

Yükseköğretimde Öneri Sistemlerine Dayalı Ders Seçme Modeli

Bülent Batmaz

DOKTORA TEZİ

İstatistik Anabilim Dalı

Kasım 2018

Course Selection Model Based On Recommendation Systems For Higher Educaiton

Bülent Batmaz

DOCTORAL DISSERTATION

Department of Statistics

November 2018

Yükseköğretimde Öneri Sistemlerine Dayalı Ders Seçme Modeli

Bülent Batmaz

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Yönetmeliği Uyarınca
İstatistik Anabilim Dalı
İstatistik Bilgi Sistemleri Bilim Dalında
DOKTORA TEZİ
Olarak Hazırlanmıştır

Danışman: Prof. Dr. Veysel Yılmaz

Kasım 2018

ONAY

İstatistik Anabilim Dalı Doktora öğrencisi Bülent Batmaz'ın DOKTORA tezi olarak hazırladığı “Yükseköğretimde Öneri Sistemlerine Dayalı Ders Seçme Modeli” başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek oy birliği ile kabul edilmiştir.

Danışman : Prof. Dr. Veysel Yılmaz

İkinci Danışman : -

Doktora Tez Savunma Jürisi:

Üye : Prof. Dr. Veysel Yılmaz

Üye : Prof. Dr. Zeki Yıldız

Üye : Prof. Dr. Necmi Gürsakal

Üye : Prof. Dr. Ali Ekrem Özkul

Üye : Dr. Öğr. Üyesi. Özer Özaydın

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun tarih ve
..... sayılı kararıyla onaylanmıştır.

Prof. Dr. Hürriyet ERŞAHAN
Enstitü Müdürü

ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. Veysel Yılmaz danışmanlığında hazırlamış olduğum “Yükseköğretimde Öneri Sistemlerine Dayalı Ders Seçme Modeli” başlıklı DOKTORA tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim.
09/11/2018

Bülent BATMAZ
İmza

ÖZET

Günümüzde, internetin kişisel olarak hayatımıza girmesi ile birlikte ulaşılabilir olan bilgi hacmi günden güne katlanarak büyümektedir. Ancak, bir şahsın çok büyük miktarlardaki verileri tek başına kümeleyip bu verilerden yararlı bilgiler elde etmesi çok zor olabilir. İnsanların daha doğru ve ilgili bilgiye ulaşma için son dönemlerde “Bilgi Filtreleme Sistemleri” kavramı tartışılmaya başlamıştır. Bilgi Filtreleme Sistemlerinin amacı; kullanıcı için gereksiz ve istenmeyen verilerden saf bilgiyi otomatik olarak sunmaktır. Bu bağlamda da, “Öneri Sistemleri” ortaya çıkmıştır. Öneri sistemlerinde temel amaç, insanların beğenisini tahmin edip onlara bu beğenileri doğrultusunda önerilerde bulunmaktır. Bu tezde Anadolu Üniversitesi İİBF öğrencileri için gelecek dönemlerde alacakları zorunlu ve seçmeli derslerin başarı notlarını tahmin ederek ders seçiminde yardımcı olacak öneri sistemi modeli oluşturulmaya çalışılmıştır. Öneri modelinde öğrenci-öğrenci ve ders-ders benzerlikleri tekniklerinden yararlanılmıştır. Bu tekniklerde benzerlik ölçütü olarak yaygın olarak kullanılan pearson korelasyon değerinden yararlanılmıştır. Benzerlik ölçütü yardımıyla model ve memory (hafıza) tabanlı öneri algoritmaları kullanılmıştır. Hafıza tabanlı işbirlikçi filtrelemede ürün tabanlı öneri sistemi teknikleri, Model tabanlı işbirlikçi filtreleme tekniklerinde ise; k-Ortalama, K En yakın Komşu ve Tekil değer ayrışımı yaklaşımlarıyla öğrenci başarı notu tahmini yapılmıştır. Uygulamada dört farklı algoritma sonucunda elde edilen sonuçlar karşılaştırılmıştır. Sonuç olarak ders seçiminde öneri sunmada yararlanılacak öğrenci başarı notlarını gizil örüntüyü belirleyen, tekil değer ayrışımı tekniği ile tahmin edilmiştir. Uygulanan Öneri sistemleri algoritmaları için en iyi tahmin doğruluk ölçütü olan (MAE) mutlak ortalama hata değerine Tekil değer ayrışımı algoritması sonucunda ulaşılmıştır.

Anahtar Kelimeler: Öneri Sistemleri, Öğrenme Analitiği, İşbirliğine dayalı filtreleme, Tekil değer ayrışımı, Kümeleme, K en yakın komşuluk

SUMMARY

Nowadays, the volume of the available information is increasing day by day with the inclusion of internet in to our personal lives. However, it may be very difficult for a person to cluster big data and obtain/interpret useful information out of them. In order to reach more accurate and relevant information, recently the discussion of “Information Filtering Systems” has been started. The purpose of Information Filtering Systems is to provide the user with pure information from unwanted and cluttered data automatically. In this respect, “The Recommendation Systems” have been emerged. The main purpose of the recommendation systems is to estimate the likes of people in order to make recommendations accordingly. In this thesis, a recommendation system model is purposed to help the students, of Anadolu University Faculty of Economics and Administrative Science, in the selection of the courses to be taken in the future by predicting the grades of future compulsory and elective courses offered in students’ curriculum. In the recommendation system model, the student-student and course-course similarities are used. In order to reflect the similarity, the well-known Pearson’s correlation coefficient is used. Model and memory-based recommendation algorithms are used via similarity measure. In memory based collaborative filtering, product-based recommendation systems are used; In model based collaborative filtering, k-means, k nearest neighbours, and singular value decomposition is used to predict the students’ grade scores. In the application part of the study, the results of four different algorithms are studied in detail. As a result, the singular value decomposition, because of its ability to show the latent pattern, is used as the prediction method for students. Among the recommendation algorithms used in this study, the minimum mean absolute error indicator is achieved by the singular value decomposition.

Keywords: Recommendation Systems, Learning Analytics, Collaborative Filtering, Singular value decomposition (SVD), Clustering, K nearest neighbours

Teşekkür

Her şeyden önce, Doktora tez çalışmamı değerlendirmeyi kabul eden Jüri üyelerine teşekkür ediyorum. Tezimi yazarken geçen süre boyunca yönlendirmeleri ve bana verdiği destekleri için danışmanım Sayın Prof. Dr. Veysel Yılmaz'a çok teşekkür ederim. Her ne kadar farklı illerde yaşamamıza rağmen e-mail, telefonla veya yüz yüze görüşmelerimizde, tavsiyeleri ile cesaretlendirici ve yönlendirici olan, çalışmam için önemli bir itici güç oluşturan Sayın Prof. Dr. Necmi Gürsakal hocama teşekkür ederim. Tez izleme toplantılarında olumlu tavsiyeleri ve pek çok yapıcı tartışmaları için Sayın Prof. Dr. Zeki Yıldız hocama çok teşekkür ederim. Tezde öğrenme analitiği bölümü ve modelleme sürecinde önemli katkıları olan Sayın Prof. Dr. Ali Ekrem Özkul hocama teşekkür ederim. Ayrıca değerli katkıları için İstatistik Bilgi Sistemleri A.B.D. başkanı Sayın Dr. Öğretim Üyesi Özer Özaydın hocama teşekkür ederim. Doktora tezin ne zaman bitti diye sürekli soran hocalarıma, meslektaşlarıma, iş arkadaşlarıma, yakın dostlarıma teşekkür ederim. Burada tek tek sıralayamadığım, emeği geçen herkese teşekkür ederim.

En önemlisi tez çalışmalarım boyunca, onlara ayıracağım zamanlardan çaldığım, Sevgili eşim Serap'a, canım oğlum Başar'a ve canım kızım Deniz'e sonsuz teşekkür ederim.

Bülent BATMAZ

İÇİNDEKİLER

Sayfa

ÖZET	vi
SUMMARY	vii
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
ŞEKİLLER DİZİNİ	xiii
ÇİZELGELER DİZİNİ	xvi
SİMGELER VE KISALTMALAR DİZİNİ	xvii
1. GİRİŞ VE AMAÇ	1
2. LİTERATÜR ARAŞTIRMASI	8
2.1. Ders Öneri Sistemleri.....	11
2.2. Ders Öneri Sistemlerinde Kestirimsel Analitik	12
2.3. Ders Öneri Sistemlerinde Öğrenme Analitiği	13
2.4. Ders Öneri Sistemlerinde Öğrenme Analitikleri Süreci	15
2.5. Ders Öneri Sistemleri Örnekleri	17
3. MATERYAL VE YÖNTEM	22
3.1. Materyal.....	22
3.2. Yöntem	24
3.3. Öneri Sistemleri.....	25
3.4. Öneri Sistemleri Bileşenleri.....	31
3.4.1. Öneri sistemlerinde veri toplama süreci.....	32
3.4.2. Öneri Modeli	34
3.4.3. Öneri Süreci	35
3.4.4. Çevrimiçi (Online) Modüller	35
3.4.5. Kullanıcı Ara yüzü	35

İÇİNDEKİLER (devam)

Sayfa

3.5. Öneri Sistemlerinde Kullanılan Yöntemler	35
3.5.1. İşbirliğine dayalı filtreleme yöntemi.....	36
3.5.2. İçerik tabanlı filtreleme	36
3.5.3. Hibrit öneri sistemleri	37
3.5.4. Demografik yöntemler	38
3.6. İşbirliğine Dayalı Filtreleme İşlem Adımları	38
3.7. İşbirliğine Dayalı Filtrelemede Problem Tanımı.....	40
3.8. Benzerlik Ve Uzaklık Hesaplamada Kullanılan Ölçütler	42
3.8.1. Öklid (euclidean) uzaklığı	43
3.8.2. Pearson korelasyon katsayısı	44
3.8.3. Kosinüs (cosine) Ölçüsü.....	45
3.9. İşbirliğine Dayalı Filtreleme Algoritmaları.....	46
3.9.1. Bellek-Tabanlı Algoritma.....	47
3.9.2. Model-tabanlı algoritma	48
3.9.3. Hibrit (melez) algoritma	51
3.10. İşbirliğine Dayalı Filtrelemede Öneri Hesaplama Temel Yaklaşımları.....	51
3.11. İşbirliğine Dayalı Filtrelemede Kullanılan İstatistiksel Teknikler.....	51
3.11.1. Kümeleme analizi	52
3.11.2. K-ortalamlar algoritması.....	54
3.11.3. En yakın komşuluk (KNN) algoritması.....	56
3.11.4. Matris ayrışımı algoritmaları	58
3.11.5. Temel bileşenler analizi.....	59
3.11.6 Tekil değer ayrışımı	60

İÇİNDEKİLER (devam)

