



T.C.
NECMETTİN ERBAKAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**E-TİCARET SİTELERİNİN SATIŞLARININ
ARTIRILMASI İÇİN BULUT BİLİŞİM TABANLI
YAPAY ÖĞRENME**

AHMET AYAN

YÜKSEK LİSANS TEZİ

Endüstri Mühendisliği Ana Bilim Dalı

**Kasım - 2018
KONYA
Her Hakkı Saklıdır**

TEZ KABUL VE ONAYI

Ahmet Ayan tarafından hazırlanan “E-Ticaret Sitelerinin Satışlarının Artırılması için Bulut Bilişim Tabanlı Yapay Öğrenme” adlı tez çalışması 9/11/2018 tarihinde aşağıdaki jüri tarafından oy birliği ile Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Ana Bilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri**İmza****Başkan**

Prof. Dr. Sabri KOÇER

.....

Danışman

Dr. Öğr. Üyesi Ali Osman ÇIBIKDİKEN

.....

Üye

Dr. Öğr. Üyesi Muhammed Abdullah BÜLBÜL

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Ahmet AVCI
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Ahmet Ayan

09.11.2018

ÖZET**YÜKSEK LİSANS TEZİ****E-TİCARET SİTELERİNİN SATIŞLARININ ARTIRILMASI İÇİN BULUT
BİLİŞİM TABANLI YAPAY ÖĞRENME****Ahmet Ayan****Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü
Endüstri Mühendisliği Ana Bilim Dalı****Danışman: Dr. Öğr. Üyesi Ali Osman ÇIBIKDİKEN****2018, 48 Sayfa****Jüri****Dr. Öğr. Üyesi Ali Osman ÇIBIKDİKEN****Prof. Dr. Sabri KOÇER****Dr. Öğr. Üyesi Muhammed Abdullah BÜLBÜL**

E-ticaret sitelerinde ürünler ile ilgili öneriler yapılması kullanıcı tercihlerinin belirlenmesinde önemli bir etkidir. Kullanıcılar bir alışveriş sitesini ziyaret ederken hızlı bir şekilde karar vermek ister. Kullanıcının daha önceki alışverişlerine göre seçilen ürünle ilişkili ürünleri gösterdikleri öneri sistemleri bulunmaktadır. Öneri sistemleri ile kullanıcıya "Bu ürünü kullananlar, aynı zamanda bu ürünü de aldı." şeklinde farklı ürün önerileri sunulabilmektedir. Böylece e-ticaret sitelerinin satışlarının artırılması hedeflenmektedir. Bu tez çalışmasında, e-ticaret sitesi için bir bulut bilişim platformu üzerinde yapay öğrenme algoritmaları kullanarak tavsiye sistemi çalışması yapılmıştır. Çalışma kapsamında; öncelikli olarak e-ticaret ile ilgili web sitelerinden satış verilerinin alınması gerçekleştirilmiştir. Alınan bu veriler bulut bilişim altyapısı üzerinde bulunan yapay öğrenme algoritmalarından içerik tabanlı öneri sistemi (content-based), işbirlikçi öneri sistemi (collaborative) ve hibrid öneri sistemi (hybrid) ayrı ayrı uygulanarak bir tavsiye modeli elde edilmiştir. Yapay öğrenme algoritmalarından elde edilen sonuçlar karşılaştırılarak en iyi sonucu sunan algoritma modeli belirlenmiştir. Bu model bir web servis arayüzüne dönüştürülerek bir e-ticaret sitesinde kullanılabilir şekilde hazırlanmıştır.

Anahtar Kelimeler: Bulut bilişim, e-ticaret, hibrid, içerik tabanlı, işbirlikçi, öneri sistemleri, yapay öğrenme

ABSTRACT**MS THESIS****MACHINE LEARNING ON CLOUD COMPUTING PLATFORM FOR
INCREASING SALES OF E-COMMERCE SITES****THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
NECMETTIN ERBAKAN UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE
IN INDUSTRIAL ENGINEERING****Ahmet Ayan****Advisor:** Assistant Prof. Dr. Ali Osman ÇIBIKDİKEN**2018, 48 Pages****Jury**

Assistant Prof. Dr. Ali Osman ÇIBIKDİKEN

Prof. Dr. Sabri KOÇER

Assistant Prof. Dr. Muhammed Abdullah BÜLBÜL

Making suggestions about products on e-commerce sites is an important factor in determining user preferences. When visiting a shopping site, users want to decide quickly. There are suggestion systems where the user shows the products associated with the product selected according to their previous purchases. Recommendation systems to the user "users of this product, also bought this product." different product recommendations can be offered. Thus, it is aimed to increase the sales of e-commerce sites. In this thesis, a recommendation system study was performed using artificial learning algorithms on a cloud computing platform for the e-commerce site. The scope of work, the sales data were obtained from e-commerce related websites firstly. This data is obtained by applying artificial learning algorithms in cloud computing infrastructure, and a recommendation model has been obtained. These data were obtained from artificial learning algorithms on cloud computing infrastructure by using content based suggestion system, collaborative suggestion system and hybrid suggestion system separately. The model is presented by comparing the results obtained from artificial learning algorithms. This model is transformed into a web service interface and can be used on an e-commerce site.

Keywords: Cloud computing, content-based recommender, collaborative recommender, e-commerce, hybrid recommender, machine learning, recommender system

ÖNSÖZ

Günümüzde e-ticaret sitelerinde kullanıcılara yapılan öneriler, sitelerdeki satışların artmasına oldukça etki etmektedir. Tez çalışmasında bu önerilerin bir e-ticaret sitesinde nasıl uygulandığı üzerinde durulmuştur. Çalışmada, e-ticaret sitesi için bir bulut bilişim platformu üzerinde yapay öğrenme ile site müşterilerine inceledikleri ürün bilgisine göre yeni tavsiyelerde bulunması amaçlanmıştır.

Çalışmam boyunca yardımlarını esirgemeyen danışmanım Dr. Öğr. Üyesi Ali Osman Çıbıkdiken'e teşekkürü bir borç bilirim.

Ahmet Ayan
KONYA - 2018

İÇİNDEKİLER

ÖZET	3
ABSTRACT	4
ÖNSÖZ	5
İÇİNDEKİLER	6
KISALTMALAR	7
1. GİRİŞ	8
1.1. Yaygın Kullanılan Öneri Sistemleri	8
1.2. Öneri Sistemlerinin Faydaları	11
1.3. E-Ticaret Nedir?	12
1.4. Çalışmanın Amacı	13
1.5. Çalışmanın Önemi	13
2. KAYNAK ARAŞTIRMASI	14
3. MATERYAL VE YÖNTEM	19
3.1. Microsoft Azure Makine Öğrenmesi Platformu	19
3.2. ML Studio’da Model Geliştirmek	22
3.3. ML Studio Train Matchbox Recommender	25
3.4. ML Studio’da Hibrid Filtreleme Tekniğini İçeren Train Matchbox Recommender Modülünü Kullanarak Ürün Önerisi Sunmak	26
3.5. Veri Seti	26
3.6. ML Studio’da İşbirlikçi Filtreleme Yöntemi ile Ürün Önerisi Sunmak	33
3.6.1. R Programlama Dili ve ML Studio’da Execute R Script Modülü	33
3.6.2. R Studio’da İşbirlikçi Öneri Modelinin Geliştirilmesi ve ML Studio’da Kullanılması	35
3.7. R Studio’da İçerik Tabanlı Öneri Modelinin Geliştirilmesi ve ML Studio’da Kullanılması	38
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA	39
4.1. Öneri Sisteminin Çalıştırılması	40
4.1.1. Testlerden Ürün Önerileri	41
4.2. Web Servis ile Tahminleme Modeli İçin Geliştirilen Arayüz	44
5. SONUÇ VE ÖNERİLER	48
KAYNAKLAR	51
ÖZGEÇMİŞ	55

KISALTMALAR

AWS	: Amazon Web Servisi (Amazon Web Services)
LA	: Otomata Öğrenme (Learning Automata)
LASA	: Önerilen Öğrenim Otomata Tabanlı Mercek Analiz Sistemi (Learning Automata-Based Lens System Analysis)
TS	: Tavsiye Sistemi
AI	: Yapay Zeka (Artificial Intelligent)
CLR	: Ortak Dil Çalışma Platformu (Common Language Runtime)
ML	: Makina Öğrenmesi (Machine Learning)
MAE	: Ortalama Mutlak Hata (Mean Absolute Error)
RMSE	: Kök Ortalama Karesi Alınmış Hata (Root Mean Squared Error)
RAE	: Göreli Mutlak Hata (<i>Relative Absolute Error</i>)
RSE	: Göreli Karesi Alınmış Hata (<i>Relative Squared Error</i>)
NIEKK	: Normalleştirilmiş İskonto Edilmiş Kümülatif Kazanç
NDCG	: Normalized Discounted Cumulative Gain

1. GİRİŞ

Öneri sistemleri günümüzde tahmini analitik yöntemlerin en güçlü ve en yaygın uygulamalarından biri haline gelmiştir. Günümüzde önerilebilecek ürünlerdeki çeşitliliğin artması ve bunun yanında kullanıcıların geçmişte hangi ürünleri kullandıkları bilgisinin tutulması öneri sistemlerinin önemini artırmaktadır. Genel olarak kullanıcıların her birinin farklı beğenileri olacağından doğru kullanıcıya doğru ürün önerisi yapılması önemli bir problem olarak karşımıza çıkmaktadır. Genel olarak kullanılan ürünlere benzer ürünlerin önerilmesi şeklinde bir adım izlense de, kullanılan ürünlerle ilgili olmayanlar arasından da bir öneri yapılıp yapılamayacağı da incelenmelidir (Marx ve ark., 2010). Öneri sistemlerinde genel olarak kullanıcıların ürünlere verdikleri puanları içeren, kullanıcı-ürün matrisleri kullanılmaktadır. Bu sistemler en popüler web uygulamalarının çoğunda bulunabilir ve pek çok tüketici platformunda başarı için önemli hale gelmiştir.

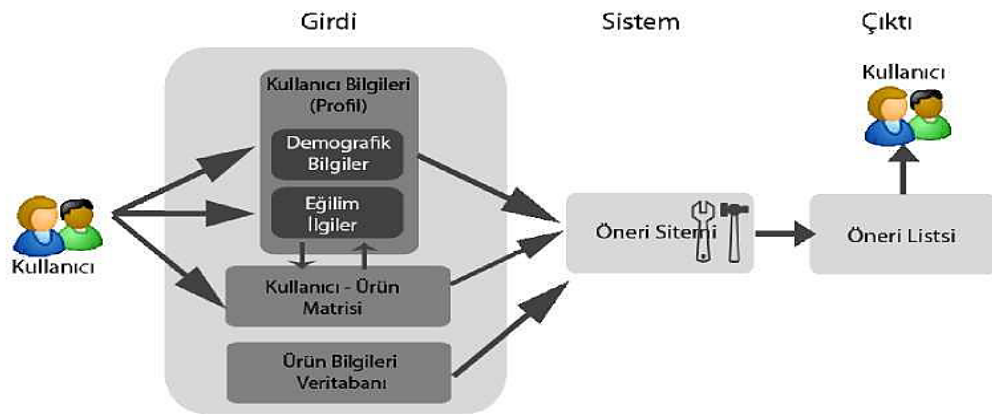
1.1. Yaygın Kullanılan Öneri Sistemleri

İçerik tabanlı (Content-Based) öneri sistemleri her kullanıcıyı tek tek ele almaktadır ve kullanılan içerik tabanlı öneri sisteminde ürün bilgileri ve kullanıcıların önceki ilgi alanındaki bilgilerinden elde edilen dolaylı değerlendirme verileri kullanılmaktadır. İşbirlikçi (Collaborative) öneri sistemi ise bir kullanıcının ilgi alanı bilgileriyle birlikte ona ürün önerisi olarak yakın olan diğer kullanıcıların bilgilerinden de faydalanır (Sarwar ve ark., 2001). Ek olarak her iki sistem birleştirilerek hibrit (Hybrid) bir öneri sistemi oluşturulur.

Puan verilmemiş olsa da, kullanıcının hangi ürünlere baktığı gibi veriler kullanılarak kullanıcının bir ürünle ilgili düşüncesi hakkında bilgi sağlanabilir. İşbirlikçi (Collaborative) öneri sistemlerinde kullanıcı-ürün puanlama matrisleri ile ürünlerin kendi aralarındaki ve kullanıcıların kendi aralarındaki benzerlikler hesaplanabilmekte ve öneri için kullanılabilir. Fakat puanlama matrisleri genelde seyrek yapıda olan matrislerdir. Bütün kullanıcıların tüm ürünlere puan vermiş olmaları çok gerçekleşmesi

bir durum değildir. Matris ayrıştırma metotları ile kullanıcı ve içerik daha az boyutlu uzaylarda gösterilerek benzerliklerin gerçeğe yakın hesaplanması sağlanmaktadır.

Öneri sistemleri, hedeflenmiş/belirlenmiş kullanıcıların en çok ilgisini çekebilecek “nesne” lerin önerilmesini sağlayan yazılım (software) araçları ve teknikleridir. Öneri sistemleri özellikle ürün yelpazesinin çok geniş olduğu sistemlerde ya da süreçlerde müşteriye ürün seçimi konusunda yardımcı olurlar. Milyonlarca ürüne sahip bir sitede müşteriler bu ürünlerin tamamını inceleyecek ne sabra ne de zamana sahiptirler. Bu yüzden öneri sistemleri kişiselleştirilmiş öneriler sunarak müşterinin karar verme problemini çözmeye çalışırlar. Böylece müşteri zaman kaybetmeden istediği ürüne ulaşabilir (Ricci ve ark., 2015). Şekil 1.1 öneri sistemlerinin temel akış şeması gösterilmiştir.



Şekil 1.1. Öneri sistemlerinin temel akış şeması (Ricci ve ark., 2015)

Öneri sistemlerinde kullanılan veri setleri ve bu veri setlerindeki parametrelerin sistemde ne ifade ettiğine dair tanımlar ve işlevlerin açıklaması aşağıda belirtilmiştir (Ricci ve ark., 2015):

- **Kullanıcı:** Karakterleri, ilgileri ve istekleri farklı birçok kullanıcıyı temsil eder.
Kullanıcı Profili: Kullanıcıdan ilk kayıt esnasında istenilen veya kullanıcının zaman içerisindeki işlemleri baz alınarak çeşitli yöntemler ile elde edilen, kullanıcının beğenilerini, isteklerini ve kullanıcının dinamik özelliklerini yansıtan verilerdir.

