kabul edersek 10 yıllık 1000 film verisi içeren bir veri seti böyle bir çalışma için eksik kalıyor olabilir. Daha büyük bir veri seti ile daha iyi sonuçlar elde edilebilmesi mümkün. Ama böyle duygu, yetenek ve insan ilişkisi gerektiren bir sektörde, yapılan çalışmada da görüldüğü üzere, bu yöntem ile kararlar alınması pek mümkün değil.

KAYNAKLAR

- <https://brcubayram.medium.com/apriori-algoritmas%C4%B1-ile-birliktelik-analizi-ve-market-sepet-analizi-%C3%B6rne%C4%9Fi-1a454b05827>
- https://tr.wikipedia.org/wiki/Film_end%C3%BCstrisi
- <https://arts.artuklu.edu.tr/en/download/article-file/2411299>
- https://tr.wikipedia.org/wiki/Film_ekibi
- [https://suatatan.wordpress.com/2019/03/28/market-sepeti-analizi-nedir/-](https://suatatan.wordpress.com/2019/03/28/market-sepeti-analizi-nedir/)
- <https://acikkaynak.bilecik.edu.tr/xmlui/bitstream/handle/11552/1498/10261472.pdf?sequence=1&isAllowed=y>

KONTROL LİSTESİ

	EVET /HAYIR
Çalışmanızı hem bir araç kullanarak hem de kod yazarak mı hazırladınız?	Evet
Raporunuzu şablonda belirtildiği gibi hazırladınız mı?	Evet
Çalışmanızın sonuçlarını (print screen) rapora eklediniz mi?	Evet
Rapor dosyanızı şablondaki gibi yeniden adlandırdınız mı?	Evet
Tek bir zip/rar dosyası içinde aşağıdaki materyalleri sisteme yüklediniz mi? <ul style="list-style-type: none">• Proje Raporu• Projenizde kullandığınız veri seti• Kodlar	Evet