Sayfa

3.12. Öneri Kalitesi	65
3.13. Öneri Sistemleri Değerlendirme Ölçümleri	68
3.13.1. Ortalama Mutlak Hata (MAE).....	69
3.13.2. Kök Ortalama Mutlak Hata (RMSE).....	69
3.14. Öneri Sistemleri Ve Karşılaşılan Problemler	70
3.14.1. Doğruluk problemi	71
3.14.2. Ölçeklenebilirlik problemi.....	71
3.14.3. Nadirlik (sparsity) problemi	71
3.14.4. Soğuk başlangıç problemi	72
3.14.5. Olağandışı kullanıcı problemi	73
3.14.6. Sınırlı içerik analizi problemi.....	73
3.14.7. Kapsam geçerliliği problemi	73
3.15. Yükseköğretimde Öneri Sistemlerine Dayalı Ders Seçme Modeli.....	74
3.16. Yükseköğretimde öneri sistemlerine dayalı ders seçme modeli bileşenleri	75
3.16.1. Ders öneri modeli veri toplama süreci.....	77
3.16.2. Ders öneri modeli	77
3.16.3. Ders Öneri Süreci	79
3.16.4. Çevrimiçi: Ders seçimi öneri yönetim sistemi	79
3.16.5. Ders seçimi öneri yönetim sistemi: Ara yüzü.....	79
3.17. Yükseköğretimde işbirliğine dayalı ders seçme öneri modeli.....	80
3.18. İşbirliğine dayalı filtreleme ile ders başarı notlarını tahmin süreci.....	84

İÇİNDEKİLER (devam)

Sayfa

4.	BULGULAR VE TARTIŞMA	86
4.1.	İktisat Bölümü Öğrencileri İçin Seçmeli Ders Öneri Algoritma Sonuçları.....	87
4.1.1.	K-NN (K En Yakın Komşu) Algoritması Sonuçları	87
4.1.2.	K-Means (K Ortalamalar) Algoritması Sonuçları	90
4.1.3.	Tekil Değer Ayırışımı (SVD) Algoritması Sonuçları.....	92
4.1.4.	Ders Benzerliğine Dayalı Algoritma Sonuçları.....	94
4.2.	İşletme Bölümü Öğrencileri İçin Seçmeli Ders Öneri Algoritma Sonuçları.....	96
4.2.1.	K-NN (K En Yakın Komşu) Algoritması Sonuçları	96
4.2.2.	K Ortalamalar Algoritması Sonuçları.....	98
4.2.3.	Tekil Değer Ayırışımı (SVD) Algoritması Sonuçları.....	100
4.2.4.	Ders Benzerliğine Dayalı Algoritma Sonuçları.....	102
4.3.	İktisat Bölümü Öğrencileri İçin Seçmeli Ders Öneri Algoritma Sonuçları.....	108
4.4.	İktisat Bölümü Öğrencileri İçin Seçmeli Ders Öneri Algoritma Sonuçları.....	109
5.	SONUÇ VE ÖNERİLER	110
	KAYNAKLAR DİZİNİ.....	115
	EK AÇIKLAMALAR	123
	Ek Açıklama - A: Knn matlab iktisat bölümü algoritma.....	124
	Ek Açıklama - B: Item Base Matlab Algoritma	128
	Ek Açıklama - C: Tekil Değer Ayırışımı (Svd) Algoritma	132
	ÖZGEÇMİŞ	136

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
1.1. Ders seçiminde öğrenci labirenti	3
1.2. Ders seçiminde ders adı ve kodlarından oluşan kelime bulutu	4
2.1. Öğrenme analitikleri süreci	16
3.1. Veri hazırlık aşamasında verilerin birleştirilmesi ve düzenlemesi	22
3.2. Hazırlık aşamasında verilerin matrise dönüşümü.....	23
3.3. Öneri sistemleri yöntemleri	27
3.4. Ders öneri sistemi temel bileşenleri.....	31
3.5. Ders seçme öneri sistemi için problemin şematik gösterimi	41
3.6. K-ortalamlar algoritması kümeleme adımları görselleştirilmesi.	56
3.7. Tekil değer ayrışımının matris ve vektör şeklindeki gösterimi	65
3.8. Yükseköğretimde öneri sistemi temel bileşenleri.....	76
3.9. Yükseköğretimde İşbirliğine Dayalı Ders Seçme Öneri Modeli.....	81
3.10. Yükseköğretimde Öneri Sistemi İle Öğrenci Ders seçim Süreci	83
3.11. İşbirliğine dayalı filtreleme ile ders başarı notlarını tahmin süreci	85
4.1. İktisat bölümü, K-En Yakın Komşuluk Algoritması sonucu MAE değerleri	88
4.2. İktisat bölümü k en yakın komşu algoritması sonucunda tahmin edilen ders notları	89
4.3. İktisat bölümü k-ortalama algoritması sonucu mae değerleri	90

ŞEKİLLER DİZİNİ (devam)

<u>Sekil</u>	<u>Sayfa</u>
4.4. İktisat Bölümü K-ortalama algoritma tahmin değeri.....	91
4.5. İktisat bölümü, tekil değer ayrışımı algoritması sonucu mae değerleri.....	92
4.6. İktisat bölümü ders notu tekil değer ayrışımı algoritması (k=16) tahmin değeri	93
4.7. İktisat bölümü, ders benzerliği algoritması sonucu mae değerleri	94
4.8. İktisat bölümü ders benzerliği ders notu tahmin değeri	95
4.9. İşletme bölümü, k-en yakın komşuluk algoritması sonucu mae değerleri	96
4.10. İşletme bölümü, k-en yakın komşuluk algoritması ders notu tahmin değerleri	97
4.11. İşletme bölümü, k-ortalama algoritması sonucu mae değerleri.....	98
4.12. İşletme bölümü K-ortalama algoritma sonucu tahmin ders notu	99
4.13. İşletme bölümü, tekil değer ayrışımı algoritması sonucu mae değerleri.....	100
4.14. İşletme bölümü tekil değer ayrışımı algoritması (k=12) tahmin ders notları	101
4.15. İşletme bölümü ders benzerliği algoritması sonucu MAE değerleri	102
4.16. İşletme bölümü ders benzerliğine algoritması sonucunda tahmin ders notları	103
4.17. İktisat bölümü tekil değer ayrışımı test verisi tahmin ders başarı notları grafiği. ...	104
4.18. İktisat bölümü verisetinden hesaplanan özdeğer – özvekök matrisi	105
4.19. İktisat bölümü verisetinden hesaplanan özdeğer – özvekök matrisi grafiği	105

ŞEKİLLER DİZİNİ (devam)

<u>Şekil</u>	<u>Sayfa</u>
4.20. İşletme bölümü tekil değer ayrışımı test verisi tahmin ders başarı notları grafiği. .	106
4.21. İşletme bölümü verisetinden hesaplanan özdeğer – özvekor matrisi	107
4.22. İşletme bölümü verisetinden hesaplanan özdeğer – özvekor matrisi grafiği.....	107
4.23. İktisat böümü tekil değer ayrışımı algoritması sonucu elde edilen tahmin başarı paunları ile ders seçme öneri örneği.....	108
4.24. İktisat böümü tekil değer ayrışımı algoritması sonucu elde edilen tahmin başarı punları ile ders seçme öneri örneği.	109

ÇİZELGELER DİZİNİ

<u>Cizelge</u>	<u>Sayfa</u>
3.1. Öneri Sistemleri Ve İşbirliğine Dayalı Filtrelemenin Literatüre Göre Tanımları.....	30
3.2. İşbirliğine dayalı Filtreleme Algoritmalarının Kullanımında Olumlu Ve Olumsuz Yönleri.....	46
3.3. Madde Tabanlı (User-İtem) Matrisi	48
3.4. Kullanıcı Tabanlı (User-İtem) Matrisi.....	50

SİMGELER VE KISALTMALAR DİZİNİ

Simgeler

Açıklama

U	Öğrenci; U_i , i'inci öğrenci olmak üzere, $i=\{1,2,\dots,n\}$
I	Ders; I_j , j'inci ders olmak üzere, $j=1=\{1,2,\dots,m\}$
R	Kestirilen başarı notu değeridir.
R(U, I)	Öğrenci geçmiş başarı notlarıyla R'yi kestirmektir.

Kısaltmalar

Açıklama

İF	İşbirliğine dayalı Filtreleme
CF	Collaborative Filtering
MAE	Ortalama mutlak hata
RMSEA	Ortalama mutlak hata karekökü
SVD	Singular value decomposition
TDA	Tekil değer Ayrışımı
K-Means	K-Ortalama kümeleme algoritması
KNN	K En yakın komşu algoritması
RARE	Recommender system based on association Rules (Birliktelik kuralına dayanan bir ders öneri sistemi)
PEL-IRT	Personalized e-learning system based on Item Response Theory (Öğe yanıt teorisine dayalı kişiselleştirilmiş e-öğrenme sistemi)

1. GİRİŞ VE AMAÇ

Yükseköğretimde bazı özel bölümler haricinde ders müfredatı genellikle esnekler. Bu esnek müfredat, zorunlu dersler, mesleki zorunlu dersler veya seçmeli derslerden oluşmaktadır. Öğrenciler her dönemin başında yönetmeliklerle belirlenen ders kredi kotası kadar zorunlu ve seçmeli dersler arasından ders seçimi yaparlar. Bu ders seçim sürecinde, öğrencilerin bazıları fazladan zorunlu dersler dışında bir üst deönemden ya da seçmeli ders seçim yaparak ders kredilerini artırılabilirler. Bazı öğrencilerde, daha sonra almak üzere bazı dersleri seçmeyerek ders kredi miktarını azaltabilirler.

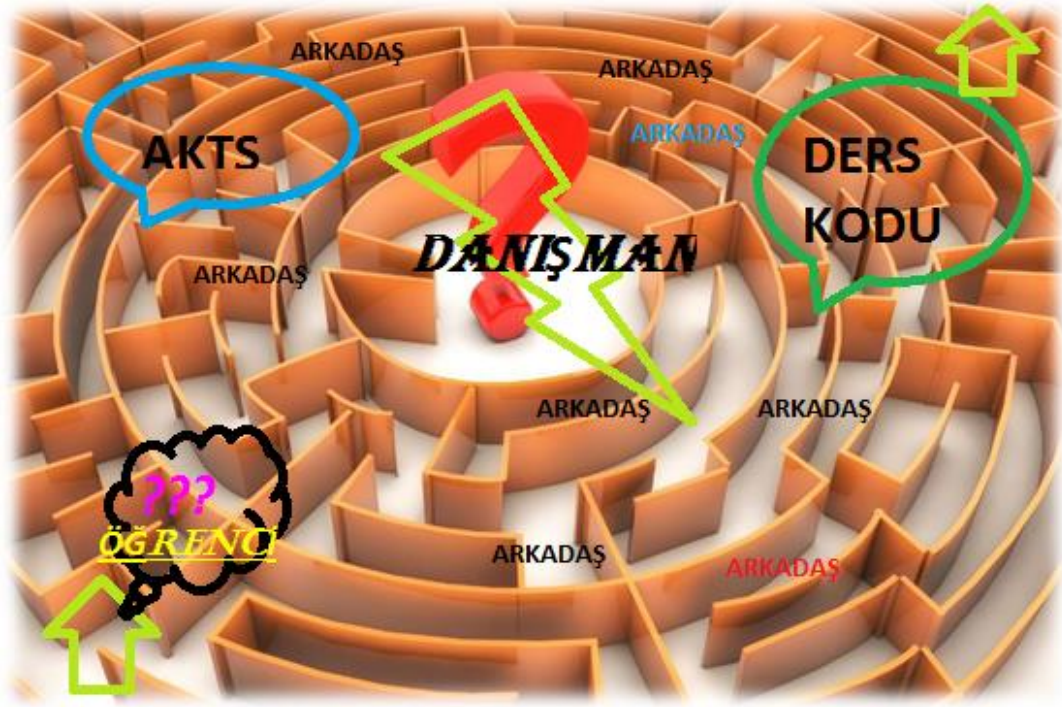
Yükseköğretimde öğrencilere ders seçimlerinde yardımcı ve yönlendirici olarak veya akademik olarak izlenmeleri amacıyla danışman öğretim elamanı görevlendirilir. Ancak gerçekte ders seçim onayı dışında akademik izleme ve bu akademik başarıya göre ders seçiminde yönlendirme yapılmadığı görülmektedir. Danışman öğretim elemanlarının çoğunluğu danışmanı oldukları öğrencilerin çoğunu çok iyi tanımazlar. Ayrıca danışmanların yoğun iş yükü nedeniyle danışmanlık hizmeti sadece ders çakışmalarını önlemek amacıyla öğrenci ders seçimini onaylama şeklinde yapmaktadırlar. Bir öğrencinin bir derse kaydolması koşulları yerine getirdiği sürece sadece onun kararına dayanır. Öğrenci ders seçmeden önce danışman öğretim elemanından akademik başarısına göre hangi dersleri alabileceği konusunda öneriler isteyebilir. Ancak öğrenciler bu önerileri nadiren istemektedirler ve genelde ders kayıtları öğrencinin bilgisine ve deneyimine dayandırılmaktadır (Vialardi, vd., 2009) Danışmanlar çok sayıda öğrenci ile yüklenmişlerdir, yeterli zamanları yoktur ve bazı öğrenciler danışmanların derslere ilişkin bilgi düzeylerinden memnun olmazlar.