- **Kullanıcı- Ürün Matrisi:** Genellikle çok seyrek olan ve kullanıcıların sistemdeki içerik hakkındaki fikirlerini yansıtan matristir (Şekil 1.2). Kullanıcı-Ürün matrisi kullanıcı tarafından doldurulur ve bu matristeki bilgi kullanıcıdan değişik biçimde istenebilir:
 - Belirli bir sayı aralığında sayısal bir değerlendirme şeklinde istenebilir.
 - 1-5 yıldız (amazon.com)
 - 1-10 yıldız (imdb.com)
 - Sıralı sözlü bir değerlendirme şeklinde istenebilir.
 - Çok iyi, iyi, orta, kötü, çok kötü
 - İkili formatta istenebilir.
(Beğendim-beğenmedim, Onaylıyorum-onaylamıyorum, İyi-kötü gibi)
 - Tekil değerlendirme: Kullanıcı bir ürünü incelemiş veya satın almış ise, kullanıcının bu ürün ile alakalı olumlu düşünceleri vardır. Tam tersi durumda eğer kullanıcı ürüne hiç bakmamış ise ürün ile ilgilenmiyor değerlendirmesi yapılabilir.

		Ürünler					
		1	2	...	<i>i</i>	...	<i>m</i>
Kullanıcılar	1	5	3		1	2	
	2		2				4
	:			5			
	<i>u</i>	3	4		2	1	
	:					4	
	<i>n</i>			3	2		
<i>a</i>		3	5		?	1	

Şekil 1.2. Kullanıcı-ürün değerlendirme matrisi (Sammur ve ark., 2017)

Bilgi edinme şekli de; kullanıcıdan direk alınan ve kullanıcıdan dolaylı yoldan elde edilen bilgi olmak üzere iki türdür:

- **Açık Veriler:** Kullanıcının doğrudan kendisinin yaptığı yorumlardır. Kullanıcı bir ürün hakkında bir yorum yapmış veya bir derecelendirmede bulunmuşsa bu tür verilere açık veriler denir.
- **Kapalı Veriler:** Kapalı veriler aslında bir çeşit yorumlamadır. Kullanıcının sistem içerisindeki hareketlerini analiz edilerek bu kullanıcının neye ilgi duyduğunu ve nelerden hoşnut olmadığını anlamaya çalışmaktır. Örneğin kullanıcı online olarak sürekli belirli bir tür film izliyorsa kullanıcı bu filmler hakkında hiç yorum yapmasa veya bu filmleri hiç değerlendirmese bile kullanıcının bu tür filmlere ilgisinin olduğu fikrine varılabilir (Sammut ve ark., 2017).

1.2. Öneri Sistemlerinin Faydaları

Öneri sistemleri kullanıcılara ürünlerin sunulduğu platformlarda birçok yarar sağlamaktadır. Bunlardan birkaçı aşağıda belirtilmiştir :

- **Satılan ürün sayısını artırır:** Bu özellik ticari bir site için bu sistemlerin en önemli faydasıdır. Önerilen ürünün çevrimiçi olan müşterinin hem ihtiyacı hem de istediği bir ürün olma ihtimali yüksek olacağından müşterinin bu ürünü alma ihtimali de yükselir ve dolayısıyla bu da satışlara pozitif yönde etki eder.
- **Daha çeşitli ürün satışı olur:** Ticari sitelerin en büyük sorunlarından bir tanesi sürekli aynı müşterilere benzer ürünler satma problemidir. Öneri sistemleri müşterilere ilgi duydukları ürünlerin yanında, bu ürünlerden tamamen farklı ürünler de önererek bu sorunu büyük oranda çözmüştür.
- **Müşterinin güvenini artırır:** Müşteri sistemin önerdiği ürünleri ilginç, yerinde ve kendine yakın buldukça zaman içerisinde müşterinin sisteme ve firmaya olan güveni de artacaktır.
- **Müşterinin sadakatini artırır:** Bir mağazaya gittiğinizde satıcı sizi tanıyor ve sizin istediğiniz şeyleri aşağı yukarı tahmin edebiliyorsa o mağazaya gitmekten hoşlanırsınız ve bir ürüne ihtiyacınız olduğunda ilk aklınıza gelen yer yine o mağaza olur. Bu örnekten yola çıkarak eğer bir e-ticaret sitesi müşterisini çevrimiçi olduğunda tanıyor, kişisel tercihlerini biliyorsa, müşterinin istediği

doğrultuda ürünler önerebiliyorsa; yani kısaca müşteriye değerli olduğunu hissettirebiliyorsa müşteri ihtiyacı olduğunda tekrar bu siteyi kullanacaktır. Tavsiye sistemleri e-ticaret, e-iş, e-öğrenme, e-turizm ve e-kaynak servisleri gibi birçok alanda kullanılabilir. (Jannach ve ark.,2011 ; Ricci ve ark., 2015).

1.3. E-Ticaret Nedir?

E-Ticaret internet ortamında insanların belirli sistemler üzerinde gerçek ya da sanal ürünleri para karşılığında satın alması işlemidir. Yani e-ticaret, ticaretin elektronik ortama aktarılmış halidir.

Elektronik Ticaret veya "E-Ticaret" anlam olarak genellikle internet ortamında mal veya hizmet satın alma işlemi olarak tanımlanmaktadır. Özellikle günümüz dünyasında her web sitesi üzerinde sanal mağazalar veya sanal alışveriş merkezleri görülebilmektedir. Bunun yanı sıra sunulan ürün için talep oluşturmak (pazarlama), müşteri desteği sağlamak (satışın bütün aşamalarında), ticari kurumlar ile müşterileri arasındaki ticari ve lojistik iletişimi yine küresel bilgisayar ağları aracılığıyla sağlamaktır. Bunun yanı sıra şirketlerin yönetim akışını kolaylaştıran sistemlerdir.

E-Ticaretin başlıca yararları şu şekilde ifade edilebilir (Laudon ve ark., 2013):

- Şirketlerin internet ağı üzerinden müşterilerine bilgi sunabileceği 7/24 açık bir sistem oluşur.
- Elektronik ortamda veriler kısa zamanda aktarıldığı için zamandan tasarruf sağlanır.
- Müşteriye fiziki boyutlar olmadığı için daha fazla ürün seçeneği sunulabilmektedir.
- Mağazanın internet ortama aktarılması, maddi yükü azaltır.
- Stok, maliyet bilgileri internet ortamında bulunduğu için tedarikçi firma stok yapmak zorunda kalmaz. Deposunda ne kadar ürün bulunduğunu ve bu ürünlerden kaç tane sattığını bilir.
- Tedarikçinin büyük bir mağazaya ihtiyaç duymadan internet üzerinden online alışveriş merkezini oluşturmasına imkân sağlar.

- İnternet gibi küresel bir ortamda çok fazla piyasa dalgalanması olmayacağı için ticari zarar oluşmaz.

1.4. Çalışmanın Amacı

Tez çalışmasında, e-ticaret sitesi için bir bulut bilişim platformu üzerinde yapay öğrenme ile site müşterilerine inceledikleri ürün bilgisine göre yeni tavsiyelerde bulunarak, müşterinin ilgisinin çekilmesi ile satış oranının artırılması amaçlanmıştır. Yapılan literatür taramasında bu tarz çalışmaların kullanımının henüz yeni olduğu görülmektedir. Çalışma kapsamında; öncelikli olarak e-ticaret ile ilgili web sitelerinden çeşitli verilerin alınması gerçekleştirilmiştir. Daha sonra alınan bu verilere bir bulut platformu üzerinde (bu çalışmada örnek olarak Microsoft Azure seçilmiştir) yapay öğrenme algoritmaları uygulanarak bir öneri modeli elde edilmiştir. Bu model bir web servisi yapısına dönüştürülerek herhangi bir web sitesinde kullanılabilir hale getirilmiştir.

1.5. Çalışmanın Önemi

E-ticaret sitelerindeki tavsiyeler kullanıcılar için büyük bir önem arz etmektedir. Öneriler kullanıcıların sitede gezinirken daha hızlı karar vermesinde güçlü bir araç haline gelmiştir ve satışları artırmak için iyi bir teknik olabilir. Muhtemelen, bir e-ticaret perakende sitesinde baktığınız ürünle birlikte sıklıkla bir araya gelen öğeleri gösterdikleri tavsiyeleri görürsünüz. Tez çalışmasında bir e-ticaret sitesi verileri üzerinde öneri sistemi uygulanmıştır ve kullanıcıya "Bu ürünü kullananlar, aynı zamanda bu ürünü de aldı." şeklinde farklı ürün önerileri sunulmuştur. Bu sayede e-ticaret sitelerinin satış oranlarına ekonomik anlamda büyük katkı sağlanması hedeflenmiştir.

2. KAYNAK ARAŞTIRMASI

Türkçe “bulut bilişim” olarak kullanılan “cloud computing” teriminin nereden çıktığı tam olarak bilinmemektedir. Fakat bilişim ve iletişim sistemlerinin yapısındaki ağları belirtmek için kullanılan bulut benzeri çizimlerden türetilmiş gibi görülmektedir. Bulut simgesi 1994 yılına gelinceye kadar İnternet’i sembolize etmektedir.

1990’larda önceden noktadan noktaya özel veri devreleri sağlayan telekomünikasyon firmaları, artık daha iyi hizmet kalitesine ve daha az masrafa sahip olan “Sanal Özel Ağ” hizmetlerini sağlamaya başlamışlardır. Kullanımı dengelemek için uygun gördüklerinde trafik anahtarlama ile toplam ağ bant genişliğini daha aktif kullanmayı başardılar. Bulut simgesi sağlayıcıların sorumlulukları ile kullanıcıların sorumlulukları arasındaki sınırı belirtmek için kullanıldı. Bulut bilişim bu sınırı sunuculardan ağ altyapısına kadar kapsayacak şekilde genişletmiştir. Bulut Bilişim denildiğinde tüm dünyada akla ilk gelen iki isim Microsoft Azure ve Amazon AWS dir.

Gupta ve ark. (2016) yaygın olarak kullanılan Microsoft Azure ve Amazon AWS bulut bilişim altyapıları ile ilgili karşılaştırmalar yaparak hangisinin kullanılmasının daha uygun olacağını belirlemişlerdir.

Barnes (2015) bir bulut altyapısı olan Microsoft Azure’da yer alan makine öğrenmesi ile ilgili çalışmalar yapmıştır.

Grounound ve ark. (2014) Microsoft Azure bulut altyapısı üzerinde makine öğrenmesinin nasıl yapılacağı araştırmışlardır.

Bobadilla ve ark. (2011) işbirlikçi (collaborative) filtreleme öneri sistemleri için bir framework geliştirmişlerdir. Önerilen çalışma, öneri yeniliğini, işbirlikçi filtreleme yaklaşımı için denklemleri ve söz konusu ölçümleri ve denklemleri kullanan bir çatıyı (framework) değerlendirmek için ölçümler sağlamaktadır.

Bobadilla ve ark. (2011) kullanıcılar arasında bir benzerlik hesaplayarak, genetik algoritmalar kullanarak işbirlikçi filtreleme öneri sistemi performansını geliştirmek için bir başka çalışma sunmuşlardır.

Mild ve Reutterer (2003), yalnızca ikili müşteri bilgilerinin mevcut olduğu durumlar için ortak bir filtreleme (colloborative) yaklaşımı sunmuştur. Önerilen yaklaşım, sepet verisinde yaygın olan verilerin seçilmesi veya seçilmemesi gibi ikili değerler için geçerlidir.

Wang ve ark. (2004) kozmetik işi için kişiselleştirilmiş bir tavsiye sistemi önermiştir. Araştırmacılar içerik tabanlı (content-based), işbirlikçi (collaborative) filtreleme ve veri madenciliği yaklaşımlarını kullanarak bir öneri sistemi geliştirmişlerdir. Müşterilerin ürünlere olan ilgilerini değerlendirmek için bir puanlama yaklaşımı da araştırmacılar tarafından önerilmiştir.

Xiang ve ark. (2007) çevrimiçi öneri sistemleri farklı bir bakış açısından incelenmiştir. Araştırmacılar, özellikle online restoran hizmetlerinde etkili bir öneri sistemi geliştirmek için dilin önemli bir faktör olduğunu göstermeyi amaçlamışlardır. Yapılan çalışma bu tür sistemler için öneriler sağlamıştır.

Diez ve ark. (2008) benzer tercihleri paylaşan kişilerin kümelenmesini keşfetme yaklaşımı getirmiştir. Bu kümeleri oluşturmak için, sıralama işlevleri bireylerin tercih yorum kümelerinden oluşturulmuştur. Araştırmacılar, bu kümeleri insanları farklı pazar segmentlerine haritalamak için kullanmayı amaçlamışlardır.

Bir gruptaki her bireyin öneri memnuniyetini artırmayı amaçlayan Kim ve ark. (2010) bir öneri sistemi geliştirilmiştir. Bu hedefe ulaşmak için araştırmacılar, başlangıç önerileri oluşturmak için ortak bir filtreleme yaklaşımı (colloborative) kullandılar ve daha sonra, her bir grup üyesinin tercihlerine göre ilgili olmayan öğeleri sistemden kaldırmışlardır.

Sarwar ve ark. (2000) yaptığı çalışmada işbirlikçi filtreleme tabanlı öneri sistemlerinin üç aşamada müşteriye tavsiye verdiği görülmüştür. Bunlardan ilki kullanıcının derecelendirdiği ürünlere bakılarak kullanıcı profilinin oluşturulmasıdır. İkinci aşama sistemin makine öğrenmesi veya istatistiksel teknikler kullanarak benzer davranışlara sahip komşuluklar olarak adlandırılan kullanıcı kümelerini inşa etmesidir. Son aşama ise tahmin ve öneri hesaplama sürecidir.

Choi ve ark. (2006) makalesinde nesnelerin müşteri açısından sağladıkları fayda değerlerinin karşılaştırılmasıyla ürünlerin benzerliklerini hesaplayan bir model önerisi bulunmaktadır. Bu model kapsamında çok kriterli karar verme tekniği ile nesnelerin müşteriye olabilecek minimum ve maksimum faydası müşteri geri bildirimini vasıtasıyla hesaplanmaktadır. Bu noktada nesnelerin özelliklerinin önemi (ağırlığı) ve bu özelliğin marjinal faydası dikkate alınmaktadır. Sonrasında nesnelerin ortalama faydalarının birbirine yakınlığına ve fayda aralıklarının birbirinin ne kadar içinde olduğuna (yukarıda bahsedilen minimum/maksimum fayda aralığı) bakılarak nesnelerin arasında benzerlikler hesaplanmakta en benzer olan k adet ürün müşteriye önerilmektedir. Makale kapsamında deneysel bir çalışma yapılmakta ve sonuçlar öklidyen uzaklık (Euclidian Distance) benzerlik ölçüsü ile kıyaslanmaktadır. Deneye katılan kullanıcıların önerilen nesnelerin benzerliğini ve tatmin düzeylerini skorlaması istenmektedir. Sonuç itibarıyla kurulan model daha başarılı olduğu gözlemlenmektedir.

Wang ve ark. (2008) bir web sitesini gezen kullanıcılara ilgilerini çekebilecek yeni sayfaları önermek için hazırlanmış bir modeli kapsamaktadır. Buna göre siteyi daha önceden gezen kullanıcıların gezinme verileri analiz edilerek modele girdi olarak verilmektedir. Model kapsamında kullanıcıların gezinme adımları kümelemeye tabi tutulmaktadır. Bu kapsamda Ward Minimum Varyans (hierarchical agglomerative clustering) tekniği kullanılmıştır. Kümeler oluştuktan sonra her bir kümenin medoidi tespit edilmektedir. Kurulan model kapsamında, bu kümeleme çalışması çevrimdışı olarak yapılmaktadır. Daha sonra çevrimiçi olarak gezinen bir kullanıcı belirli bir gezinme dizisi büyüklüğüne ulaştıktan sonra, bu kullanıcının gezinme dizisinin daha önce oluşturulmuş kümelerin medoidi ile mesafesi (yakınlığı - distance) ölçülmektedir. En yakın mesafedeki küme, kullanıcının da potansiyel olarak dahil olduğu küme olarak kabul edilmektedir. Bu tespit sonrasında bu küme içerisindeki ziyaret edilebilecek sayfalar (düğümler/node) erişim frekansına göre çoktan aza doğru sıralanmakta ve en yüksek frekansa sahip N tanesi kullanıcıya tavsiye edilmektedir.

Davidson ve ark. (2010) dünyanın en popüler çevrimiçi video yayını yapan Youtube'un kullandığı video öneri sistemi üzerinde çalışmışlardır. Sistem kullanıcıların sitedeki etkinliklerine göre kişiselleştirilmiş video kümelerini önermektedir. Öneri

sisteminin karşılaştığı problemlerden ve kullanıcıya nasıl hitap öneri yapıldığından bahsedilmiştir. Ek olarak, yeni algoritmaları test etmek ve kullanmak için yapılan deney ve değerlendirme çerçevesi hakkında ayrıntılı bilgi vermişlerdir.

Zhang ve ark. (2005) makalesinde, B2B (işten-işe) e-ticarete öneri sisteminin uygulanması tartışılmıştır.. Öncelikle, öneri sisteminin B2B katılımcılarının işlemlerini nasıl kolaylaştırdığı incelenmiştir, daha sonra B2B iş uygulamaları ve iş zekası temelinde B2B e-ticaretin öneri sistemi için etkin bir sistem çerçevesi tasarlanmış; ve sonra, model bileşenlerini ve süreçlerini tanımlanmıştır; Sonunda, uygulamanın problemleri ele alınmıştır.

Linden ve ark. (2003), Amazon.com web mağazası üzerinde her müşteri için ayrı bir mağazanın olması gerektiğini belirtmiştir. Müşterinin ilgisine göre bu mağazadaki ürünler değişiklik gösterecektir.

Schafer ve ark. (2001) makalesinde tavsiye sistemlerinin e-ticaret sitelerinin satışları nasıl artırdığını incelemiş ve altı adet e-ticaret pazar lideri sitelerde öneri sistemlerini analiz etmişlerdir. Bu örneklere dayanarak, tüketicilerin ihtiyaç duyduğu girdiler, gerekli ek bilgi dahil olmak üzere, bir öneri sistemleri yapısı oluşturulmuştur. Veritabanından, tavsiyelerin tüketicilere sunulma şekilleri, tavsiyeleri oluşturmak için kullanılan teknolojiler ve tavsiyelerin kişiselleştirilme durumu incelenmiştir. Yaygın olarak kullanılan beş E-ticaret öneri uygulama modeli belirtilmiştir.

Işık ve ark. (2017) sahte hesapların e-ticaret ortamlarında oluşturdukları olumsuz etkilerin azaltılması için, kullanıcılar arasındaki ilişkileri analiz ederek diğer kullanıcılar üzerinde etkisi olan ve gerçekten güvenilir olduğu düşünülen kullanıcılar bularak bir tavsiye modeli oluşturmuşlardır. Böylece, güvenilir kullanıcıların düşüncelerinden yola çıkılarak e-ticaret ortamlarında kullanıcılara tavsiyelerde bulunan Tavsiye Sistemlerinin (TS) kalitesini artıracak bir tavsiye sistemi geliştirilmiştir.

Krishna ve ark. (2013) öğrenme otomasını kullanarak bir öneri sistemi geliştirmişler ve duygu analizi yapmışlardır. Öğrenme otoması analizi kullanılarak önerilen sistem tarafından üretilen tavsiye puanı optimize edilmiştir. Önerilen Öğrenim Otomata Tabanlı Mercek Analiz Sistemi, kullanıcıların mevcut konumlarının

yakınındaki yerleri, yerlerden gelen geri bildirimleri analiz ederek önerir ve buna bağı olarak hesaplama yapar. Öğrenme otomayı kullanarak, önerilen sistemin performansının artırılabilceğini düşünmüşlerdir.

Bu tez çalışmasında e-ticaret sitesi için bir bulut bilişim platformu üzerinde yapay öğrenme ile site müşterilerine inceledikleri ürün bilgisine göre yeni tavsiyelerde bulunarak, müşterinin ilgisinin çekilmesi ile satış oranının artırılması amaçlanmıştır. Tez çalışmasında öneri yöntemi olarak kullanılan içerik tabanlı öneri sistemi, işbirlikçi öneri sistemi ve hibrid öneri sisteminin ne olduğundan bahsedilmiştir. Üç farklı öneri yöntemi ile de Microsoft Azure platformunda öneri sistemi geliştirilmiştir. Çalışma sonucunda üç farklı öneri yönteminin öneri performansından bahsedilmiştir. Son olarak performans sonuçları (Mae, Rmse, Ncdg) ve analizleri gösterilmiştir.

3. MATERYAL VE YÖNTEM

Verilerin işlenmesi için tercih edilen bulut bilişim altyapısı ve üzerinde öneri sistemlerinde yaygın olarak kullanılan yapay öğrenme algoritmaları adım adım tanıtmıştır. Burada bulut altyapısı olarak Microsoft Azure platformu kullanılmıştır. Bu platformun tercih edilmesindeki sebep hem hazır öneri sistemi algoritmalarının bulunması hem de kullanıcının istediği bir algoritmayı platformun sağladığı dillerde (C#, R, Python vb.) çalıştırmaya imkan sağlamasıdır.

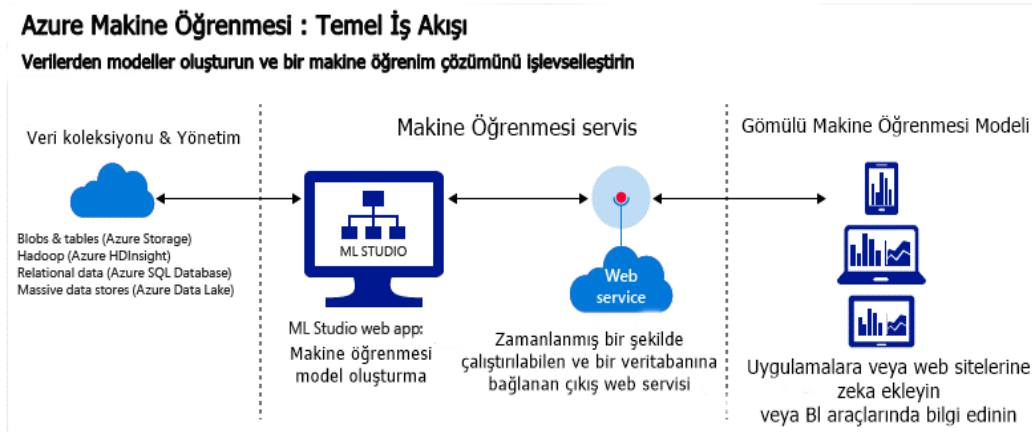
3.1. Microsoft Azure Makine Öğrenmesi Platformu

Microsoft Azure hem açık çevre ortamlarından hem de internetten (kendisi açık çevre için hazır hâle getirilmeyen platform olsa da) tüketilebilen çok çeşitli internet hizmetini sağlamakta olan bir bulut platformu hizmetidir.

Windows Azure hosting ortamına, ikili büyük nesnelere, queue, sürücüler ve ilişkisiz olmayan tablolar gibi yapılandırılmamış verileri depolamak gibi pek çok uygulamanın gerektirdiği diğer hizmetleri de sağlar. Uygulamalar, Azure Services Platform'un bir parçası olan diğer uygulamaları da kullanabilir. Bulut platformu üzerine uygulamaları dağıtmak için ASP.NET uygulama çatısını ve birleşmiş dağıtım metodlarını desteklerken, CLR için derlenmiş .NET Framework uygulamalarını hâlen çalıştırabiliyor. Microsoft Azure servislerinin en önemli özelliklerinden birisi de “ne kadar kullanırsan o kadar öde” yaklaşımıyla ücretlendirilmesidir. Örneğin bir sanal makineyi ayda sadece 100 saat çalıştırılmış olabilirsiniz. Bu durumda sadece 100 saatlik kullanım üzerinden ücretlendirme yapacaktır. Tüm Azure servislerinin benzer şekilde kendine özel hesaplama yöntemleri vardır ve fiyatlandırma ayrıntıları ilgili sayfada açıkça belirtilmiştir.

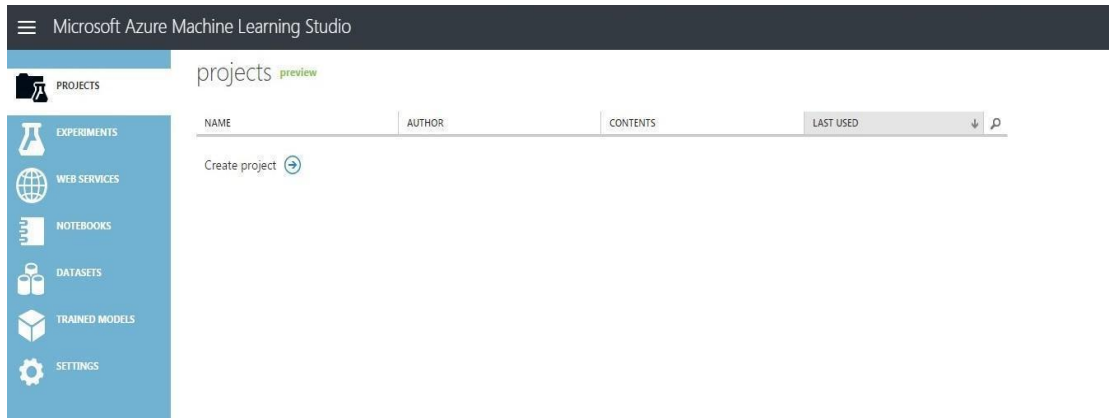
Windows Azure bulut platformu üzerinde sunulan hizmetlerden bir tanesi de Azure Makine Öğrenmesi hizmetidir. Azure Makine Öğrenmesi, tahmine dayalı modelleri analiz çözümleri olarak hızlı bir şekilde oluşturmayı ve dağıtmayı mümkün kılan bulut tabanlı ve tahmine dayalı analiz hizmetidir. Hizmet toplanan verileri

üzerinden, kullanıcının kendi makine öğrenmesi modelini ya da hazır sunuluna makine öğrenmesi modellerini kullanarak çalışacak şekilde tasarlanmıştır. (Şekil 3.1).



Şekil 3.1. Azure makine öğrenmesi çalışma şeması

Microsoft Azure Machine Learning Studio (ML Studio), makine öğrenmesi modellerini geliştirmek, eğitmek, test etmek, dağıtmak, yönetmek ve izlemek için kullanabileceğiniz bulut bilişim tabanlı bir ortam sağlamaktadır. ML Studio bir microsoft hesabı alınarak deneme sürümü kullanılabilir. (Şekil 3.2).

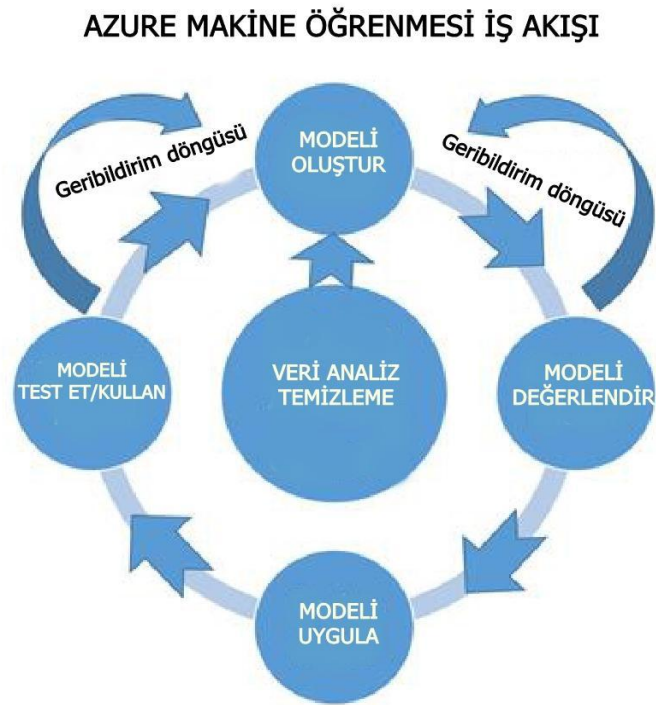


Şekil 3.2. Azure Machine Learning arayüzü.