Bu nedenle öğrencilerin akademik başarı beklentilerini karşılayacak, seçeceği dersten başarılı olup olmayacağı gibi önsel bilgilere ihtiyaç duyulur. Üniversitelerde öğrenim gören öğrenciler, bir ders seçmekle karşı karşıya kaldıklarında yüzlerce ders arasından seçim yapabilirler, bir anlamda bu iyi bir şeydir. Ancak öğrenciler açısından bu kadar çok sayıda derse ilişkin bilgi toplanmasını kolay olmadığını da kabul etmeliyiz. Bu önsel bilgiler genellikle arkadaş tavsiyesi niteliğindedir. Bir öğrenci genellikle ders seçim sürecinde tavsiyelerini istediği arkadaşları arasında kendince bir sınıflama yapar, bu

sınıflama, çalışkan ve derslerinde başarılı ya da derslerinde başarısız şeklindedir. Bu nedenle daha çok aynı bölümdeki arkadaşlarının en çok hangi dersi seçtiklerini öğrenmek ister, ya da öğrenci kendi ders başarısına yakın olan arkadaşlarında görüş ister. Bazen de daha önceden o dersi almış üst sınıftaki diğer öğrencilerin görüşlerini dikkate alır. Çünkü çoğunluğun fikri veya öngörüsü ortalamaya en yakın değer olarak karşımıza çıkmaktadır.

Topluluğun görüşünün ortalaması gerçeğe en yakın görüş “kalabalığın bilgeliği” kavramı olarak tanımlanır. Sosyal tabanlı ya da işbirliğine dayalı öneri sistemlerinde kalabalığın bilgeliği kavramı burada daha da önem kazanmaktadır. Öğrencilerin ders seçimlerinde kalabalığın gücünden faydalanmalarını sağlayacak, bu kalabalık topluluktan elde edilecek bilgiyi filtreleyip öğrenciye uygun bir saf bilgi şeklinde sunacak bir sisteme gereksinimi vardır.

Öğrenciler yüzlerce dersin arasından ders seçmek zorundadırlar. Bu konuda danışmanlar yardımcı olsa bile, uygulamada danışmanlara çok öğrenci yüklenmekte ve öğrenciler bazen danışmanların ders seçimi konusundaki bilgi düzeylerine güvenmemektedirler (Unelrød, 2011). Üniversiteler tarafından sunulan çok sayıda ders arasından, başarılı olacağı dersleri veya ilgi duyduğu dersleri seçmek öğrenci için oldukça zor bir olaydır. Üniversite ders tanıtımlarında derslerin içerikleri bulunsa da, öğrenciler bu derslerden ne ölçüde başarı sağlayacağını bilemez. Öğrencilerin bu memnuniyetsizlik durumlarını ve çok sayıda ders içerisinde ders seçmelerine ve bu seçecekleri derslere ilişkin önceden tahmin edilmiş performans (başarı notu) bilgisine dayalı bir öneri sistemine ihtiyaç vardır.



Şekil 1.1. Ders seçiminde öğrenci labirenti

Öğrencinin ders seçmedeki ilk hedefi seçeceği dersten başarılı olmaktır. Basit gibi görünen ders seçimi sürecinde öğrenciler bir Şekil (1.1)'de verilen labirentte dolaşırlar. Labirent Kelimesi TDK sözlüğünde “ Çıkış yeri kolay bulunamayacak kadar karışık koridorları olan yapı” olarak tanımlanmaktadır. Hatta bu tanımın yanında ünlü yazar Yakup Kadri Karaosmanoğlu'nun “Bu labirentte yolumu kaybedip gitmem işten değildi” (TDK, 2018) sözünü de örnek olarak vermektedir. Ünlü yazarın bu sözü ders seçimlerindeki belirsizlik ortamında kaybolan öğrencileri çok iyi anlatmaktadır. Çünkü öğrencilerin seçtikleri derslerden başarılı olup olamayacaklarını önceden bilemezler, genellikle kulaktan kulağa duyuma dayalı veya arkadaş grupları arasında tavsiye niteliğindeki bilgilere göre ders seçimi yapmaktadırlar.

Öğrenci ders seçme aşamasında, labirent içinde dolaşırken uygun verileri toplayamaz, elde edemez veya filtreleyip bilgiye dönüştürerek karar veremez. Bu verilerin yükseköğretim kurumları tarafından filtreleyerek bir ders öneri modelinin geliştirilmesi ve bir ders seçiminde tavsiye niteliğinde öğrenciye bilgi olarak sunulması gereklidir. Bir öğrencinin kendi ilgi ve yeteneklerine veya birkaç arkadaş tavsiyesine göre uygun dersleri seçerek bu derslerin tamamından yüksek başarıyı elde etmesi, öğrencinin ya çok çalışkan

olmasına veya birçok şans faktörünün birlikte olumlu yönde gerçekleşmesi ile mümkün olabilir. Her ne kadar ders programları bölüm veya program çıktıları göz önüne alınarak belirlenmiş olsa da, bir bölümde veya bir sınıftaki öğrencilerin hepsinin aynı başarıyı elde etmeleri nadirdir.

Bir öğrencinin seçeceği her ders için ise AA notu alması olasılığı oldukça küçüktür. Kısaca, bu durum öğrencinin Şekil 1.2.'de verilen dört yapraklı yoncayı bulmasına benzetilmektedir. Dört yapraklı yonca doğada mutasyon sonucu 1/10000 oranında görülmektedir.



Şekil 1.2. Ders seçiminde ders adı ve kodlarından oluşan kelime bulutu

(WordCloud R programlama)

Bu tez çalışmasında; öğrencilerin dönemlik ders seçimlerinde öneri sistemlerine dayalı ders seçme modeli oluşturulmuştur. Tezde; Yükseköğretimde kredili sistemde eğitim

gören öğrenciler için öneri sistemlerine dayalı ders seçme modeli oluşturulmasında öğrencilerin henüz seçmedikleri ders başarı puanları tahmin edilerek, tahmin algoritmaları karşılaştırılmıştır. İşbirliğine dayalı filtreleme tekniği ile tahmin edilen başarı puanları ile bir öneri sistemi geliştirilmesi öğrencilerin akademik hayatlarına yardımcı olacağı düşünülmektedir. İşbirliğine dayalı öneri sistemleri, kullanıcılara ürünler arasından seçmesi için öneriler yapan sistemlerdir. Bu sistemler benzer özelliklere sahip olan bireylerin, genelde aynı şeyleri seçtikleri düşüncesine dayanmaktadır. İşbirliğine dayalı öneri sistemleri, film, kitap, müzik, haber, ders seçimi gibi çok sayıda öneri sunan uygulama alanında yüksek derecede kabul görmüş ve iyi sonuçlar vermiştir.

Bu tez çalışmasında, kredili ders geçme sistemi uygulanan üniversitelerde, öğrencilere ders seçimlerinde de yardımcı olacak danışman öneri sistemi modeli oluşturulması amaçlanmıştır. Danışman öneri sistemi bir filtreleme tekniğidir (Venkatesan ve Thangadurai, 2016). Öğrencinin ihtiyacına ve ilgisine göre verileri filtreler ve onlara rehberlik eder. Öğrencilere tahmin değerine göre başarılı olacağı en uygun dersi öğrenciye önerir. Çünkü öğrenciler ders seçimi aşamasında (Bkz.Şekil 1.1) bir labirent içerisinde kendilerine yol ararlar. Öğrencilere bu labirentin içinden çıkmalarına ve hedeflerine erişimi sağlayacak bir yol göstericiye (Pathfinder) ihtiyaç vardır. Bu ihtiyacın karşılanmasında öğrencilerin o dersi daha önceden almış diğer öğrencilerin bilgisine, ders forumlarından elde ettikleri bilgilere benzer üretilecek bilgi karar vermelerine yardımcı olacaktır. Öğrencinin diğer derslerdeki performanslarına bağlı olarak yeni seçeceği dersleri önerme sürecinde yararlanılacak tahmine dayalı bilginin elde edilmesi araştırmanın çözülmesi gereken problemdir. Söz konusu problemin çözülmesinde öğrencilerin geçmiş ders başarı puanları ile henüz seçmedikleri ders başarı puanları tahmin edilerek, ders kaydı sistemine entegre edilmek üzere çevrimiçi ders seçiminde yol gösterici bir öneri sistemini modellemektir.

Tezde bu amaçla;

1. Literatür taraması,
2. Ders öneri sistemleri hakkında yapılmış çalışmaların bazılarını özetlemek,
3. Öneri sistemini tanıtmak,
4. Öneri sisteminin tarihsel gelişimini özetlemek,
5. Öneri sistemlerinin özetini vermek,
6. Öneri sistemleri yöntemleri ve algoritmaları tanıtmak,
7. Öneri sistemlerinde karşılaşılan sorunları özetlemek,

8. Öneri sistemleri teknikleri yardımıyla öğrencilerin ders başarılarını tahmin edilmesi,
9. Tahmin edilen ders başarı puanları ile ders seçiminde yol gösterici bir öneri sistemi modellemektir.

Tezin uygulama aşamasında veri olarak, Anadolu Üniversitesi İktisadi İdari Bilimler Fakültesi'nin İktisat ve İşletme bölümü öğrencilerinin 2009 ve 2014 dönemleri arasındaki geçmiş ders puan verileri kullanılmıştır. Tezde yararlanılan veriler 2015 yılında görev yapan Anadolu Üniversitesi Bilgisayar Araştırma ve Uygulama Merkezi (BAUM) Müdürünün izni ile BAUM'deki görevli eleman tarafından hazırlanarak mail yoluyla gönderilmiştir.

Uygulama aşamasında kullanılacak öğrenci geçmiş başarı notu verileri ile gelecekte seçebileceği dersler için başarı puanlarının tahmini yapılmıştır. Böylece gelecekte başarılı olacağı öngörülen derslerden bir ders öneri sistemi için gerekli olan başarı puanları tahmin eden algoritma sonuçları karşılaştırılarak tartışılmıştır.