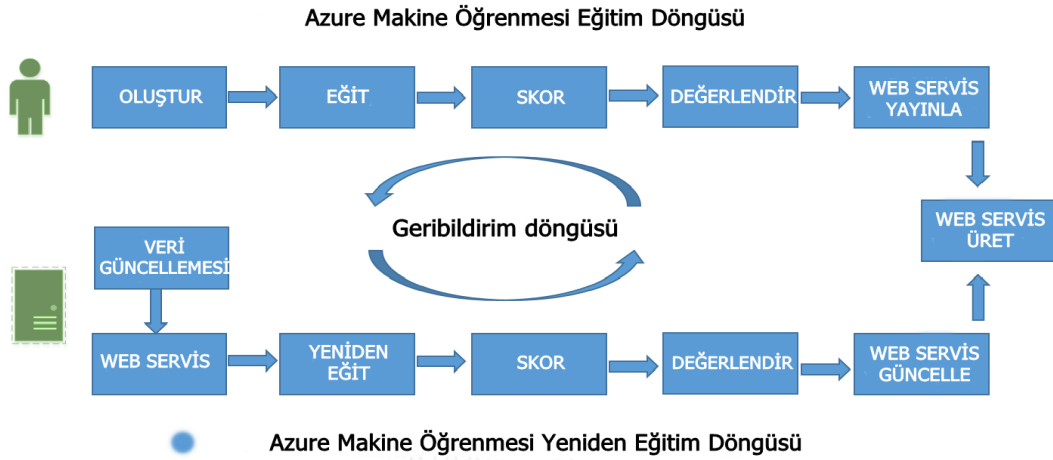
Etkileşimli, görsel bir çalışma alanı kullanılarak veri kümeleri ve analiz modülleri tuvale sürüklenip bırakılır.

ML Studio’da çalıştırılabilen bir deneme oluşturmak için veri kümeleri birbirine bağlanır. Bu bağlamlar ile model oluşturulabilir, eğitilebilir, puanlanabilir ve test edilebilir.

İstenilen sonuçlar alınana kadar model tasarımını yinelenebilir, deneme düzenlenebilir ve çalıştırılabilir. Model hazır olduğunda, bu modele dışarıdan yeni veriler göndererek tahminler alabilmesi için bir web hizmeti olarak yayımlanabilir. Model oluşturma iş akışı Şekil 3.3’de ve modeli çalıştırma süreci Şekil 3.4’de gösterilmiştir (Barnes, 2015).



Şekil 3.3. Microsoft Azure Machine Learning’de model oluşturma iş akışı (Barnes , 2015).



Şekil 3.4. Microsoft Azure Machine Learning’de model oluşturma iş akışı (Barnes , 2015).

3.2. ML Studio’da Model Geliştirmek

ML Studio’da basit bir işlev için için aşağıdaki adımların izlenmesi yeterlidir. Temel olarak adımlar model oluşturma, modeli eğitime, modeli puanlama ve test etme şeklinde gruplandırılabilir:

- **Model oluşturma**
 1. Adım: Verileri alma
 2. Adım: Verileri hazırlama
 3. Adım: Özellikleri tanımlama
- **Modeli eğitime**
 4. Adım: Bir öğrenme algoritması seçme ve uygulama
- **Modeli puanlama ve test etme**
 5. Adım: Yeni verileri tahmin etme

1. Adım: Verileri alma

Makine öğrenimi için ihtiyaç olan ilk şey verilerdir. ML Studio'da kullanabilecek birçok örnek veri kümesi bulunur ve birçok kaynaktan veriler bu ortama aktarılabilir.

2. Adım: Verileri Hazırlama

Genellikle bir veri kümesi analiz edilmeden önce biraz ön işleme gerekir. Örneğin, çeşitli satırların sütunlarında bulunan eksik değerler fark edilmiş olabilir. Modelin verileri doğru şekilde analiz edebilmesi için bu eksik değerlerin temizlenmesi gerekir.

3. Adım: Özellikleri tanımlama

Veriler temizlendikten sonra, tahmine dayalı modelde hangi özellikler kullanılacağı belirtilmelidir. Machine Learning' de özellikler, ilgilendiğiniz bir şeyin tek tek ölçülebilir özellikleridir. Tahmine dayalı bir model oluşturmaya yönelik iyi bir özellikler kümesi bulmak için, çalışma ve çözmek istediğiniz sorun hakkında bilgi sahibi olmak gereklidir. Örneğin birtakım özellikleri kullanmak, hedefi tahmin etmede diğer özellikleri kullanmaktan daha uygun olabilir.

4. Adım: Bir öğrenme algoritması seçme ve uygulama

Bu aşamada veriler hazır hale getirilmiş olur. Bundan sonraki işlemde, tahmine dayalı bir model oluşturmak için eğitim ve test etme işlemlerinin yapılması gerekir. Verilerin bir kısmı eğitim bir kısmı da test işlemi için kullanılır. Algoritma seçerken verinin ne olarak tahmin edileceğine göre bir öğrenme algoritması seçilir. Sınıflandırma ve regresyon, denetimli makine öğrenimi algoritmasının iki türüdür. Sınıflandırma; renk gibi (kırmızı, mavi veya yeşil) tanımlanmış bir kategori kümesinden yanıt tahmin eder. Bir sayıyı tahmin etmek için regresyon kullanılır.

5. Adım: Yeni verileri tahmin etme

Modelde eğitim ve test bağlantılarını da doğru olarak yaptıktan sonra artık model çalıştırılabilir. Modelde bilinen değerler ile tahmin edilen değerler arasındaki farkları inceleyerek tahmin doğruluğu hakkında yorum yapabiliriz.

Model için aşağıdaki istatistikler kullanılmıştır.

- **Ortalama Mutlak Hata** (*Mean Absolute Error - MAE*): Mutlak hataların ortalamasıdır (hata, tahmin edilen değer ile gerçek değer arasındaki farktır).

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - r_i| \quad (3.1)$$

- **Kök Ortalama Karesi Alınmış Hata** (*Root Mean Squared Error - RMSE*): Test veri kümesinde yapılan tahminlerin karesi alınmış hata ortalamasının kareköküdür.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2} \quad (3.2)$$

- **Görelî Mutlak Hata** (*Relative Absolute Error - RAE*): Gerçek değerler ve tüm gerçek değerlerin ortalaması arasındaki mutlak hataların mutlak farka görelî ortalamasıdır.

$$RAE = \frac{\sum_{i=1}^N |\hat{\theta}_i - \theta_i|}{\sum_{i=1}^N |\bar{\theta} - \theta_i|} \quad (3.3)$$

- **Görelî Karesi Alınmış Hata** (*Relative Squared Error - RSE*): Gerçek değerler ve tüm gerçek değerlerin ortalaması arasındaki karesi alınmış hataların karesi alınmış farka görelî ortalamasıdır.

$$RSE = \frac{\sum_{i=1}^n (p_i - a_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2} \quad (3.4)$$

- **Determinasyon Katsayısı** (*Coefficient of Determination*): Karesi alınmış değer olarak da bilinen ve modelin verilere ne kadar iyi uyumlu olduğunu gösteren istatistik ölçümleridir

Her bir hata istatistiği ne kadar küçük olursa o kadar iyidir. Daha küçük olan bir değer, tahminlerin gerçek değerlerle daha yakından eşleştiğini gösterir. Determinasyon katsayısı değeri 1.0'e kadar yakınsa tahminler o kadar iyi olur.

Son olarak da makine öğrenimi öğreticisi tamamladığına ve deneme kurulduğuna göre, model iyileştirmeye devam edebilir ve ardından tahmine dayalı web hizmeti olarak dağıtılabilir.

3.3. ML Studio Train Matchbox Recommender

Train Matchbox Recommender modülü, ML Studio'da yüklenen veri setinden **kullanıcı-öğe-puanlama** (*user-item-rating*) üçlülerinin oluşturduğu bir veri kümesini ve isteğe bağlı olarak belirli kullanıcı ve öğe özelliklerini okur. Eğitilmiş bir Matchbox Recommender sonucu döndürür. Daha sonra, eğitilmiş modeli kullanarak tavsiyeler üretebilir, ilgili kullanıcıları bulabilir veya ilgili öğeler bulunabilir.

Bir öneri sisteminin ana amacı, sistem kullanıcılarına bir veya daha fazla öğe önermektir. Bu öğeye bir film, restoran, kitap, müzik veya herhangi bir örnek ürün verilebilir. Kullanıcı bir kişi veya kişi grubu olabilir. Öneri sistemlerine iki temel yaklaşım vardır. Birincisi, kullanıcılar ve öğeler için özellikler kullanan içerik tabanlı (*content-based*) yaklaşımdır. Kullanıcılar yaş ve cinsiyet gibi özelliklerle öğelerde yazar ve üretici gibi özellikleriyle açıklanabilir. İçerik tabanlı öneri sistemlerinin tipik örnekleri, sosyal çöpçatanlık sitelerinde bulunabilir. İkinci yaklaşım, yalnızca kullanıcıların ve öğelerin tanımlayıcılarını kullanan ve kullanıcıların öğelere verdiği derecelerin bir matrizen oluştuğu ve bu varlıklar hakkında örtülü bilgileri elde eden collaborative filtreleme yöntemidir. Bir kullanıcı derecelendirdiği öğelerden ve aynı öğeler derecelendiren diğer kullanıcılardan nitelendirilebilir.

Matchbox öneri modeli, **collaborative filtrelemeyi** içerik tabanlı (*content-based*) tabanlı bir yaklaşımla birleştirir. Bu nedenle, melez bir tavsiyeci sayılır. Bir kullanıcı sisteme nispeten yeni olduğunda, kullanıcı hakkındaki özellik bilgilerini kullanarak bilinen "cold-start" sorununu ele alarak tahminler geliştirilir. Bununla birlikte, belirli bir kullanıcıdan yeterli sayıda puan (derecelendirme) topladıktan sonra, tek başına özelliklerinden ziyade belirli derecelendirmelerine dayalı olarak tamamen kişiselleştirilmiş öngörüler yapmak mümkündür. Dolayısıyla, content-based önerilere, collaborative- filtrelemeye dayalı önerilerle uyumlu bir geçiş olmaktadır. Kullanıcı veya öğe özellikleri bulunmasa bile, Matchbox yine de collaborative filtreleme modunda çalışmaktadır.

3.4. ML Studio'da Hibrid Filtreleme Tekniğini İçeren Train Matchbox Recommender Modülünü Kullanarak Ürün Önerisi Sunmak

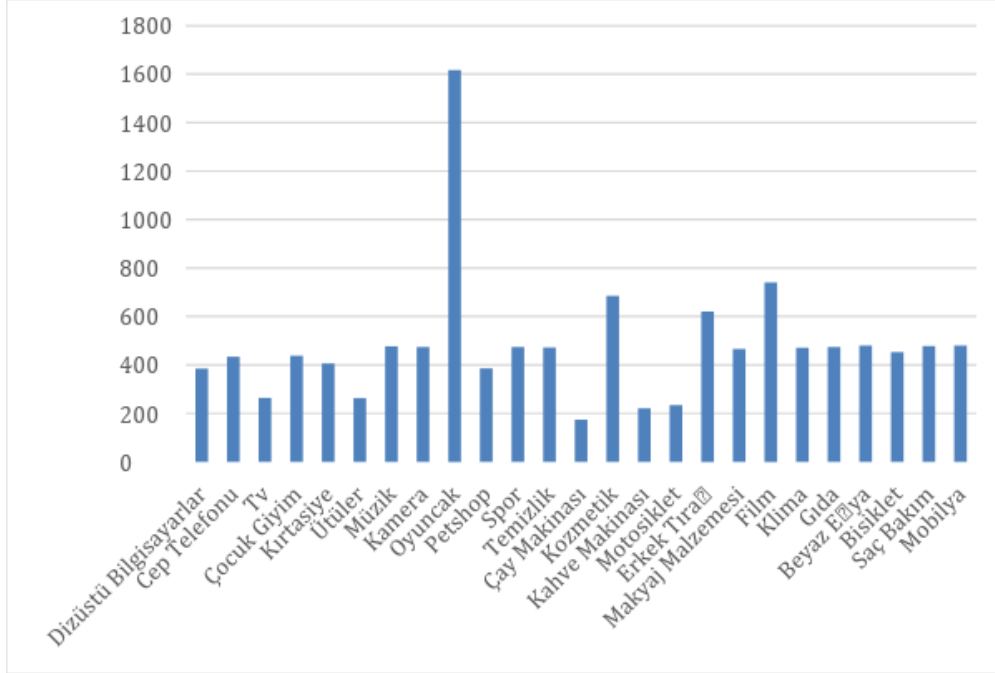
Bir e-ticaret öneri sistemi geliştirmek için Matchbox öneri modüllerinin kullanımının gösterildiği bir uygulama yapılmıştır.

Çalışmada özgün bir filtreleme mantığı kullanılmıştır. Model belirli kullanıcılar tarafından derecelendirilen ürün bilgilerinden oluşturulmuştur. Model bize sonuç olarak kullanıcıların en çok hangi ürünleri beğendiklerini göz önünde bulundurarak bir kullanıcıya hangi ürünlere göz atması gerektiği üzerinde tavsiyede bulunmuştur. Böylece kullanıcının ilgisini çekebilecek en iyi ürünler tavsiye edilmiştir. Çalışmada, bir tavsiyenin kullanılabilmesi ve değerlendirilebileceği farklı modların gösterilebilmesi için, model eğitilip yeni veriler puanlanmıştır.

3.5. Veri Seti

Çalışmada öneri algoritmasının çalışma mantığını daha iyi anlamak için oldukça fazla kullanıcı ve ürün içeren bir veri seti kullanılmıştır. Veri seti 25 farklı kategoriden en çok puan ve yorum alan ürünleri içermekte ve 94875 kullanıcı, 12066 farklı ürün, bu ürünlere kullanıcılar tarafından verilen puanlardan oluşan 207411 puan grubundan oluşmaktadır. Bu veriler bir Türk e-ticaret sitesinden C# ile geliştirilen bir program

yardımıyla toplanarak elde edilmiştir. Ürünlerin kategorileri ve her kategoride kaç ürün olduğu Şekil 3.5 ve Çizelge 3.2 de belirtildiği gibidir.



Şekil 3.5. Kategorileri göre ürün sayılarının grafiksel dağılımı

Çizelge 3.2. Veri setindeki ürün kategorileri ve ürün sayıları

Ürün Türü	Ürün Sayısı
Dizüstü Bilgisayarlar	385
Cep Telefonu	434
Tv	264
Çocuk Giyim	438
Kırtasiye	406
Ütüler	263
Müzik	477
Kamera	474
Oyuncak	1617
Petshop	386
Spor	474
Temizlik	472
Çay Makinası	174
Kozmetik	685
Kahve Makinası	221
Motosiklet	234
Erkek Tıraş	620
Makyaj Malzemesi	466
Film	740
Klima	471

Gıda	474
Beyaz Eşya	480
Bisiklet	453
Saç Bakım	478
Mobilya	480




Her veri örneği bir kullanıcı bilgisi, bir ürün bilgisi ve birde ürüne verilen puan (derece) bilgisinden oluşan bir gruptur.