Yükseköğretim öğrencileri için ders seçiminde yardımcı olacak başarı puanlarının tahmin edilmesinde işbirliğine dayalı filtreleme tekniklerinden yararlanılmıştır. Öneri sistemlerinde İşbirliğine dayalı filtreleme ile ders-ders benzerliğine dayalı teknikleri ile öğrenci benzerliğine dayalı kümeleme tabanlı K-Ortalamlar (k-Means) , En Yakın Komşu (KNN) ve boyut indirgeme ve gizil yapıyı belirlemeye yarayan Tekil Değer Ayrışımı (SVD) algoritmalarından yararlanılmıştır.

İşbirliğine dayalı filtreleme teknikleri ve kullanılan algoritmalar karşılaştırılarak bir Yükseköğretimde ders seçme modeli geliştirilmiştir. Yükseköğretimde ders seçme modelinde, öneri sistemleri tekniklerinden birisi olan işbirliğine dayalı filtreleme teknikleri yardımıyla öğrenciye en çok hangi derslerden daha başarılı olacağı tahmin edilecek ve öğrenciye başarılı olacağı dersler öneri olarak sunulacaktır. Bu sistemde tahmin modelinde geçmiş dönem derslerinden almış oldukları başarı puanları bilgilerinden yararlanılmıştır. Bu tez çalışmasında tıpkı film veya kitap öneri sistemlerindeki benzer derecelendirme öğrenci-kullanıcı (user) ve ders başarı notu (beğeni puanları gibi ele alınarak) yardımıyla henüz seçmedikleri derslerin başarı notları tahminleri yapılmıştır.

Elde edilen tahminler yüksek puandan küçüğe sıralanarak, öğrenciler için ders seçiminde alacakları tahmini notlar ile öneri örnek olarak sunulmuştur. Öneri sistemleri üretim sistemlerinde, büyük miktarda bilgiyi filtreleme amacıyla başarıyla kullanılmıştır. Öneri sistemlerinin en yaygın olarak kullanıldığı alanlar, web siteleri, haber, müzik, film, ders ve diğer eğlence medyası ürünleridir. Bu bağlamda tezde, film önermede kullanılan aynı teknikleri üniversiteler de ders önerme amacıyla kullanmak için modellenmiştir.

Yükseköğretimde öğrencilere seçmeli ders seçiminde öneri sistemleri yardımıyla ders seçiminde öneri sistemleri modeli oluşturmak bölümlerinden oluşmaktadır. Literatürde öneri sistemleri ve ders öneri sistemlerine ilişkin daha önce yapılan çalışmalarla ilgili bilgiler bir sonraki bölümde verilmiştir.

2. LİTERATÜR ARAŞTIRMASI

İletişim teknolojilerinde meydana gelen gelişmeler günün eğitim sistemlerini doğrudan etkilemektedir. Çünkü iletişim, öğrenen ile öğretene arasındaki başarılı bir eğitimin olmazsa olmazlarından biridir. Bir iletişimin oluşması için en az bir kaynak, bir alıcı ve bir mesaja ihtiyaç vardır. Eğer iletişimdeki bu mesaj eğitsel ise bunun için öğretmen ve öğrenciyle birlikte iletişimin oluşabileceği bir ortamda gereklidir. Geleneksel eğitim sistemlerinde bu ortamlar, sınıflar gibi öğrenen ile öğretene aynı ortamda bulunmasını gerektiren mekânlar olmuştur.

Öğrenen ile etkileşim, teknolojik gelişmelere bağlı olarak eğitimdeki yeni yaklaşımlar sonucunda farklı sınıf ortamlarının yaratılması sonucunu doğurmuştur. Bu yeni ortamlarda elde edilen akıllı veriler öğrenen ile öğretene arasındaki etkileşimde bilgi olarak karşılıklı akışa sahip olması gerekmektedir. Akıllı veri, gelişmiş uygulama teknolojileri (akıllı sensörler, akıllı dönüştürücüler, elektronik aygıtlar, gelişmiş işlemlere sahip donanımlar vb.) ile otomatik olarak elde edilen, sistematik olarak işlenen ve aktif paylaşım, ileri değerlendirmeler ve yorumlamalar için anlamlı bir formda sunulan veridir. Bu toplanan akıllı veriler bağlamında öğrenme analitiği kavramı burada daha da önem kazanmaktadır. Öğrenme analitiğinin Siemens tarafından yapılan tanımına göre öğrenme analitiği, "akıllı verilerin ve öğrencinin ürettiği verilerin kullanımı, enformasyonu ve sosyal bağlantılarını keşfetme amacıyla kullanılan modellerin analizleri ve öğrenme konusunda kestirim ve önerilerin üretilmesidir" (Siemens, 2010). Bu tanım öğrenme analitiği ile tahmine dayalı istatistik tekniklerin uygulandığı öneri sistemlerinin ve eğitsel veri madenciliğinin çok yakından ilişkili olduklarını göstermektedir (Manouselis, vd., 2012).

Öğrenciler, cinsiyet, dil, genetik kökenleri, etnik köken, sosyal ve kültürel çevreleri, bilişsel becerileri, önceki bilgi ve deneyimleri, akademik performans ve öğrenme ihtiyaçları ile güdülenmeleri açısından farklılıklar gösterirler. Öğrencilerde bu farklılıkların çoğu değişmezken, eğitim-öğretim etkileri sonucunda sosyal ve kültürel çevreleri ile yaşam koşulları sürekli olarak değişir. Bu değişen koşullarda eğitim ve öğretim etkisiyle, öğrencilerin öğrenme ihtiyaçları ve sahip olmak istedikleri kazanımlar açısından bir iyileşme süreci başlar. Mezun olduklarında öğrenim gördükleri alanda sahip oldukları kazanımlar onların gelecekteki yaşantılarını şekillendirecektir. Öğrencilerin, geleceğe ilişkin yön

belirlemeleri ve güdülenmeleri sonucunda eğitim-öğretim dönemlerinde ilgi duydukları veya sahip olmak istedikleri alanlardaki bilgi gereksinimlerinin artmasını sağlar. Öğrenme analitiği kapsamında, öğrencilere yatkın oldukları ve bilgilenmek istedikleri alanlarda ihtiyaçlarının karşılanması için öğrenme çabalarının izlenmesi ve yönlendirilmeleri gerekmektedir. Bu nedenle öğrencilere başarılarını artırmak amacıyla rehberlik ve danışmanlık hizmeti önem taşır. Öğrenme rehberlik ve danışmanlık hizmeti sürecinde başarıyı getirir, sözünün etkisi anlam kazanır. Böylece, öğrencilerin almış oldukları derslerden başarılı bir dönem geçirmelerine, onların akademik dönemde zaman kaybını önlediği gibi öğrenme güdüsünü artırarak daha fazla bilgi ile donanma istekleride artırır. Ayrıca yeni alanlara yönelmelerini böylece bireysel gelişimlerinde katkı sağlar. Bu nedenle öğrencilerin, eğitim öğretim dönemlerinde ilgi duydukları, yeteneklerine ve performanslarına göre başarılı olacakları dersler konusunda danışmanlık ve önerilere ihtiyaçları vardır.

Akademik danışmanlık, yükseköğretimde eğitim ve öğrenim amaçlarının yerine getirilmesinin ayrılmaz bir parçasıdır. Akademik danışmanlık yoluyla, öğrenciler yükseköğretim topluluklarının bir üyesi olmayı, öğrenciler olarak rolleri ve sorumlulukları, demokratik ve küresel bir topluluğun bir parçası olarak eleştirel düşünmeyi ve eğitimli birer vatandaş olmaya hazırlanmayı öğrenirler. Akademik danışmanlar, öğrencilerin eğitim ortamından neler bekledikleri ve bu ortamda yaşadıkları durumlar arasındaki uyumsuzluğa aracılık eder (Habley, 1981).

Fakülte öğrencilerinin, akademik danışmanlar ile sık sık iletişim kurması, öğrenci motivasyonunda ve derslere katılımında en önemli faktördür. Fakülte öğretim üyeleri arasından görevlendirilen akademik danışmanlık, öğrencilerin mezun olmak için gerekli şartları tamamlamalarına yardımcı olduğu görevleri içerir. Bu görevlerin yerine getirilmesinde, hem öğrencilerin hem de danışmanların önemli ölçüde planlanma yapmasını gerektirir. Akademik danışmanlar, teknoloji üniversite kampüslerinde daha yaygın hale geldikçe çeşitli fırsatlara, iyileştirmelere, sorunlara ve seçeneklere maruz kalmaktadır (Steele, vd., 2013).

Dünyadaki çeşitli üniversiteler ve diğer eğitim kurumlarının bazıları otomatik danışmanlık sistemlerini kullanmaktadırlar. Bu sistemler; müfredat oluşturma, öğrenci

başarısını izleme ve yönlendirme gibi hizmetlerin yanında ders seçiminde de öneri sunulmasında yardımcı olmaya katkıda buldukları için hem akademik danışmanlar hem de öğrenciler için yararlıdır. (Murray ve Renaud, 1995). Literatürde bazı üniversite ve eğitim kurumlarında çevrimiçi danışmanlık sistemleri oluşturulmuş ve öğrencilerinin dijital izleri sürekli takip edilmektedir. Bu Sistemler, öğretilere bir yetkilendirme ile bir öğrencinin not dökümünü görüntülemesine ve öğrenci ödevlerinin ve sınavlarının not girişine olanak tanır. Ayrıca bu sitemde öğrenci danışmanı olarak görevlendirilen bir kişi, öğrencinin aldığı tüm dersler ve o öğrencinin mezuniyetini bitirmesi için hala karşılanması gereken dersler hakkında bilgi sağlar. Web tabanlı danışmanlık sistemlerinde, öğrencinin (danışan) ve akademik yetkilinin (danışman) ilişkisini yükselten, erişimi kolay web tabanlı bir karar destek aracı 'Çevrimiçi Danışman' olarak tanımlanmaktadır. Çevrimiçi Danışman'ın görevi, personelin büro yükünü hafifletmek ve danışmanların öğrenci merkezli olmasını sağlamak, akademik danışmanların rutinin ötesinde öğrencilere yardımcı olmalarını sağlamak ve danışmanların öğrenci gelişimine odaklanmalarını sağlamaktır (Davenport, vd., 2010) Etkili danışmanlık, tavsiyeye ihtiyaç duyan öğrencinin plan ve hedeflerinin bilgisine bağlıdır. Her bir öğrenciye, görevin eksik kısımlarını başarmak için açık adımlar atmaları için tavsiye verilmelidir (Golumbic, vd., 1986). Başarılı akademik danışmanlık için çok sayıda beceri kritik öneme sahiptir. Bunlar arasında, fakülte-öğrenci iletişimi veya öğrencinin kurumla kişisel bir bağlantı kurması için bir fırsat sunan, danışan ve danışman arasındaki bire bir ilişki vardır (Heisserer ve Parette, 2002). Ayrıca, bir danışman kurumdaki akademik programlar ve müfredat gereksinimleri hakkında bilgi sahibi olmalıdır. Doğru ve bilgiye dayalı akademik rehberlik vermesi beklenir (Davenport, vd., 2010) Öneri sistemlerinde Öğrencinin ders seçmesinde tavsiyede bulunacak akademik danışmanlık sistemi için, öncelikle öğrencinin başarısının tahmin edilmesi gerekmektedir.