ML Studio'da ilk olarak Şekil 3.6 daki verileri içeren kullanıcı, ürün ve ürün derecesi bilgisinden oluşan veri grubu eklenir.

rows	columns			
207409	3			
		urunId	userId	rating
view as				
		ABO0098001YSL	ARZU KUTANCI	2
		ABO0098001YSL	GÜLŞAH KARAKAYA	4
		ABO0098001YSL	GÜLŞAH KARAKAYA	3
		ABO0098001YSL	MERVE KUNDAK	1
		ABO0098001YSL	MERHAMET ALI DOĞRU	3
		ABO0098001YSL	YASAR ÇETİNKAYA	3
		ABO0265101	MERVE KUNDAK	1
		ABO0265102	AHMET DEMİRELİ	3
		ABO0265102	EEBUT KAYA	4
		ABO0265102	EMRE NURCANI DOĞRU	5
		ABO0265102	ERDİNÇ TUBAN	2
		ABO0265102	ERMIN BİNİR	5
		ABO0265102	GÜLŞAH KARAKAYA	1
		ABO0265102	GÜLŞAH KARAKAYA	5

Şekil 3.6. Veri setindeki örnek kullanıcı özellikleri

Daha sonra bu veriye Şekil 3.7 deki gibi derecelendirme verilerinde ürünleri tanımlayan alınan ürün isimleri ve türlerini içeren bir dosya eklenmiştir.

rows	columns	urunId	title
12066	2		
view as			
		ZYHEN100529785	Perwoll Yenilenen Renkler Hassas camaslr Deterjanl 4 lt 66 Ylkama
		ZYHEN4438	Perwoll Yenilenen Renkler Hassas camaslr Deterjanl 3 lt
		HBV00000BA514	Perwoll Yenilenen Renkler Hassas camaslr Deterjanl 3 lt 4'lu Set
		SGGILLETTE5486	Gillette Fuslon ProGilde Flexball Tiras Paketi
		SPORALTIS552	Altis SD 5000 Plus 3 hp Masajll Otomatik Eglml Kосу Bandl
		HBV000007SI33	Hattrick SD5000 Prime M 2
		HBV000006FL8Z	Altis SD5000 Auto 2
		SGGIL2018029655	Gillette Mach3 Tiras Makinesi
		TELCEPLGG432GR-DK3	LG G4 32 GB Dual Sim
		HBV000008S8AE	Samsung Galaxy Note Fan Editlon Dual Sim
		HBV0000012S8T	Apple iPhone 7 32 GB

Şekil 3.7. Veri setindeki örnek ürün bilgileri

Burada derecelendirme alanı bir tamsayı olarak değil *numeric* olarak görülmektedir. Eğitim modelimiz tamsayı derecelendirme gerektirdiğinden puan alanını tam sayıya çevirmek gerekmektedir. Bu işlemi yapmak için Meta Veri Editörü kullanılmıştır. Meta Veri Editörünün parametreleri Şekil 3.8 deki gibi ayarlanmıştır. Burada işlem yapılacak kolon Rating olarak seçilmiştir.

Bu veriler yüklendikten sonra Şekil 3.9'da görüldüğü gibi çalışmada birleştirilmiştir.

▲ Metadata Editor

Column

Selected columns:
Column names: Rating

Launch column selector

Data type

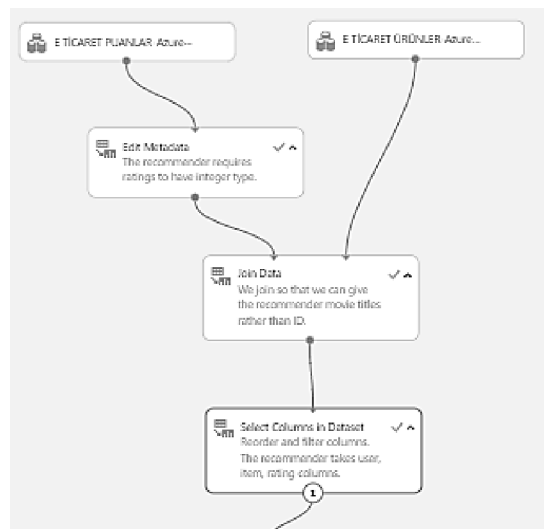
Categorical

Fields

New column names

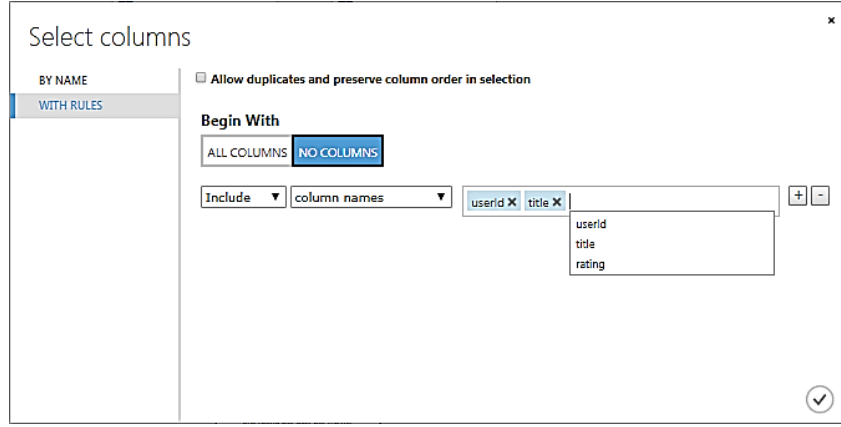
Şekil 3.8. Meta veri parametreleri

Train Recommender modülü kullanıcı ve öge tanımlayıcılarına göre daha esnekler. Bu nedenle genellikle bu model için tamsayı tipinde değerler kullanılır. Sonuçların daha kolay anlaşılması için derecelendirmeler ve ürün başlığı veri kümeleri birleştirilmiştir. Bu birleştirme yapılırken her iki veri kümesinde de bulunan ortak bir alan olması ve bunun Join bileşeninde belirtilmesi gerekir. Aksi takdirde hatalı bir birleştirme işlemi olacaktır.



Şekil 3.9. Veri seti ve meta veri parametrelerinin ilişkilendirilmesi

Train Recommender modülü girdinin eğitim için kullanılan üç alanını içermesi gerekmektedir. Bu nedenle yalnızca **userId** (*Kullanıcı Ad*), **title** (*ürün ismi*) ve **rating** (*ürüne verilen puan*) alanlarını seçmek için Project Columns bileşeni kullanılır. Ancak bu veri kümesi aynı kullanıcı-ürün çiftleri için çelişen birkaç derece içermektedir. Bu eğitim ve değerlendirmede gürültüye neden olduğundan, aynı değerler kaldırılır, karşılaşılan her bir kullanıcı-ürün çifti için keyfi olarak yalnızca verilerden bir tanesi korunur (Şeki 3.10).



Şekil 3.10. Veri setinde gürültüye neden olabilecek değerlerin kaldırılması

Herhangi bir istatistiksel modelde olduğu gibi bir veri setinde bulunan değerlerin bir kısmı test bir kısmı da modelin eğitilmesi (öğrenmesi) için kullanılmıştır. Ama bu verilerinde rastgele değilde anlamlı olarak dağılması gerekir. ML Studio da bu görevi Split bileşeni üstlenmektedir. Split modülünde Şekil 3.11 da görüldüğü gibi verilerin yüzdeler olarak ne kadarının eğitim ve test için kullanılacağı belirlenir. Burada Splitting mode parametresini varsayılan olarak verilen değer olarak bırakmamız yeterlidir.

▲ Split Data

Splitting mode

Recommender Split ▼

Fraction of training-onl... ☰

0.75

Fraction of test user rat... ☰

0.25

Şekil 3.11. Veri setinin öğrenme ve test için ayarlanması

Burada verilerin %75 i eğitim, %25 ise test olarak kullanılmıştır. Bu işlemlerden sonra artık model eğitime hazır hale gelmiştir. Modelimize Train Recommender eğitim modülü eklenir. Bu Eğitim modeli üç adet giriş parametresi gerektirmektedir.

Özellik sayısı (*Number of traits*): Bu, her bir kullanıcı ve her öge için öğrenilecek gizli parametrelerin sayısını belirler. (Teknik olarak, önerilen algoritma, kullanıcı-ürün etkileşim matrisini çarpanlara ayırmaya dayalıdır: Bu parametre, yaklaşımın sıralamasını belirler.) Daha fazla özellik, daha güçlü bir model sağlar, ancak eğitim verilerinin uyumu üzerinde risk oluşturur.

Parametre, genellikle kabul edilebilir bir performans elde eden en küçük sayı bulma amacı ile deneme yoluyla belirlenir. Bu sorun için, genellikle 20 özellikten oluşan değer kullanılmaktadır, ancak bizim verilerimiz fazla olduğu için modelimizde 10 olarak aldık. Diğer değerleri de denemek önerilir.

Yineleme sayısı (*Number of recommendation algorithm iterations*): Model parametreleri, rastgele başlatma ile bulgulanır ve bunu takiben iteratif bir degradasyon iniş tekniği kullanılarak kalan bir hatayı (her bir kullanıcı-ürün çifti için doğru ve tahmini derecelendirmeler arasındaki fark) asgariye indirilir. Hata genelde katlanarak azalır; bu, faydanın çoğunun başlangıç yinelemelerin de meydana geldiği anlamına gelir. Matchbox öneri, girdi verisinin üzerinde birden çok kez yineleyebilen bir ileti gönderme algoritması kullanarak eğitilir. Bu nedenle, optimizasyonu yakınsaklığa kadar

çalıştırmak değil, eğitim süresini sınırlamak için yinelemeleri makul bir sayı ile sınırlamak yaygın bir uygulamadır. Bu sayı ne kadar yüksek olursa, tahminler o kadar doğru olur; Ancak, eğitim yavaşlayacaktır. Genellikle yinelemelerin sayısı 1- 10 aralığındadır. Modelimizde bu değer 6 olarak kullanılmıştır.

Eğitim gruplarının sayısı (*Number of training batches*): Bu değere dayalı olarak, kullanıcı ögesi puanlama üçlülerinin veri seti, eğitim sırasında birden fazla parçaya veya partilere bölünmüştür.

Train Matchbox Recommender, toplu iş paketlerini çalıştırdığından, tüm eğitim modeli verileri belleğe sığarsa, eğitim toplu işlem sayısının kullanılabilir çekirdek sayısına ayarlanması önerilir. Aksi halde, eğitim gruplarının sayısı, eğitim verisinin belleğe sığdırıldığı mevcut çekirdek sayısının en düşük katına ayarlanmalıdır.

Varsayılan olarak, eğitim verileri dört partiye bölünmüştür. Sadece kullanıcı-ürün-derecelendirme üçlülerini içeren veri seti bölünmüştür. Kullanıcı veya öge özelliklerinin bölünmesine gerek yoktur.

3.6. ML Studio’da İşbirlikçi Filtreleme Yöntemi ile Ürün Önerisi Sunmak

İşbirlikçi (Collaborative) filtreleme yöntemi bir kullanıcının ilgi alanı bilgileriyle birlikte ona ürün tercihi olarak yakın olan diğer kullanıcıların bilgilerinden faydalanır. Bir kullanıcı derecelendirdiği öğelerden ve aynı öğeler derecelendiren diğer kullanıcılardan nitelendirilebilir. ML studio da işbirlikçi filtreleme tekniği kullanarak öneri sistemi oluşturmak için Execute R Script Modülü kullanılacaktır.

3.6.1. R Programlama Dili ve ML Studio’da Execute R Script Modülü

R istatistiksel hesaplama ve grafikleri için bilgisayar programı olup aynı zamanda bir programlama dilidir. Yeni Zelanda Auckland Üniversitesinden Ross Ihaka ve Robert Gentleman tarafından ortaya çıkarılan R hali hazırda R Geliştirme Çekirdek Ekibi tarafından geliştirilmekte. S programlama dilinin açık kaynak kodlu versiyonu olan R bundan dolayı bazen GNU S olarak da anılmaktadır.

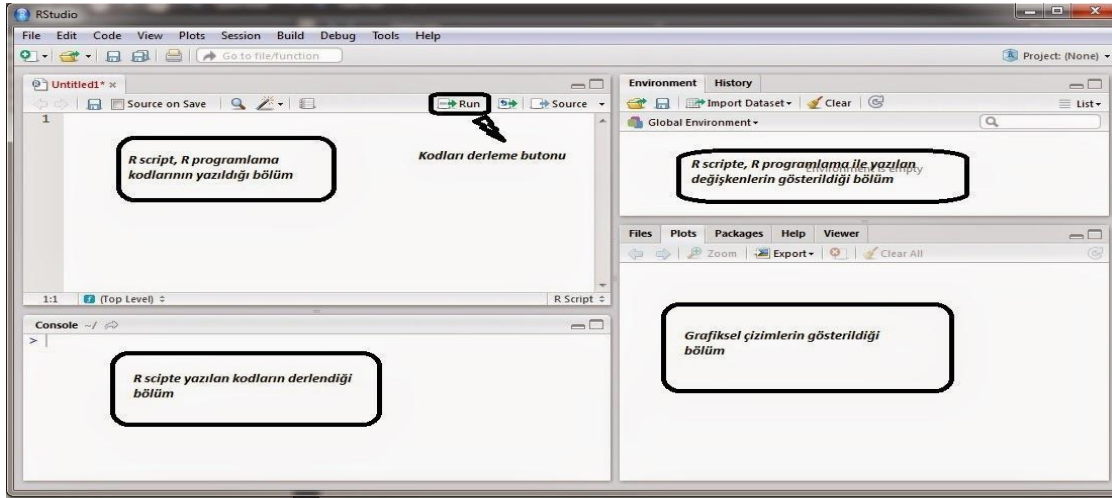
İstatistiki yazılım geliştirme için istatistikçiler arasında de fakto standart haline gelen R istatistiki yazılım geliştirme ve veri analizi alanında kullanılmaktadır.

GNU projesinin bir parçası olan R nin kaynak kodları GNU Genel Kamu Lisansı altında olup değişik işletim sistemleri için sürümleri mevcuttur. R komut satırı arayüzü kullanıyor olsa da değişik grafik kullanıcı arayüzleri de bulunmaktadır.