Öğrenci Danışmanlık hizmeti, danışman için akademik yaşamında önemli ve zaman alıcı bir çabadır. Öğrencilere danışmanlık hizmeti öğrencilerin gerekli şartları tamamlayarak mezun olmalarına yardımcı olacak görevler içermektedir. Bu nedenle danışman öğrencinin başarılı bir akademik dönem geçirmesinde önemli bir bileşendir. Danışman ve öğrenci bu süreçte birlikte akademik bir planlama yapmalarını gerektirir. Ancak danışman öğretim elemanlarının çoğunluğu danışmanlık yaptığı öğrencilerin çoğunu çok iyi tanımazlar, ya da yoğun iş yükünden ötürü ders seçiminde danışmanlık yalnızca sistemde ders çakışmalarını önlemek amacıyla ders seçimini onaylama şeklinde

yapılmaktadır. Akademik danışmanlık dört tür başlık altında tanımlanmaktadır. Literatürde kuralcı danışmanlık modeli, gelişimsel danışmanlık modeli, bütünlük danışmanlık modeli ve sorumlu (katılımcı) modeli olarak sıralanmaktadır.

Kuralcı danışmanlık modeli karar verme sorumluluğunun danışmanda olduğunu varsayar, öğrencilerin ders seçiminde, karar verici otorite danışmandır (Crookston, 1972). Gelişimsel danışmanlık modelinde ise, danışman öğrenciyi doğru kaynaklara yönlendiren kişidir. Böylece öğrenci ile danışman arasında sorumluluk paylaşılmaktadır. Öğrenci, kuralcı danışmanlık modeline göre gelişimsel danışmanlık modelinde daha özgür karar verme ve gelişimini yönlendirebilme yetisine sahiptir (Heisserer ve Parette, 2002). Bütünlük (entegre) danışmanlık modeli, kuralcı ve gelişimsel danışmanlık modellerinin elemanlarını birleştirir. Etkili akademik danışmanlık, katılımcı modeli olarak tanımlanır ve öğrenci gereksinimlerini tamamlamak ve öğrenci öz yeterliliğini arttırmak için öğrenci danışmanı ve öğretim üyesi arasında ilişki kurmayı içerir. (Crookston, 1972). Öğrencinin özellikle ders seçimi ve mezuniyet gereklerini yerine getirmek için öğrenci-danışman arasındaki etkileşimi oluşturma etkinliğini kapsar. Çevrimiçi danışmanlık modelinde öğrenci danışman etkileşimini arttırmak yerine öğrencinin bilgiye erişimini kolaylaştırmak için kullanılmalıdır (Dyarbrough, 2011).

2.1. Ders Öneri Sistemleri

Akademik danışmanlık öğrencinin, başarılı bir akademik süreç geçirmesinde önemli bir bileşendir. Danışmanın özellikle öğrencileri, yönlendirme çabası bir öneri niteliğindedir. Bu öneriler danışmanın bilgi deneyim, teknoloji kullanımı veya danışmanlık sisteminin ortamına bağlıdır. Eğitim ve teknoloji birlikte gelişen ve iç içe girmiş bir olgudur. Buna bağlı olarak da teknoloji tabanlı eğitim ortamları sürekli gelişmektedir. Bu teknoloji tabanlı öğrenme ve ortamları sayesinde akademik danışmanlık sistemlerinde teknoloji kullanımı daha ekonomik olabileceği gibi öğrencilere daha iyi hizmet verilebilir. Üniversite yönetimi teknolojik tabanlı akademik danışmanlık hizmetinde, öğrenci merkezli uygulamalar ile danışmanları ve öğrencileri daha iyi bilgilendirerek daha etkin karar vermelerini iyileştirebilir. Ancak, teknoloji danışman öğrenci etkileşimleri yerine geçmez. Bu nedenle, danışmanların yerini almak için değil, danışmanlık deneyimini ve hizmetini geliştirmekte bir araç olarak teknoloji kullanımı gerekir (Steele, vd., 2013).

Ders seçiminde öneri sistemleri uygulamasında temel hedef öğrencinin başarısını tahmin etmektir. Bu başarı tahmini öğrenme analitiğinin ilk hedeflerinden birisidir. Bu nedenle Öğrenci başarısının tahmin edilmesinde “ Kestirimsel Analitik” ve “Öğrenme analitiği” kavramlarının açıklanması gerekir.

2.2. Ders Öneri Sistemlerinde Kestirimsel Analitik

Analitik kavramı, Fransızca kökenli “Analytic” kelimesinden türemiş olup Türkçe ’de çözümlenmeli, çözümsel veya tahlil etme anlamlarında kullanılmaktadır. Analitik, veri odaklı kararlar almayı sağlayan kapsayıcı bir kavram olarak da nitelendirilebilir. Kestirimsel analitik ise; büyük hacimleri veriler içindeki ilişkileri, kalıpları ortaya çıkararak davranışları ve olayları tahmin etmek için çeşitli teknolojileri kullanan istatistiksel analiz alanı olarak tanımlanabilir (Barneveld, vd., 2012). Özellikle modelleme ile kestirimsel analitik tekniklerinde oldukça sık başvurulan öneri sistemlerinin (Recommender Systems) ve Veri Madenciliğinin (Data Mining) en önemli görevlerinden biri model oluşturarak bir olayın gerçekleşmesini kestirmektir

Öngörüsül veya tahmine dayalı analitik olarak Türkçeye çevrilen Predictive Analitics, İş dünyası için, gelecekteki beklentiler, riskler veya fırsatların oluşumunu belirlemek için kullanılmaktadır. Kestirimsel analitik; gelecekteki belirsiz olaylar hakkında çıkarım yapmak için kullanılan bir grup tekniktir. Eğitim Veri Madenciliği (EDM), Öneri sistemleri ve öğrenme analitiği (LA) gibi alanlarda, kestirimsel modelleme, araştırmacıların temel bir uygulaması haline gelmiştir ve öğrencinin akademik başarısını operasyonel olarak tahmin etmeye odaklanmıştır (Manouselis, vd., 2012).

Öncelikle, kestirimsel modellemeyi, açıklayıcı modellemeden ayırmak önemlidir. Açıklayıcı modelleme, amaç belirli bir sonuç için bir açıklama sağlamak üzere mevcut tüm etkileyen faktörlerin çoğunu ya da bazılarını kullanarak açıklamaktır. Örneğin, öğrencinin ders başarısı üzerimde, yaş, cinsiyet ve sosyoekonomik statüsüne ilişkin değişkenlerin nasıl katkıda bulduklarını açıklamak için bir regresyon analizi ile modellenenebilir. Böylece, bu yaklaşımları kullanılan modelde genelde nedensellik (karşı bağıntılı) olmakla birlikte sunulan sonuçlar sıklıkla deneysel çalışmalardan kaçınılmış ve nedenselliği ima etmek için teorik yoruma dayanmıştır (Brooks ve Thompson, 2017).

Buna karşın, kestirimsel modellemede amaç, gözlemlere dayalı yeni verilerin değerlerini (veya tahmini sayısal veriyle ilgilenmiyorsa sınıfı) öngören bir model oluşturmaktır. Kestirimsel modelleme, açıklayıcı modellemeden farklı olarak, bilinen veri kümesinin (eğitim verilerinin) varsayımına dayanır. Kestirimsel tahminlemede kullanılan veriler eğitim verisi ve test verisi olarak adlandırılır (Brooks ve Thompson, 2017; Johnson ve Wichern, 2007).

2.3. Ders Öneri Sitelerinde Öğrenme Analitiği

Öğrenme analitikleri literatürde yeni bir kavram olsa da eğitsel veri madenciliği (educational data mining), işletme analitikleri (business analytics), web analitikleri (web analytics) ve akademik analitikler (academic analytics) kavramları ile yakından ilgilidir. Öğrenme analitiği kavramı tıpkı İşletme analitiği veya Web analitiği kavramına benzer şekilde, eğitim sistemlerinde üretilen verinin toplanması, analiz edilmesi ve raporlanması gibi konular ve uygulamalar söz konusudur.

Öğrenme analitiklerine temel olan analitik kelimesi, veriler ışığında açık ve kesin sonuçlara ulaşabilmek için yollar sunan bilimsel bir süreci ifade etmektedir. 1980 ve 1990'larda popüler hale gelen verilere dayalı karar verme, genel olarak analitik olarak adlandırılan yazılım yaklaşımlarına dayanmıştır. Günümüzde ise analitik, sosyal ağ, metin, web izleri vb. büyük verilerden yararlanma olarak bilinen daha karmaşık bir kavram haline dönüşmektedir. Öğrenme analitiği (LA) alanı, veri analizindeki hesaplama veya algoritmalara bağlı yöntemlere doğal olarak ilgi göstermektedir. Bu perspektifte "öğrenme analitiği", teknoloji açısından zengin ortamlarda öğrenme etkileşimlerinin deneysel analizinden başka bir şey değildir. Aslında analizin bir parçası olarak belirli hesaplama ve matematiksel yaklaşımları da gerektirir (Young ve Muller, 2015). Öğrenme Analitiği kavramı haliyle henüz yeni tanışılmış olsa da bu konuda yapılmış çeşitli uygulamalar ve geliştirilmiş araçlar bulunmaktadır (Johnson, vd., 2014).

Öğretim uygulamaları için büyük veri ve analitik çalışmaları emekleme aşamasındadır ve var olmaları hâlihazırda hissedilmekte ve göz ardı edilmemesine rağmen olgunlaşması birkaç yıl alacaktır. Büyük veri ve analitik, yükseköğretimdeki yöneticilerin karşılaştığı tüm sorunları ve kararları ele almak için kapsamlı çözümler olmamasına rağmen,

yönetim ve öğretim işlevlerine bütünleşmiş çözümlerin bir parçası haline gelebilirler (Picciano, 2012).

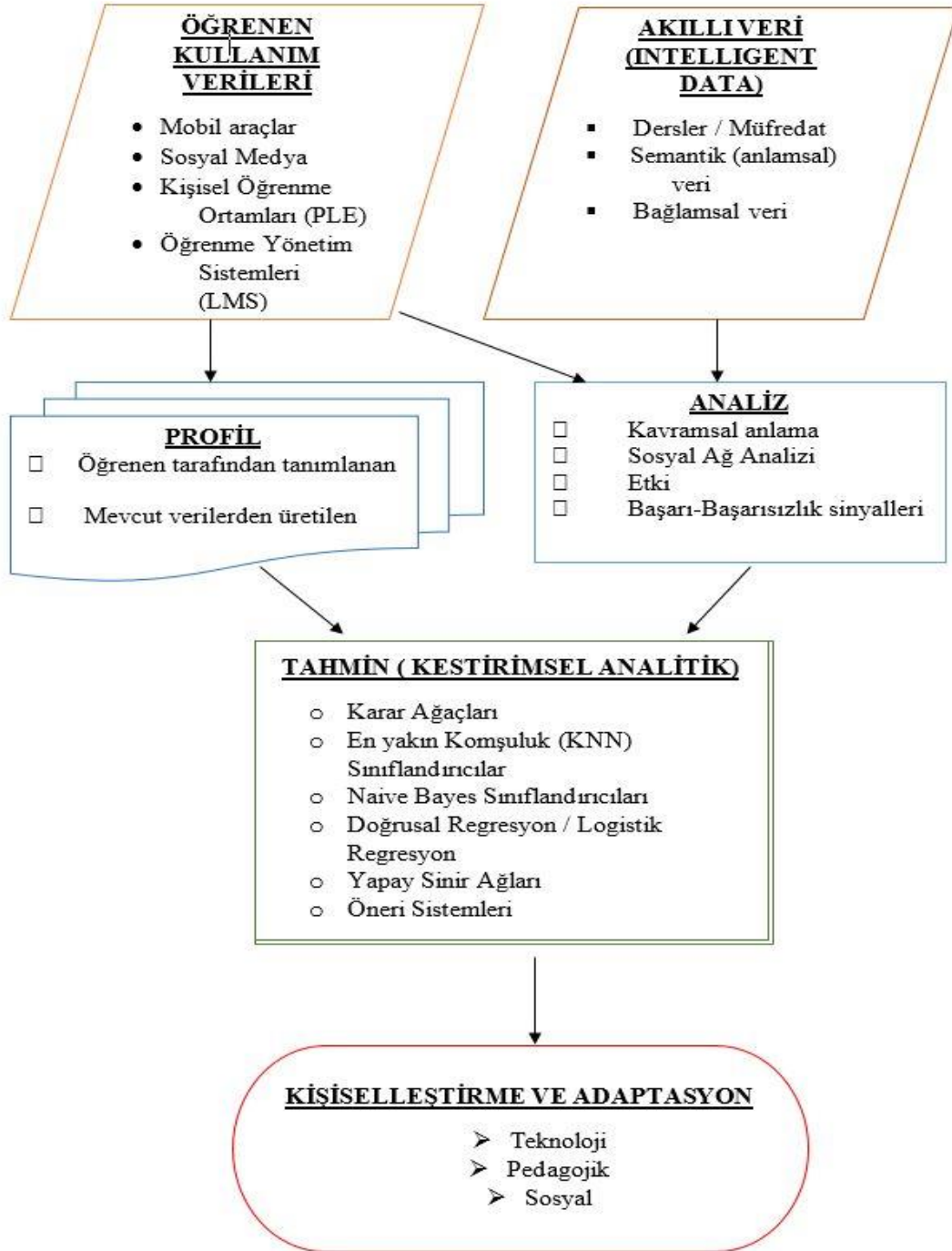
Daha genel bir ifadeyle analitik, veriye dayalı sistematik bir akıl yürütmedir (Davenport, vd., 2010). Bu süreçte ki temel amaç istatistiksel analiz yöntemlerine ek olarak yapay zekâ ve tahmin modelleme gibi farklı yaklaşımlarla verilerdeki anlamlı kalıpların keşfedilmesi, yorumlanması ve ilişkilendirilmesidir. Öğrenme Analitiği öğrenmenin ve öğrenmenin gerçekleştiği ortamın anlaşılması ve optimize edilmesi amacıyla; öğrencinin bulunduğu bağlam hakkındaki verilerin ölçülmesi, toplanması, analiz edilmesi ve raporlanması olarak tanımlanmaktadır (Chatti, vd., 2012). Ayrıca Öğrenme analitiğinin Siemens tarafından yapılan tanımına göre "akıllı verilerin ve öğrencinin ürettiği verilerin kullanımı, enformasyonu ve sosyal bağlantılarını keşfetme amacıyla kullanılan modellerin analizleri ve öğrenme konusunda kestirim ve tavsiyelerin üretilmesidir". Bu tanım öğrenme analitiği ile eğitsel veri madenciliğinin çok yakından ilişkili olduklarını göstermektedir (Manouselis, vd., 2012).

Öğrenme analitiği, Öğrenen tarafından üretilen ve öğrenenle ilgili birçok verinin analizine olanak sağlar. Bunları Öğrenen aktivite ve gelişimini izlemeyi mümkün kılar ve öğrenen çıktılarını tahmin etmeyi amaçlar. Öğrenmeye, öğreten ve öğrenenlerin müdahalelerine ve karar vermelerine olanak sağlar (Diaz ve Brown, 2012) Öğrenme analitiklerinin tanımına bakıldığında amaçların benzer olabileceği düşünülse de eğiticiler ve öğrenenler öğrenme analitiklerini farklı amaçlarla kullanabilmektedir (Sharples, vd., 2013). Sofistike analitik araçları öğrenme ve eğitimi iyileştirmek için kullanıldığı gelişmekte olan bir alan" olarak öğrenme analitiği olarak tanımlanır. Ayrıca öğrenme analitiği, bilgi ve sosyal bağlantıları keşfetmek ve tahmin etmede öğrenme öneri akıllı veri, öğrenen üretilen veri, analiz ve modellerin kullanımı" olarak tanımlamaktadır (Siemens, 2010). Yeni nesil öğrenmede "öğrenci ilerlemesini ve performansını tahmin etmek için veri ve modellerin kullanımı ve bu bilgilerle hareket yeteneği" öğrenme girişimi olarak tanımlanır. Burada model kullanımı öğrenen tarafından üretilen veriden, akademik ilerlemeyi değerlendirmek öğrencinin gelecek performansını tahmin ve potansiyel sorunları belirlemek adına toplanan büyük verilerin yorumlanması anlamında kullanılmaktadır. Bazı ayrıntılarda farklı olmakla birlikte, bu tanımlar eğitim verilerini öğrenmeyi teşvik etmede yararlı eylemlere dönüştürmek olarak da adlandırılabilir (Chatti, vd., 2012).

2.4. Ders Öneri Sistemlerinde Öğrenme Analitikleri Süreci

Öğrenme analitiği, öğrencinin ürettiği verilerden öğrenme sürecini ve sosyal bağlantılarını keşfetmek için yapay zekâ veya akıllı öğrenme analizleri yardımıyla tahmin ve tavsiye amacıyla modellenmesidir. Kısaca, analitik programları en iyi ders belirlemeye yardımcı olabilecek bilgileri karar vericilere sunmak için büyük veri kümelerini değerlendiren bir organizasyondur. Eğitsel veri madenciliği son yıllarda ilgi çekiyor (Manouselis, vd., 2012). Öğrenme analitiği çeşitli bilgisayar sistemleri ile öğrenci etkileşimi arasındaki faaliyetlere ilişkin bilgileri toplar ve öğrenme çıktıları arasındaki korelasyon aramak için "dijital kırıntıları" analiz eder. Öğrenme analitiği yazılımları her bir öğrenci için bir model oluşturmada, dersi alan öğrenci ile daha önce sınıfta dersi alan diğer öğrencilerin aktivitelerini karşılaştırır. Bu şekilde, öğrenme analitiği çoğu yükseköğretim ve üniversite öğretimi geliştirmek için kullanılan desenleri bulmak için büyük miktarlarda toplanan verilerden istifade eder. Öğrencilerin öğrenme deneyimlerinin kalitesi tüm eğitim kurumları için kritik bir konudur. Modern teknolojik gelişmelerin yardımıyla, eğitim kurumları öğrenme programlarının güçlü ve zayıf yanlarını daha iyi anlama yeteneğine sahiptir (Singh, vd., 2008).

Öğrenme analitik en yaygın kullanımı akademik olarak başarılı ve etkinleştirmek-hatta daha iyi sonuçlar elde yardımcı olmak için öğrencileri tespit etmektir. Aynı zamanda, bazı Fakültelerde belirli birimlerin genellikle öğrencilerin bir ders çalışma veya atamalarındaki zorluk nedenleri tanımlamak için Öğrenme analitiği araçları kullanır. Böylece tüm öğrencilerin bir kısmında öğrenme geliştirmek için ders değişiklikleri yapmak veya sonra müfredat ya da öğrenme faaliyetlerini değiştirebilir (Chatti, vd., 2012). Öğrenme Analitikleri Süreci Siemens, G'nin "What are Learning Analytics?" başlıklı makalesinden Şekil 2.1.'deki gibi uyarlanmıştır.



Şekil 2.1. Öğrenme analitikleri süreci (Siemens, 2010).

2.5. Ders Öneri Sistemleri Örnekleri

Günümüzde yeni teknolojilerin gelişimi sürekli olup, bu gelişen teknolojiler öğrenme süreçlerine dâhil edilmiştir. Öğrenme süreçleri içerisinde özellikle yükseköğretimde uzaktan eğitim ve örgün üniversite eğitiminde de kısmen dâhil edilmiştir. Böylece yükseköğretimde bilgi ve iletişim teknolojilerinin dâhil edilmesi, geleneksel üniversite eğitimine sayısız avantaj sağlamaktadır. Bu avantajlar arasında, esneklik vurgulanabilir, böylece öğrenciler, herhangi bir zamanda, bir içeriğe erişmek için bir bilgisayar veya mobil bir cihaza sahip olarak, eğitim içeriğine erişebilirler. Yani bilgiye erişim çok daha hızlı ve daha üretken ve gerektiğinde bilgi alınabilir bir yapıdadır.

Bilgi teknolojileri sayesinde öğreten ve öğrenen arasındaki iletişim ve etkileşimde değişmiştir. İletişim ve etkileşimdeki bu değişim sonucunda, bilgi alışverişi ve aralarındaki işbirliği daha akıcı olmaya başladığı gibi, öğrenen için öğrenimini kişiselleştirilmiş bir şekilde geliştirmesinin yolunu açmıştır. Böylece, eğitim kurumları hem öğrencilerin ilgisini ve performanslarını, içeriğe erişim ve öğreten ile iletişimdeki iyileşme artırmak için ve yeni metotları geliştirmiştir. Üniversitelerde eğitim ortamı bağlamında ortaya çıkan birçok yeni teknoloji vardır. Özellikle dijital platformda uygulanan Web tabanlı öğretim teknolojileri, yığın veri deposu haline gelmiştir. Bu yığın verinin analizi ve bu alandaki tekniklerin gelişmesine yeni uygulamalara neden olmuştur. Öğrenci her türlü bilgi ve belgeye hızlıca ulaştığı gibi, öğrenci bu süreçte daha yakın izlenebilir olmuştur. Öğrencinin izlenebilirliği, özellikle öğrenme sürecini faaliyetlerinin analizi anlamına gelmektedir. Öğrenme faaliyetlerine ilişkin izlerin takibi Öğrenme analitiği kavramını ve üniversitelerde bu alanda çalışmaları ve gelişmeleri beraberinde getirmiştir. Öğrenme analitiği kavramı tanımı, öğrenmeyi ve gerçekleştiği ortamları anlamak ve optimize etmek amacıyla, öğrenciler ve bağlamları hakkındaki verilerin ölçümü, toplanması, analizi ve raporlanmasıdır (Chih, vd., 2005).

Öğrenme analitiği çeşitli bilgisayar sistemleri ile öğrenci etkileşimi arasındaki faaliyetlere ilişkin bilgileri toplar ve öğrenme çıktıları arasındaki korelasyon aramak için "dijital izleri" analiz eder (Manouselis, vd., 2013). Bu dijital izler, yapılandırılmış veya yapılandırılmamış verilerden oluşmaktadır. Öğrenme analitiği yazılımları her bir öğrenci için bir öneri modeli oluşturmada, bütün öğrencilerden toplanan yapılandırılmış ve

yapılandırılmamış verilerden yararlanır. Böylece dersi alan öğrenci ile daha önce dersi alan diğer öğrencilerin aktivitelerini karşılaştırır. Bu şekilde, öğrenme analitiği çoğu yükseköğretim ve üniversite öğretim sürecini geliştirmek ve iyileştirmek amacıyla kullanılan desenleri bulmak için büyük miktarlarda toplanan verilerden istifade eder. Bu nedenle öğrenme analitiği bağlamında eğitsel veri madenciliği son yıllarda ilgi çekmektedir. (Manouselis, vd., 2013).