R çok geniş istatistiki (doğrusal ve doğrusal olmayan modelleme, klasik istatistik testleri, zaman serileri analizi, sınıflandırma, kümeleme ve diğer) ve grafik çizim teknikleri sunmaktadır. R sisteminin büyük kısmı aynı dille yazıldığından kullanıcının yapılan algoritmik seçimleri takip etmesi kolay olmaktadır. Yoğun hesaplamalar gerektiren görevler için C,C++ ve Fortran kodu çalışma zamanında bağlanıp çalıştırılır hale gelmektedir. İleri seviye kullanıcılar R nesnelere manipülasyonu için C kodu yazabilirler.

R, kullanıcıların eklediği özel fonksiyonlar veya çok özel araştırma alanlarına ait paketlerle oldukça geliştirilebilir. R diğer birçok istatistiki hesaplama dilinden daha kuvvetli bir nesneye yönelik programlama kabiliyetine sahiptir.

R'nin bir diğer güçlü yönü matematiksel sembolleri de içeren yayın kalitesinde grafikler çizebilen grafik imkânlarıdır. R kodu yazmak için R Studio derleyicisi kullanılır. Şekil 3.12'de R Studio arayüz ekranı gösterilmiştir. ML Studio'da R kodları çalıştırmak için Execute R Script modülü mevcuttur. Bu modül geliştirdiğimiz R kodunu ML Studio'da çalıştırmanızı ve oluşturduğumuz modelimizde kullanmaya yaramaktadır (Özdemir ve ark., 2010).



Şekil 3.12. R arayüz görünümü

3.6.2. R Studio'da İşbirlikçi Öneri Modelinin Geliştirilmesi Geliştirilmesi ve ML Studio'da Kullanılması

İlk adımda Bölüm 3.5.1'de tanıtılan ürün veri setinin okunması gerçekleştirilmiştir. Daha sonra R Studio da işbirlikçi filtreleme yönteminin kullandığı ilişkilendirme kurallarını belirlemek ve sık kullanılan ürünleri sunmaya yarayan “arules” ve “arulesViz” paketleri yüklenmiştir.

Ürün verilerinin sayılarına göre frekans dağılımı yapılmıştır ve bu ürünler arasında ilişkilendirme kuralları apriori algoritmasına göre belirlenmiştir. Burada sadece en az 0.001 destek ve en az 0.8 güven içeren kurallar kullanılacaktır. Daha sonra kullanıcı veri setinin okunması gerçekleştirilmiştir. Kullanıcı veri seti ile ürün veri seti arasındaki kurallar belirlenerek işbirlikçi filtreleme yöntemi oluşturulmuştur. Çalıştırılan komut ML Studio da kullanılmaya hazır hale getirilmiştir.

Bütün adımları çalıştıran komutlar aşağıda verilmiştir:

```
library(arules) // arules paketinin yüklenmesi
library(arulesViz) // arulesViz paketinin yüklenmesi

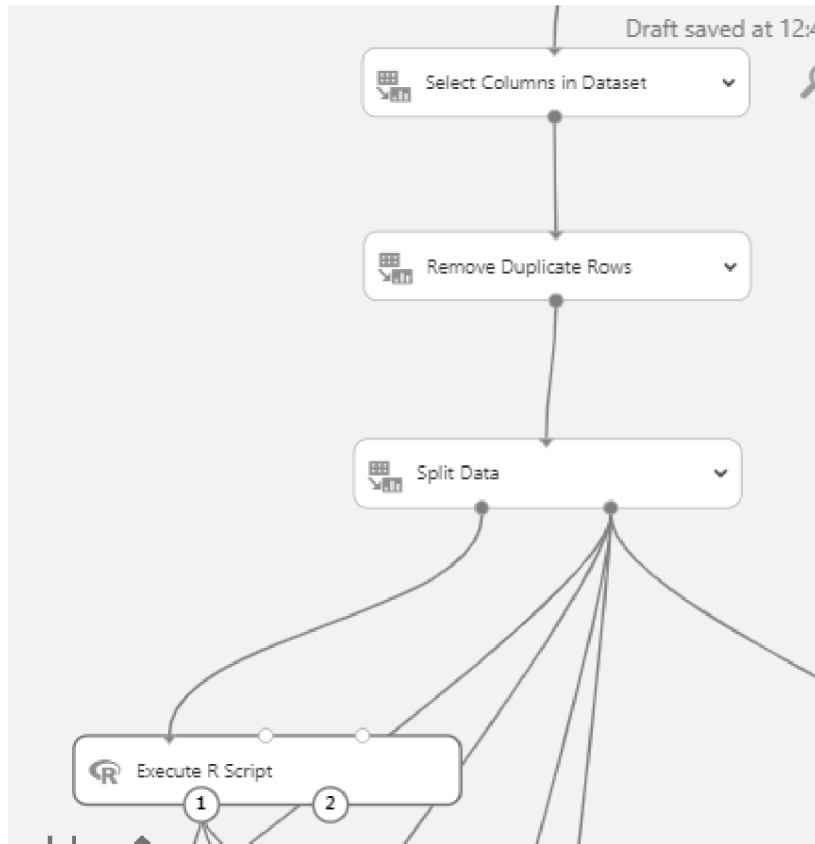
tr = read.transactions("E_Ticaret_Urunler.csv", format="basket", sep=",")
rules <- apriori(tr, parameter = list(supp = 0.001, conf = 0.8))
```

```
Puanlar = read.csv("E_Ticaret_Puanlar.csv")

dataset1 <- as.vector(t(Puanlar))

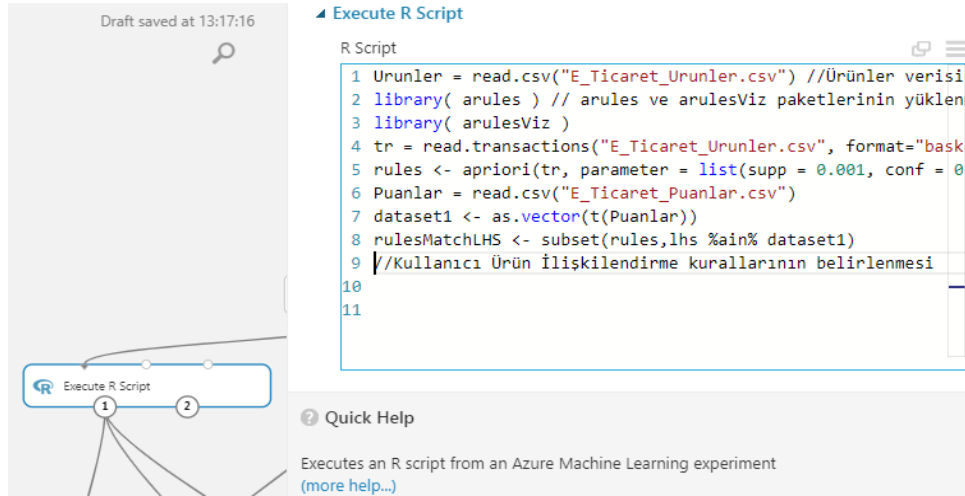
rulesMatchLHS <- subset(rules,lhs %ain% dataset1) //Kullanıcı Ürün
İlişkilendirme kurallarının belirlenmesi
```

ML Studio'da aynı veri setleri ile oluşturduğumuz hibrid model üzerinde Train Matchbox Recommender modülü yerine Execute R Script modülü kullanılmıştır. Modelin diğer kısımları ve yapılan işlemler aynı olarak kalmıştır (Şekil 3.13).



Şekil 3.13. ML Studio'da Execute R Script modülünün kullanılması

Yazılan R kodları Execute R Script modülünün içine eklenip modelimiz tekrar çalıştırılarak öneri sonuçları tekrar alınmıştır (Şekil 3.14).



Şekil 3.14. Execute R Script modülünde R kodlarının kullanılması

Çizelge 3.3 de modelimizin çalıştırılması sonucu elde edilen MAE, RMSE ve NCDG değerleri gösterilmiştir.

Çizelge 3.3. İşbirlikçi filtreleme sonuç değerleri

Öneri Yöntemi	MAE	RMSE	NCDG
İşbirlikçi (Collaborative)	2,24	2,85	0,75

Elde edilen öneri sonuçları Şekil 3.15 de gösterilmiştir.

rows	columns	User	Item 1	Item 2	Item 3	Item 4	Item 5
100	6	F***** Y*****	Apple iPhone SE 16GB	Apple iPhone 6 32 GB	Samsung Galaxy A5 2017	Samsung UE-55MU7000 55" 139 Ekran Uydu Alicili 4K Ultra HD Smart LED TV	Kingshark KS-523 HD-DVR 2.5" Hareket Sensörlü Çift Lensli Araç İçeri Kamera
		P*****	Lumberjack A3374333	ucer Ahsap Eglitci	Learning Toys Geometrical	UK Polo Club Outdoor Erkek Bot - Slyah	Korkmaz A 359 Demkolik Elektrikli caydanlık - Slyah
			Snow Uc Gunluk cocuk %100 Su Geclmez Lastik Kar Botu	Oyuncak Lablrent Koordinasyon Oyunu	Building Block		

Şekil 3.15. İşbirlikçi filtreleme öneri sonuçları

3.7. R Studio'da İçerik Tabanlı Öneri Modelinin Geliştirilmesi ve ML Studio'da Kullanılması

İçerik tabanlı model yapısı da işbirlikçi filtreleme modeline benzer şekilde kurulmuştur. Fakat yüklenen paketler farklılık göstermiştir. İçerik tabanlı filtreleme için *recommenderlab* paketi yüklenmiştir.

Sistemin diğer kod yapısı ve ML Studio ya yüklenmesi işlemi işbirlikçi filtreleme yöntemindeki adımları içermektedir.

Çizelge 3.4 de modelimizin çalıştırılması sonucu elde edilen MAE, RMSE ve NCDG değerleri gösterilmiştir.

Çizelge 3.4. İçerik tabanlı filtreleme sonuç değerleri

Öneri Yöntemi	MAE	RMSE	NCDG
İçerik Tabanlı (Content-Based)	2,89	2,95	0,71

Elde edilen öneri sonuçları Şekil 3.16 de gösterilmiştir.

User	Item 1	Item 2	Item 3	Item 4	Item 5
F***** Y*****	Apple iPhone SE 16GB	Apple iPhone 8 Plus 64 GB	HP 14-BS006NT Intel Celeron N3060 4GB 500GB Freedos 14" Taşınabilir Bilgisayar 2B599EA	Xiaomi Mi Note 3 64 GB	Apple iPhone X 64 GB
P*****	Lumberjack A3374333 Snow Uc Gunluk çocuk %100 Su Geçirmez Lastik Kar	Jump 16704 Icl Termal Kurklu Fermuarlı Kiz çocuk Bot Ayakkabı	Grisport 11520 Dakar Unisex Trekking Deri Bot	Sozzy Oyuncaklı Çıngıraklı Çorap Patik Ve El Bilekliği	Faber-Castell 188920 Silgi 7089/20 Siyah

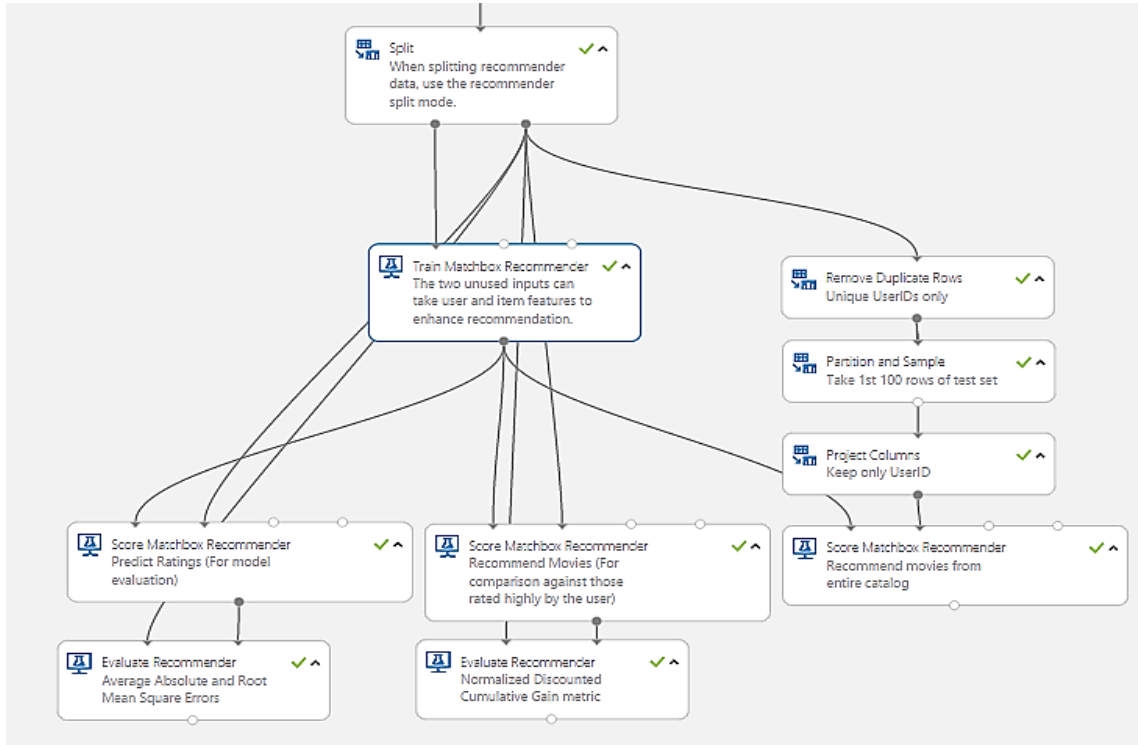
Şekil 3.16. İçerik tabanlı filtreleme öneri sonuçları

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Yapılan örnek çalışmada eğitilmiş öneri modülünün nasıl kullanılacağı ile ilgili üç farklı yol verilmiştir:

- Ürün derecelerini tahmin etmek,
- Her kullanıcı tarafından daha önce puanlanmış bir listeden en iyi n adet ürünü tahmin etmek,
- Her bir kullanıcı için tüm ürün kataloglarından n adet ürün önerisi yapmak.

İlk iki yöntem sadece öğrenilen modelin performansını değerlendirmek için kullanılırken, son yöntem tipik bir tavsiye durumunu temsil eder. Modelin şeması Şekil 4.1’de verilmiştir.