Öğrenme analitik en yaygın kullanımı öğrencilere akademik olarak başarılı olmalarında yardımcı olacak öğrenme etkinliklerini artırıcı, geliştirici ve yol gösterici öneriler sunmak amacıyla tespitlerde bulunmaktadır. Aynı zamanda, Yükseköğretim’de genellikle öğrencilerin ders çalışma ve başarılarının izlenmesiyle, ders seçimlerinde yaşadıkları zorluk nedenlerini tanımlamak için Öğrenme analitiği araçları kullanılmaktadır.

Öğrenme Analitiği kavramı ile öneri sistemleri henüz yeni tanışmış olsa da bu konuda yapılmış çeşitli uygulamalar ve geliştirilmiş araçlar bulunmaktadır (Johnson, vd., 2014). Öğrenme analitiği bağlamında geliştirilen Öneri sistemi araçlarından bazılarının işlevleri ve uygulama biçimleri aşağıda kısaca açıklanmıştır.

Öğrenme analitiklerinin yükseköğretimdeki ilk kullanım örneklerinden biri olan ve ilk deneme projesi 2007 yılında gerçekleştirilen Course Signal’s, (ders sinyalleri) belirli bir ders kapsamında başarısız olma riski yüksek öğrencilerin önceden belirlenebilmesi amacıyla geliştirilmiş bir projedir (Harmelen ve Workman, 2012; Arnold ve Pistilli, 2012). Purdue Üniversitesi tarafından 2007 yılında hayata geçilen projede ders notları, çalışma zamanı ve geçmiş öğrenme performansı gibi öğrenci bilgi sisteminden elde edilen bilgiler öğrenme yönetim sistemi kullanım verileri ile eşleştirilerek öğrencilerin başarısızlık riski farklı seviyelere göre sınıflandırılmaktadır. Bu sayede risk seviyesi yüksek öğrencilerin önceden tespit edilerek gerekli akademik, teknik veya sosyal yardım mekanizmalarının zamanında devreye sokulması sağlanmaktadır (Johnson, vd., 2011). Course Signal’s ile bir yarıyıl boyunca okutulan dersler kapsamında A ve B notu alan öğrenenlerin sayısını %10 arttırdığı, D ve F alan öğrenenlerin sayısını ise %6,41 azalttığı sonucu görülmüştür (Arnold ve Pistilli, 2012). Course Signal, hangi öğrencilerin kötü performans gösterebileceğini tahmin etmek için bireysel öğrenci performansını analiz eder ve böylece bu riski azaltmak için olumlu müdahalelerde bulunmaya olanak sağlar (Harmelen ve Workman, 2012).

Öğrenme analitikleri, öğrencilerin akademik başarılarının iyileştirilmesi ve yönlendirilmesi amacıyla da kullanılabilir. Bu amaçla, Tennessee Eyalet Üniversitesi Austin Peay'de Tristan Denley tarafından tasarlanan Degree Compass bir ders öneri sistemi ilk örneklerden biridir. Degree compass öneri sistemi Arizona'daki Peay State Üniversitesi'ndeki öğrencilerin hem güçlü yanlarını hem de gelişmelerine ışık tutacak bir penceredir. Degree Compass öneri sistemi; Amazon, Netflix ve Pandora gibi şirketler tarafından kullanılan öneri sistemlerinden esinlenilerek tasarlanmıştır. Degree compass, hangi dersleri başarıyla tamamlayabileceklerini önerisi sunarak öğrencilerin ders seçiminde karar vermelerine yardımcı olmak amacıyla kullanılmaktadır (Johnson, vd., 2013).

Degree Compass öneri sisteminin amacı, sistemdeki öğrencileri yeteneklerine uygun biçimde derslerle eşleşmelerini sağlamak ve öğrencilere gelecek dönemlerdeki çalışmaları için programlar önermektir. Degree Compass, öğrencilerin geçmiş dönemlerdeki verileri ışığında kişiselleştirilmiş bir öneri sunmaktadır. Degree Compass sisteminde algoritması yararlanılan algoritma, öğrenci ders puanları ve kayıt verilerine dayanan tahmin analitikleri kullanılarak bir öğrencinin kayıtlı olduğu programı boyunca akademik gelişimine hangi derslerin ne düzeyde katkı sağlayabileceği hesaplanmaktadır.

Bu hesaplamalar sonucunda oluşturulan öneri listesi kayıtlı olunan programın müfredatına uygun bir çalışma planı ile öğrenciye 1'den 5'e kadar derecelendirilmiş ders önerisi sunulmaktadır. Degree Compass uygulandığı tüm kampüslerde bir öğrencinin A ve B notu alma olasılığının 0,62 ile 0,63 arasında olduğu hesaplanmıştır. Öneri sisteminde B notu alacağı tahmin edilen ve bu dersi alan öğrencilerin %90'ından fazlasının A ve B ü harf notunu aldığı gözlemlenmiştir (Denley, 2013).

Öğretimi özelleştirme, özellikle web tabanlı öğrenim için öğretici sistemler için çok önemli bir hedeftir. Genel olarak, çoğu özel sistem kişiselleştirilmiş hizmetler sağlamak için bilgi, ilgi alanları ve seçeneklerin tarihini dikkate alır. İşbirliğine dayalı filtreleme yaklaşımı olan PEL-IRT, sunulan içeriğin kullanıcı bilgisi düzeyine uyarlanması için ders materyallerinin zorluk parametrelerini analiz eder. PEL-IRT, kişiselleştirilmiş öğrenmeyi sağlamak için hem ders materyallerinin zorluklarını hem de öğrencilerin yeteneklerini analiz eden madde cevap teorisine dayanan kişiselleştirilmiş bir öğrenme sistemidir (Chih, vd., 2005).

Montreal Üniversitesi Bilgisayar Bilimleri bölümündeki lisansüstü öğrenciler için RARE (Recommender system based on Association RuLES) olarak tanımlanan bir ders öneri sistemi geliştirilmiştir. RARE ders öneri sistemi, veri madenciliği teknikleri ile öneri sistemleri tekniklerinin birleştirilmesi yoluyla çalışmaktadır. Geliştirilen RARE ders öneri sistemi, Bilgisayar bölümündeki öğrenciler için ilgili oldukları düşünülen ders önerilerini sunar ve öğrencilerin değerlendirmelerini istemektedir. Böylece öğrencilerden gelen geribildirimler ile öğrencilerin davranışlarını örtük olarak analiz etmektedir. (Bendakir, 2006). Akademik kullanıcılar yeterlilikleri, bağlılığı ve değerlendirme cevaplarını detaylı olarak inceleyebilmektedirler. Çeşitli model yönetimi, tahmin ve veri görselleştirme yetenekleri sistem bireyselleştirilerek kullanılabilirler (Harmelen ve Workman, 2012).

CourseRank, Stanford Üniversitesi tarafından uygulanan bir ders öneri sistemidir. CourseRank, öğrenciler tarafından önceden alınmış popüler kursları ve kursları önerir. Özellikle, bu sistem, ders önerilerinde öğrencinin mezun olması için, alacağı derslerin gerekli koşulları sağlaması için, bir yönlendirme sistemi geliştirmek ve belirli kısıtlamaları dikkate almaya odaklanmaktadır. Bu kısıtlamaları kontrol etmek için maksimum akış algoritması kullanılır. Bu sistemde öğrencilere sadece kısıtlamaların karşılanmasına yardımcı olmakla kalmayıp aynı zamanda da (örneğin, popüler ya da benzer öğrenciler tarafından alınmak) arzu edilen dersleri önermektedir. Öğrencilerin Stanford Üniversitesi'nde akademik programları planlama süreci CourseRank sistemi kapsamında yapılmaktadır. (Parameswaran, vd., 2011).

Önerici sistem, karmaşık bilgi ortamları altında kullanıcılar için bir karar verme stratejisi olarak tanımlanmaktadır. Müşterilerin ihtiyaç ve tercihlerine göre uyarlanmış ürünler için benzersiz bir satış noktası sağladığından, kişiselleştirilmiş öneri sistemleri sıklıkla e-ticarette kullanılmaktadır (Schafer, vd., 1999). Bununla birlikte, son yıllarda kişiselleştirilmiş öneri sistemlerinin ticari olmayan amaçlar için potansiyeli araştırılmaya başlanmıştır. Bunlar arasından eğitim bağlamında; Eski kütüphane sistemlerinde dağıtık öneri hizmetleri için bir bütünleşme stratejisi (Geyer, vd. 2001). Geniş bir ürün yelpazesi öneren çeşitli eğitim öneri sistemleri tasarlanmıştır (Lawlws, 2003)., Web 2.0 kaynakları (Drachler, vd., 2010).,Yabancı dil dersleri öneri sistemi (Hsu, 2008)., Bilinçli öğrenme önerisi için nesne bağlamında işbirliğine dayalı filtreleme ve çıkarım kuralları. (Lemire, vd.,

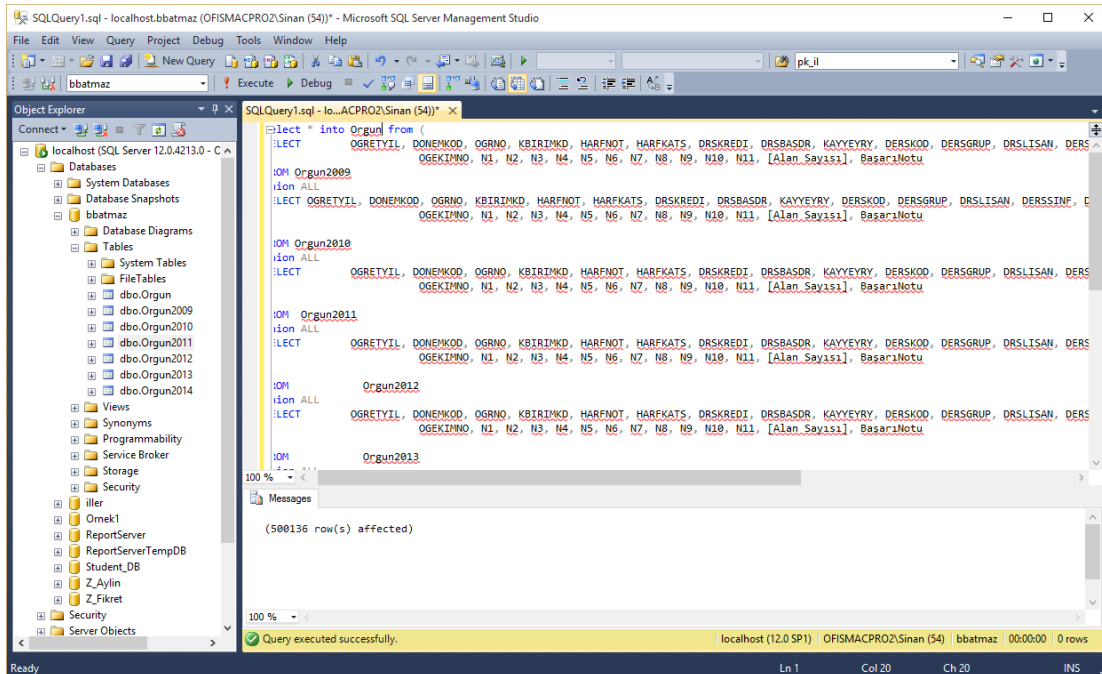
2005)., Sınav öğeleri ve ödevler (Rafaeli , vd., 2004)., E-öğrenmede sosyal navigasyon desteği: Gerçek ayak izleri nelerdir? (Farzan ve Brusilovsky, 2006) (Farzan ve Brusilovsky,2005)., Bir ders tavsiyesi sisteminde sosyal navigasyon desteği (Farzan ve Brusilovsky,2006) ., Uygulamalar kütüphane sistemlerinin kullanımı gibi çok farklı öğrenme ve eğitim alanlarını kapsamaktadır(Geyer , vd 2003)., Informal öğrenme (Drachsler, vd., 2009),, Mobil öğrenme (Andronic, vd. 2003)., ve benzer birçok araştırma yapılmıştır. Kişiselleştirilmiş öneri sistemleri ile ilgili araştırmaların bazıları uygulamaları sistemlerine eklenmiş ve başarılı sonuçları rapor edilmiştir.