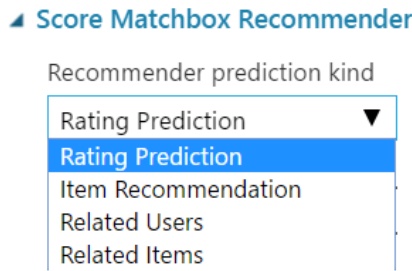
Şekil 4.1. Model şeması

Üç farklı tahmin türü gerçekleştirmek için Score Recommender modülü kullanılır. Modülün iki gerekli ve iki isteğe bağlı girişi vardır.

- İlk gerekli girdi eğitim modelinden gelen veridir. Bu durumda, eğitim modelinin çıktısı doğrudan bağlanır.
- İkinci girdi, puanlanmış bir veri kümesidir. Bu veri kümesinin biçimi aşağıdaki gibi işleve bağlı olmuştur.
- İki isteğe bağlı bağlantı noktası, eğitim sırasında isteğe bağlı girdilere benzer kullanıcı ve öğe meta verileri içindir. Eğitilirken hangi verileri kullandıysanız, puanlama yaparken de aynı verileri kullanmanız gerekmektedir.

4.1. Öneri Sisteminin Çalıştırılması

Ürün önerisi için Score Recommender modelden **Related Items (İlişkili ürünler)** parametresi seçilmiştir. Eğitim için kullanılan üç maddeli tanımlama grubu kullanılarak sonucu alınmak istenilen girdi veri kümesi ile öneri sağlanmıştır (Şekil 4.2).



Şekil 4.2. Veri kümesinin tahminleme seçeneği

Score Recommender modülünde, kullanıcı-ürün çifti için ürün derecesi tahmin etmek istersek **Rating Prediction** seçeneği kullanılır. Eğer sonuç olarak ilişkili kullanıcılar veya öğeleri döndüreceksek **Related Users** ve **Related Items** seçenekleri kullanılır.

Önerilerin doğruluğunu değerlendirmek için **Evaluate Recommender** modülünü kullanırız. İlk girdi, eğitim için kullanılan verilere benzer tanımlama grubu

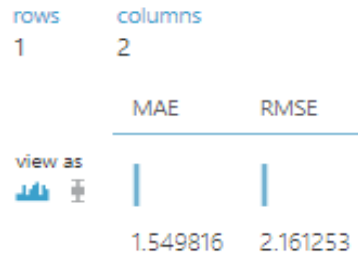
(ürün, kullanıcı derecelendirme) içeren test veri kümesidir. Genellikle, bu verileri, çalışmayı kurarken kullanılan Split modülünün test çıkış bağlantı noktasından çıkan veri kümesi çıktısı kullanarak sağlanır. Evaluate Recommender model iki adet parametreye ihtiyaç duyar:

- Minimum öge sayısı
- Minimum kullanıcı sayısı

Bu parametreler kullanılarak değerlendirmeyi en az n öge derecelendiren kullanıcılarla sınırlandırma yapılabilir ve en az m adet kullanıcı tarafından derecelendirilen sınırlandırma yapılabilir.

Bu çalışmada, ikinci girdi, modelin başında kullanılan aynı dizi tanılama grubunu içerir; bu nedenle değerlendirme, aşağıdaki ölçümleri kullanarak öngörülen derecelendirmeleri gerçek derecelendirmelerle karşılaştırılmıştır.

Çalışmada elde edilen bu değerler Şekil 4.3 de gösterilmiştir.



Şekil 4.3. Elde edilen MAE ve RMSE değerleri

4.1.1. Testlerden Ürün Önerileri

Çalışmanın bu bölümünde, her bir kullanıcı için en çok n adet ürün sıralamalı bir listesini oluşturmak için elde edilen model kullanıldı. Ancak bu ürünler daha önce puanlanan ürünlerden seçildi. Veri girdisi, eğitim için kullanılan aynı ürün kullanıcı derecelendirme formatındadır. Score Matchbox Recommender modülü, kullanıcı kümesini çıkarmak için tanımlama grubu kullanacak ve sonra her kullanıcıya sıralı listeyi oluşturmak için kullanılacak bir ürün seti oluşturmuştur.

Burada iki adet parametre belirlenmesi gerekmektedir:

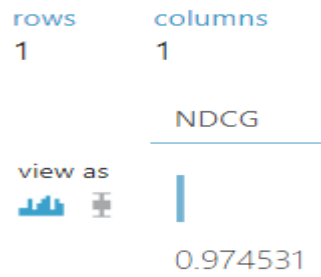
- **Maksimum öneri ögesi sayısı:** Model, her kullanıcı grubundaki tüm ürünleri puanlayacak ve bu sayıya kadar çıkacak ve tahmini dereceye göre yüksekten düşüğe göre sıralanmıştır. Burada model için 5 seçilmiştir. Yani 5 adet ürün tavsiye edilmiştir.
- **Önerilen kümenin minimum boyutu:** Puanlanmış ürünler ile ilgilendiğimizden, derecelendirilmiş öğelerin sayısı 2'den az olan kullanıcıları dâhil etmek mantıklı değildir. Modelin performansını değerlendirmek için bu sayı artırılabilir.

Bir öneri sistemi için tipik bir kullanım örneği, bir kullanıcının ilgi duyduğu n adet öğeyi tüm öğelerin ürün listesinden talep etmektir. Bu süreçte, sonuca yapılan girdi, öneriler üretmek için kullanıcı kimlikleri içeren yalnızca bir sütun içermelidir. Bu yaklaşımı göstermek için, test verileri alınarak ve benzersiz kullanıcı kimliklerinin listesi çıkartılarak 100 kullanıcı kimliği listesi oluşturulmuş ve ilk 100'ü seçmek için Bölüm ve Örnek modülünde Head seçeneği kullanılmıştır. Sonuç, her bir kullanıcı için bir satır içeren bir veri kümesidir ve istenen her bir tavsiye için bir sütun bulunur. Çıktı, sağlanan 100 kullanıcı kimliği için beş öneriyi göstermektedir (Şekil 4.4).

rows	columns						
100	6	User	Item 1	Item 2	Item 3	Item 4	Item 5
		A***** S*****	Apple iPhone 7 Plus 128 GB	Fairy Platinum Bulasik Makinesi Deterjanli Kapsulu Limon Kokulu 90 Yikama 2'li Paket + Platinum Sivi Bulasik Deterjanli Limon 870 ml + Fairy Sunger	Apple iPhone 7 Plus 32 GB	Gillette Fuslon ProGilde Flexball Tiras Paketi	Apple iPhone SE 16GB
		G***** O*****	Apple iPhone 7 128 GB	Rovi Everyday Parlak Fotograf Kagidi - 200Gsm - 50Yp - 10X15cm	Fairy Platinum Bulasik Makinesi Deterjanli Kapsulu Limon Kokulu 90 Yikama 2'li Paket + Platinum Sivi Bulasik Deterjanli Limon 870 ml + Fairy Sunger	Pantene Sac Bakim Kremi Nem Terapisi 470 ml	SanDisk Ultra 32GB 98MB/S Class 10 microSDXC Hafiza Karti + Adaptor SDSQUAR-032G-GN6MA
		M*****	Bluezen slsme Pilates Esegli	Taba Grup Manyetik Meyveler Sebzele 1307	Masal Anlatan Boncuk Kopecik		
		M*****	Bluezen slsme Pilates Esegli	Helen's slsme Pilates Hayvanli-nek			

Şekil 4.4. Önerilen ürün listesi

Son değerlendirme için Evaluate Recommender modülü tekrar kullanılır. İlk girdi, verilerin test bölünmesidir, ancak bu sefer puanlanan veri kümesi kullanıcı başına bir öneri olarak tanınır, bu nedenle modül Normalleştirilmiş İskonto Edilmiş Kümülatif Kazanç'ı (NIEKK) hesaplar. NIEKK, arama sorgusu ve tavsiye sonuçlarını değerlendirmek için sıklıkla kullanılan bir metriktir. 0 ile 1 arasında bir sayı üretir. Burada; 1, sıralamanın mükemmel olduğunu 0 ise, döndürülen öğelerden hiçbirinin kullanıcının gerçek değerinde olmadığını belirtir. Orta değerler ise listedeki en iyi dereceye sahip değerleri gösterir.



Şekil 4.5. Sıralamanın değerlendirme puanı

Sonuç olarak Şekil 4.5 de görüldüğü gibi 0.97 değeri elde edilmiştir ve bu değer analiz edildiğinde 1'e yakın olduğu için oldukça iyi bir değer olarak görülmektedir.

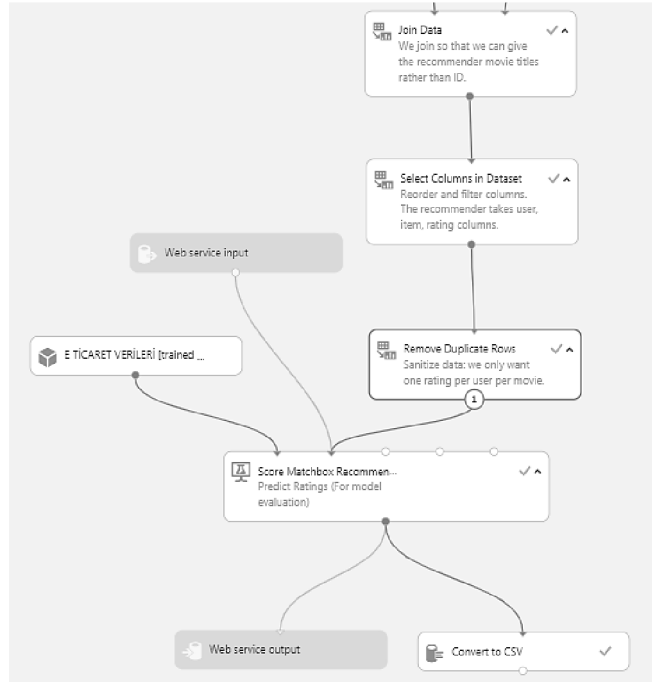
Train Matchbox Modülünde en iyi öneri sonuçları ideal parametre değerlerinde elde edilmiştir. Bu değerler Şekil 4.6 da koyu renkle belirtilmiştir. Eğitim modelimizdeki parametreler değiştiğinde MAE, RMSE ve NDCG değerlerindeki değişim gösterilmiştir (Çizelge 4.1).

Çizelge 4.1. Train Matchbox Modülünde Değişen Parametre Değerlerinin MAE, RMSE VE NDCG Sonuçlarına Yansıması

Number of Traits	Number of Recommendation	Number of Train Batches	MAE	RMSE	NDCG
10	6	4	1.54	2.16	0.97
10	15	4	1.48	2.59	0.96
10	6	10	2.57	3.58	0.90
30	3	4	1.51	2.17	0.84
60	3	4	1.91	2.63	0.87

4.2. Web Servis ile Tahminleme Modeli İçin Geliştirilen Arayüz

Azure Makine Öğrenmesinin en önemli özelliği, modelleri Windows Azure' da web hizmetleri olarak kolayca yayınlama olanağıdır. Recommender'i yayınlamak için ilk adım eğitim modelini kaydetmektir. Bunu, Train Recommender' ın çıkış bağlantı noktası tıklanarak ve Eğitim Modeli Olarak Kaydet seçeneği seçilerek yapılabilir. Daha sonra yalnızca puanlama modülü olan yeni bir deneme oluşturulur ve kaydedilen model eklenir. Ayrıca örnek girdi verilerinin sağlanması gerekir. bu nedenle 100 kullanıcı kimliğini örneklemek için oluşturulan veri hattı yeniden kullanılır. Web hizmeti giriş ve çıkış noktalarını belirlemek için özel Web Servis modüllerini kullanır. Web servisi giriş modülünün, giriş verilerinin çalışmaya gireceği düğüme eklendiğini unutulmamalıdır. Model ile web servis bağlantısının kurulması Şekil 4.6' da gösterilmiştir.



Şekil 4.6. Model ile web servis bağlantısının kurulması

Çalışma başarıyla çalıştırdıktan sonra, deneme tuvalinin altındaki Web Hizmetini Yayınla' yı tıklanarak yayımlanabilir. Web servis üç farklı eğitim modeli içinde çalıştırılarak öneri sonuçları listelenmiştir.

Web servisin çalıştırılması sonucu F**** Y**** kullanıcısı için web sitesinde listelenen öneriler Şekil 4.8, Şekil 4.9 ve Şekil 4.10 da üç farklı eğitim modeli içinde gösterilmiştir.

ÖNERİ MODELİNİ SEÇİNİZ : Train Matchbox Content-Based Colloborative

ÜRÜN SEÇİNİZ : ▼

ÜRÜN ÖNER

ÖNERİLEN ÜRÜNLER (Seçilen Kullanıcı/Model : F** Y**** / Train Matchbox)**

1--Apple iPhone SE 16GB

2--Apple iPhone 8 64 GB

3--Apple iPhone X 64 GB

4--Samsung Galaxy S8 Plus

5--SJCAM SJ6 Legend 4K Akslyon Kamerası - Silyah

Şekil 4.8 Train Matchbox eğitim modeli ile web servisin çalıştırılması sonucu elde edilen kullanıcı bazlı öneri listesinin web sitesindeki örnek kullanımı

ÖNERİ MODELİNİ SEÇİNİZ : Train Matchbox Content-Based Colloborative

ÜRÜN SEÇİNİZ :

ÜRÜN ÖNER

ÖNERİLEN ÜRÜNLER (Seçilen Kullanıcı/Model : F***** Y***** / Content-Based)

- 1--Apple iPhone SE 16GB
- 2--Apple iPhone 8 Plus 64 GB
- 3--HP 14-BS006NT Intel Celeron N3060 4GB 500GB Freedos 14" Taşınabilir Bilgisayar 2BS99EA
- 4--Xiaomi Mi Note 3 64 GB
- 5--Apple iPhone X 64 GB

Şekil 4.9 Content-Based eğitim modeli ile web servisin çalıştırılması sonucu elde edilen kullanıcı bazlı öneri listesinin web sitesindeki örnek kullanımı

ÖNERİ MODELİNİ SEÇİNİZ : Train Matchbox Content-Based Colloborative

ÜRÜN SEÇİNİZ :

ÜRÜN ÖNER

ÖNERİLEN ÜRÜNLER (Seçilen Kullanıcı/Model : F***** Y***** / Colloborative)

- 1--Apple iPhone SE 16GB
- 2--Apple iPhone 6 32 GB
- 3--Samsung Galaxy A5 2017
- 4--Samsung UE-55MU7000 55" 139 Ekran Uydu Alıcılı 4K Ultra HD Smart LED TV
- 5--Kingshark KS-523 HD-DVR 2.5" Hareket Sensörlü Çift Lensli Araç İçi Kamera

Şekil 4.10 Collobarative eğitim modeli ile web servisin çalıştırılması sonucu elde edilen kullanıcı bazlı öneri listesinin web sitesindeki örnek kullanımı

En dođru ve mantıklı önerilerin Train Matchbox öneri modeli seçildiğinde listelendiđi görölmüştür. Çünkü bu model diđer iki modelin birleřtirilmiř halidir ve veriler üzerinde hem kullanıcı bazlı hem içerik bazlı ilişkilendirme yapmaktadır.

Bu web servis kullanılarak hazırlanan web sitesinde, bir ürün seçildiğinde bu ürünle ilişkili diđer ürünler kullanıcıya önerilmiştir.

5. SONUÇ VE ÖNERİLER

Firmalar; küresel pazarda varlığını koruyabilmek ve çalışma potansiyeli yüksek personel yaratma adına, önemli yatırımlar yapmaktadırlar. Bu nedenle insandan en verimli şekilde faydalanma görüşü hakimdir. Dünya ekonomisi; firmaların rekabet direncini sağlamasının müşteri memnuniyetine dayandığını göstermektedir. Özellikle e-ticaret firmalarının müşteri memnuniyetini sağlamasında web sitelerindeki öneri sistemleri artık büyük bir rol oynamaktadır. Yapılan incelemelerde firmaların satışlarında gözle görülür oranda katkı sağladığı sonucuna ulaşılmıştır.

E-ticaret sitelerindeki öneriler kullanıcılar için büyük bir önem arz etmektedir. Bu tip öneriler, kullanıcıların sitelerde gezinirken daha hızlı karar vermesine yardımcı olur. Bir e-ticaret sitesine bakıldığında seçilen ürünle birlikte sıklıkla birbirleriyle ilişki öğeleri gösterdikleri öneriler görülmektedir. Öneri sistemleri ile kullanıcıya "Bu ürünü kullananlar, aynı zamanda bu ürünü de aldı." şeklinde farklı ürün önerileri sunulabilmektedir. Bu sayede e-ticaret sitelerinin satış oranlarına ekonomik anlamda büyük katkı sağlaması hedeflenmektedir. Bu çalışmada, e-ticaret sitesi için bir bulut bilişim platformu üzerinde yapay öğrenme algoritmaları ile site müşterilerine inceledikleri ürün bilgisine göre yeni tavsiyelerde bulunarak, müşterinin ilgisinin çekilmesi ve satış oranının artırılması amaçlanmıştır. Çalışma kapsamında; öncelikli olarak e-ticaret ile ilgili web sitelerinden satış verilerinin alınması gerçekleştirilmiştir. Daha sonra alınan bu verilere bulut bilişim altyapısı üzerinde yapay öğrenme algoritmaları uygulanarak bir öneri modeli elde edilmiştir. Bu model bir web servis yapısına dönüştürülerek herhangi bir web sitesinde kullanılabilir hale getirilmiştir.

Öneri yöntemleri kullanılarak elde iki farklı kullanıcı için elde edilen öneriler Çizelge 5.1 de verilmiştir.

Çizelge 5.1. Üç Farklı Öneri Yöntemi İçin Elde Edilen Bazı Öneri Sonuçları

Kullanıcı	Ürün 1	Ürün 2	Ürün 3	Ürün 4	Ürün 5	Yöntem
F**** Y****	Apple iPhone SE 16GB	Apple iPhone 8 64 GB	Apple iPhone X 64 GB	Samsung Galaxy S8 Plus	SJCAM SJ6 Legend 4K Aksiyon Kamerası - Siyah	Train Matchbox
P**** A****	Lumberjack A337433 3 Snow Uc Gunluk cocuk %100 Su Gecirmez Lastik Kar Botu	Eksen sehIrde Yasam Ahşap Puzzle / 35 Parça	Korkmaz A359 DemkolIk ElektrIkli cayıdanlık - Siyah	Neva Toys Süper Arabalar	LearnIng Toys Three Shape Small TraIns	Train Matchbox
F**** Y****	Apple iPhone SE 16GB	Apple iPhone 8 Plus 64 GB	HP 14-BS006N T Intel Celeron N3060 4GB 500GB Freedos 14" Taşınabilir Bilgisayar 2BS99EA	Xiaomi Mi Note 3 64 GB	Apple iPhone X 64 GB	İçerik Tabanlı
P**** A****	Lumberjack A337433 3 Snow Uc Gunluk cocuk %100 Su Geçirmez Lastik Kar Botu	Jump 16704 içi Termal Kürklü Fermuarlı Kız cocuk Bot Ayakkabı	Grisport 11520 Dakar Unisex Trekking Deri Bot	Sozzy Oyuncaklı Çıngıraklı Çorap Patik Ve El Bilekliği	Faber-Cas tell 188920 Silgi 7089/20 Siyah	İçerik Tabanlı
F**** Y****	Apple iPhone 6 32 GB	Apple iPhone 8 64 GB	Samsung Galaxy A5 2017	Samsung UE-55MU 7000 55" 139 Ekran Uydu Alıcılı 4K Ultra HD Smart LED TV	Kingshark KS-523 HD-DVR 2.5" Hareket Sensörlü Çift Lensli Araç İçi Kamera	İşbirlikçi

PINAR AYDIN	Lumberjack	üçer Ahşap	LearnIng	UK Polo	Korkmaz	İşbirlikçi
	A337433 3 Snow Uc Gunluk cocuk %100 Su Geçirmez Lastik Kar Botu	Eğitici Oyuncak Labirent Koordinasyon Oyunu	Toys Geometrika l Shape BuildIng Block	Club Outdoor Erkek Bot - Siyah	A359 Demkolik Elektrikli çaydanlık - Siyah	

En doğru ve mantıklı önerilerin Train Matchbox öneri yöntemi seçildiğinde listelendiği görülmüştür.

Çalışmada kullanılan içerik tabanlı, işbirlikçi ve hibrid (Train Matchbox) yaklaşımları kullanılarak çalıştırıldığında oluşan değerlerin karşılaştırma tablosu Çizelge 5.2 de verilmiştir.

Çizelge 5.2. Kullanılan Eğitim Modellerinin Öneri Sistemindeki Sonuçlarının Karşılaştırılması

Öneri Modeli	MAE	RMSE	NDCG
Train Matchbox	1.54	2.16	0.97
İşbirlikçi Filtreleme	2.24	2.85	0.75
İçerik Tabanlı Filtreleme	2.89	2.95	0.71

Çizelge 5.2.'de de görüldüğü üzere en başarılı model olarak Train Matchbox modeli görülmektedir. Çünkü bu model diğer iki modelin birleştirilmiş halidir ve veriler üzerinde hem kullanıcı bazlı hem içerik bazlı ilişkilendirme yapmaktadır.

KAYNAKLAR

- Azure AI Gallery, 2010, Solution, [Ziyaret Tarihi:11 Ağustos 2017].
- Barnes J., 2015, Microsoft Azure Essentials: Azure Machine Learning, Microsoft Press.
- Bobadilla, J., Hernando, A., Ortega, F., and Bernal, J., 2011, A framework for collaborative filtering recommender systems. *Expert Systems with Applications*, 38(12), 14609-14623.
- Ceviz Bilgi Teknolojileri, 2014, Microsoft Azure Machine Learning–Örnek Deney ve MVC Arayüzü, <http://www.cevizbilgi.com.tr/microsoft-azure-machine-learning-ornek-deney-ve-mvc-arayuz>, [Ziyaret Tarihi:10 Mayıs 2017].
- Choi, S. H., Kang, S., and Jeon, Y. J., 2006, Personalized recommendation system based on product specification values. *Expert Systems with Applications*, 31(3), 607-616.
- Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., and Sampath, D., 2010, The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296). ACM.
- Dooms, S. Pessemier, T.D., Martens, L., 2013, MovieTweatings: a Movie Rating Dataset Collected From Twitter. *Workshop on Crowdsourcing and Human Computation for Recommender Systems, CrowdRec. RecSys*.
- Gronlund C., Whitlatch K., Nevil T., Ericson G., Franks L., McKay P., 2014, What is Machine Learning on Azure?, Microsoft Docs.
- Gupta S., 2016, Cloud Computing Technologies Overview and Comparison, Master Thesis, San Diego State University.
- Hepsi Burada., 2018, Tüm Ürünler, <http://hepsiburada.com>, [Ziyaret Tarihi:01 Ekim 2018].
- Işık, M , Dağ, H., 2017, An Effective Recommender Model For E-Commerce Platforms. *Mugla Journal of Science and Technology*, 3 (2), 143-149. DOI: 10.22531/muglajsci.357313
- Jannach, D. Zanker, M. Felfernig, A. and Friedrich, G., 2011, *Recommender systems: an introduction*: Cambridge University Press, New York.
- Johnston B., Kim K., Nevil T., Makadia A., Steen H., 2016, Create an Azure Search index using the .NET SDK, Microsoft Azure Documentation.

- Kim, J. K., Kim, H. K., Oh, H. Y., and Ryu, Y. U., 2010, A group recommendation system for online communities. *International Journal of Information Management*, 30(3), 212-219.
- Krishna, P. V., Misra, S., Joshi, D., and Obaidat, M. S., 2013, Learning automata based sentiment analysis for recommender system on cloud. In *Computer, Information and Telecommunication Systems (CITS), 2013 International Conference on* (pp. 1-5). IEEE.
- Laudon, K. C., and Traver, C. G., 2013., *E-commerce*. Pearson.
- Liam C, 2016, Using Machine Learning and Azure Search to improve conversion rates through recommendations, Microsoft Azure Documentation.
- Linden, G., Smith, B., and York, J., 2003, Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, (1), 76-80.
- Machine Learning Repository., 2010, Statlog (German Credit Data) Data Set, [http://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)), [Ziyaret Tarihi:11 Ağustos 2017].
- Machine Learning Studio, 2010, Experiments, [Ziyaret Tarihi:11 Ağustos 2017].
- Marx, P. Hennig-Thurau, T. and Marchand, A., 2010, "Increasing consumers' understanding of recommender results: a preference-based hybrid algorithm with strong explanatory power", *Proceedings of the Fourth ACM Conference on Recommender Systems*, 26 - 30 September 2010, New York, 297-300.
- Microsoft Azure Platform, <https://azure.microsoft.com/tr-tr/documentation/> [Ziyaret Tarihi:1 Eylül 2016].
- Microsoft Azure Platform, Train Matchbox Recommender , <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/train-matchbox-recommender> [Ziyaret Tarihi:1 Eylül 2016].
- Mild, A., and Reutter, T., 2003, An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data. *Journal of Retailing and consumer Services*, 10(3), 123-133.
- Özdemir, A. F., Yıldıztepe, E., ve Binar, M., 2010, İstatistiksel yazılım geliştirme ortamı: R. XII. Akademik Bilişim Konferansı, 10-12.

- Recommender and introduction systems,
http://www.recommenderbook.net/media/Recommender_Systems_An_Introduction_Chapter07_Evaluating_recommender_systems
 [Ziyaret Tarihi:2 Ekim 2018].
- Ricci, F. Rokach, L. and Shapira, B., 2015, Recommender Systems Handbook, Second Edition, Springer, New York.
- Sammut, C. ve Webb, G.I., 2017, Encyclopedia of machine learning and data mining, Springer, New York.
- Sarwar, B. Karypis, G. Konstan, J. and Riedl, J., 2001, "Item-based collaborative filtering recommendation algorithms", Proceedings of the tenth international conference on World Wide Web - WWW '01, New York., 285-295.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., 2000, Application of dimensionality reduction in recommender system-a case study (No. TR-00-043). Minnesota Univ Minneapolis Dept of Computer Science.
- Schafer, J. B., Konstan, J. A., and Riedl, J., 2001, E-commerce recommendation applications. Data mining and knowledge discovery, 5(1-2), 115-153.
- Soon, W. M., Ng, H. T., Lim, D. C. Y., 2001, A machine learning approach to coreference resolution of noun phrases. Computational linguistics, 27.4: 521-544.
- Stern, David H., Ralf Herbrich, and Graepel T., 2009, "Matchbox: large scale online bayesian recommendations." Proceedings of the 18th international conference on World wide web. ACM.
- Tataru, P., Sand, A., Hobolth, A., Mailund, T., Pedersen, C.N.S., 2013, Algorithms for Hidden Markov Models Restricted to Occurrences of Regular Expressions, Biology 2013, 2(4), 1282-1295.
- Wang, Y. F., Chuang, Y. L., Hsu, M. H., and Keh, H. C., 2004, A personalized recommender system for the cosmetic business. Expert Systems with Applications, 26(3), 427-434.
- Wang, Y., Dai, W., and Yuan, Y., 2008, Website browsing aid: A navigation graph-based recommendation system. Decision Support Systems, 45(3), 387-400.
- Xiang, Z., Kim, S. E., Hu, C., and Fesenmaier, D. R., 2007, Language representation of restaurants: Implications for developing online recommender systems. International Journal of Hospitality Management, 26(4), 1005-1018.

- Zhang, X., and Wang, H., 2005, Study on recommender systems for business-to-business electronic commerce. *Communications of the IIMA*, 5(4), 8.
- Zhao, W. X., Li, S., He, Y., Chang, E. Y., Wen, J. R., and Li, X., 2016, Connecting social media to e-commerce: Cold-start product recommendation using microblogging information. *IEEE Transactions on Knowledge and Data Engineering*, 28(5), 1147-1159.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı: Ahmet AYAN
Uyruğu: T.C
Doğum Yeri ve Tarihi: Konya 01.09.1992
Telefon: 0538-771-1545
Faks:
e-mail: ahmetayan@selcuk.edu.tr

EĞİTİM

Derece	Adı, İlçe, İl	Bitirme Yılı
Lise:	Hekimoğlu Anadolu Lisesi Selçuklu/Konya	2010
Üniversite:	Selçuk Üniversitesi Selçuklu/Konya	2014
Yüksek Lisans:		
Doktora:		

İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
2014-(Halen Devam Ediyor)	Selçuk Üniversitesi	Bilgisayar Mühendisi

UZMANLIK ALANI

- **Programlama Dilleri:** C, C++, C#, JavaScript, Java
- **İşletim Sistemleri:** Windows, Linux

YABANCI DİLLER

- İngilizce (Okuma: İyi, Konuşma: İyi, Yazma: İyi)

YAYINLAR

Ahmet Ayan, Ali Osman Çıbıkdiken, Bulut Bilişim Platformunda Yapay Öğrenme ile Film Öneri Sistemi (Ankara/Türkiye : ISMSIT, 2018).