3. MATERYAL VE YÖNTEM

3.1. Materyal

Bu tez çalışmasında, Yükseköğretimde öneri sistemlerine dayalı ders seçme öneri sistemi modellenmiştir. Sözkonusu modelin geliştirilmesi sürecinde Anadolu Üniversitesi 2009-2014 yılları arasında 6 eğitim yılı ve 12 yarıdönem öğrenci ders notlarından yararlanılmıştır. Anadolu Üniversitesi İktisadi İdari Bilimler Fakültesi öğrenci verileri yardımıyla hem zorunlu derslerde hem de seçmeli dersler için alabilecekleri notları tahmin edilecektir. Öneri tahmin algoritmaları için, kullanılacak veri; Anadolu Üniversitesi Örgün Öğrenci Bilgi Sisteminden (BAUM) izni ile alınarak veri tabanına aktarılmıştır. Anadolu Üniversitesi İİBF öğrencilerine ait 2009-2014 öğretim yıllarındaki öğrenci not bilgileri yıllar bazında alınarak MS SQL Server veri tabanı yönetim sistemine aktarılmıştır.

Verinin hazırlanması aşamasının birinci adımında, farklı yapılarıdaki veriler düzenlenerek standart hale getirilerek tek bir tablo haline getirilmiştir. Bu işlem adımıdaki MS SQL Server Veri tabanı yönetim sistemi ekran görüntüsü Şekil 3.1’de verilmiştir.



Şekil 3.1. Veri hazırlık aşamasında verilerin birleştirilmesi ve düzenlemesi

Verinin hazırlanmasının ikinci adımında ise; 14.111 farklı öğrencinin not bilgilerini içeren yapı her satır da bir öğrenci ve her bir sütunda ders kodu olacak şekilde

satır ve sütunlara dönüştürülerek bir matris veri elde edilmiştir. Öğrencilerin birden fazla almış oldukları ders ile elde edilen dikey matriste 1031711 satır veri içermektedir. Bu işleme ait ayrıntılar aşağıdaki tablonun ekran görüntüsü Şekil 3.2’de verilmiştir

Kimlik	OGRENİD	OGROZLN	OGROZLHS	OGRETYL	DONEMKOD	KBRIMKID	DERSSINF	YARILKO	DERSKOD	DRSLISAN	DERSGRUP	DRSKREDI	KRBASHOT	HARFNOT	DRSDEVMK	DRSSTKD	DRSORTGR	ds	
1031684	270189 206***5640	1	0	2002	100 5208070000	2,3	YPL103	1 A	3,0	35 20 FF	1	8	1 M						
1031685	270189 206***5640	1	0	2002	300 5208070000	2,3	YPL103	1	3,0	58 00 CB	1	8	1 M						
1031686	270190 206***5640	1	0	2001	100 5208070000	1	YPL161	1 A	3,0	51 80 DC	3	8	1 M						
1031687	270191 206***5640	1	0	2001	300 5208070000	1	YPL161	1 A	3,0	42 40 DC	3	8	1 M						
1031688	270192 206***5640	1	0	2004	300 5208070000	1,1	YPL161	1	3,0	72 00 BB	1	8	1 M						
1031689	270193 206***5640	1	0	2002	200 5208070000	1,2	YPL162	1 B	2,0	46 00 CD	1	8	1 M						
1031690	270194 206***5640	1	0	2003	200 5208070000	1,2	YPL162	1 B	2,0	69 20 BB	1	8	1 M						
1031691	270195 206***5640	1	0	2001	200 5208070000	1,0	YPL162	1 A	2,0	27 50 FF	1	8	1 M						
1031692	270196 206***5640	1	0	2005	200 5208070000	2,4	YPL201	1 B	4,0	54 80 CC	1	8	1 M						
1031693	270197 206***5640	1	0	2002	100 5208070000	2,3	YPL202	1 A	3,0	16 80 FF	1	8	1 M						
1031694	270198 206***5640	1	0	2002	300 5208070000	2,3	YPL202	1	3,0	14 50 FF	1	8	1 M						
1031695	270199 206***5640	1	0	2003	100 5208070000	2,3	YPL202	1 B	3,0	24 80 FF	1	8	1 M						
1031696	270200 206***5640	1	0	2003	150 5208070000	2,3	YPL202	1	3,0	00 FF	1	8	1 M						
1031697	270201 206***5640	1	0	2004	100 5208070000	3,5	YPL202	1 A	3,0	57 00 CB	1	8	1 M						
1031698	270202 206***5640	1	0	2002	100 5208070000	2,3	YPL203	1	3,0	66 00 CC	1	8	1 M						
1031699	270203 206***5640	1	0	2003	100 5208070000	3,5	YPL301	1 B	3,0	29 50 FF	1	8	1 M						
1031700	270204 206***5640	1	0	2003	150 5208070000	3,5	YPL301	1	3,0	00 DC	1	8	1 M						
1031701	270205 206***5640	1	0	2004	100 5208070000	3,5	YPL301	1 B	3,0	49 80 CD	1	8	1 M						
1031702	270206 206***5640	1	0	2004	300 5208070000	3,5	YPL301	1	3,0	57 20 CB	1	8	1 M						
1031703	270207 206***5640	1	0	2005	200 5208070000	3,6	YPL302	1 A	4,5	66 00 AB	1	8	1 M						
1031704	270208 206***5640	1	0	2005	100 5208070000	3,5	YPL305	1	3,0	51 00 CC	1	2	1 M						
1031705	270209 206***5640	1	0	2005	100 5208070000	3,5	YPL361	1	3,0	58 40 CB	1	2	1 M						
1031706	270210 206***5640	1	0	2005	200 5208070000	4,8	YPL402	1 A	4,5	39 50 FF	1	8	1 M						
1031707	270211 206***5640	1	0	2005	300 5208070000	4,8	YPL402	1	4,5	60 00 CC	1	8	1 M						
1031708	270212 206***5640	1	0	2005	100 5208070000	4,7	YPL404	1 A	4,5	66 00 CC	1	8	1 M						
1031709	270213 206***5640	1	0	2005	200 5208070000	4,8	YPL405	1 A	4,5	27 50 FF	1	8	1 M						
1031710	270214 206***5640	1	0	2005	300 5208070000	4,8	YPL405	1	4,5	63 50 BC	1	8	1 M						
1031711	270215 206***5640	1	0	2003	300 5208070000	4,8	YPL406	1	3,0	82 00 BB	1	8	1 M						
1031712																			
1031713																			
1031714																			

Şekil 3.2. Hazırlık aşamasında verilerin matrise dönüşümü

Veri tabanında dikey olarak depolanan verilerin yatay hale gelebilmesi için veri tabanı yönetim sisteminde pivot komutları kullanılmıştır. Her bir öğrencinin tek bir satır da yer alması sağlanarak her bir derslerin bir sütun olacak şekilde veri yapılandırılmıştır. Veri setinden 335 ders kodundan oluşan sütun ile 14.111 öğrenci TC kimlik numarasından satırdan oluşan 14.111x335 boyutunda bir matris elde edilmiştir. Öğrencileri aynı dersi birden fazla alma durumlarında ise aldıkları en yüksek harf notuna denk gelen katsayısı seçilerek tek bilgi haline dönüştürülmüştür. İktisat Bölümü ve İşletme bölümü öğrencileri ayrı ayrı algortmalar uygulanacağı için veri seti Bölüm koduna göre iki ayrı veri setine dönüştürülmüştür.

İktisat bölümü için 2196 öğrenci ve 108 ders kodundan oluşan öğrenci-ders notu bir veri matrisi oluşturulmuştur. Ayrıca İşletme bölümü için 2239 öğrenci ve 111 ders kodundan oluşan öğrenci-ders notu bir veri matrisi oluşturulmuştur. İlk düzenlenen 14.111 öğrenci – ders matrisinden, Fakülte ve bölüm ders kataloğundaki mevcut açık olan dersler seçilmiştir. Bu derslerin bazıları öğrenci işleri merkezi sisteminde kayılıdır. Ancak aktif dersler değildir. Bu nedenle iktisat 108 ders ve işletme 111 ders olmak üzere 219 derse göre veri matrisi temizlenmiştir. Veri seti İİBF öğrencilere ait olduğunu belirtmiştik, ancak

fakültenin bazı bölümlerinde çok az öğrenci vardır. Bu nedenle model geliştirmek ve test etmek amacıyla öğrenci sayısı en çok olan İktisat bölümü ile İşletme bölümü öğrenci - ders veri matrisi ile öneri sistemi algoritmaları uygulanmıştır.

3.2. Yöntem

İnternet dünya üzerinde yaygınlaştıkça, insanlar web üzerinden daha fazla bilgi paylaşabilir hale geldi. Bunun yanı sıra, neredeyse her internet kullanıcısı paylaşılan verilere daha hızlı bir şekilde erişebilmek ve aşırı veri yoğunluğuna karşı direnebilmek amacıyla kullanıcılar arasında veri paylaşımının olduğu farklı veri ağlarını kullanmaya başladı

Günümüzde bilgisayar ve internet hemen hemen her alanı büyük ölçüde etkilemiş ve etkilemektedir. Peta byte düzeyindeki verilerin ucuz maliyetlerle saklanıp işlenmesinin yolunu açan “Bulut Bilişim” ve büyük verinin anlamlandırılmasına yönelik analiz çalışmaları konumuzla ilgili çok önemli gelişmeler arasındadır. İktisat bize tüketici tercihlerinin nasıl oluştuğunu belirli bir kuramsal çerçevede açıklamaktadır. Benzer şekilde pazarlama bilimi bu alandaki bilgiye katkıda bulunmuştur. Ancak günümüz toplumlarında veri ve bilgi bombardımanı ile akli karışmış durumdaki tüketici tercihlerini şekillendirmek için yardıma gerek duymaktadır.

Web’ de kullanıcı kendisi için gerekli olan bilgiyi arama sürecinde, çok farklı ve ilgisiz bilgi ile de karşı karşıya kalmaktadır. Öneri sistemleri kullanıcıların ihtiyaçlarına göre ürün veya servis öneren sistemlerdir. Şirketlerin öneri sistemleri kullanmadaki temel amacı satışları artırmaktır. Bu amaçla e-ticaret satışlarını, Web sitesini inceleyen, kişilere kişiselleştirilmiş öneriler sunarak potansiyel müşteriye dönüştüren, çapraz satış sağlayan ve müşteri bağlılığı yaratmada Öneri sistemlerinden faydalanırlar. Özellikle büyük verinin iş dünyasını dönüştürdüğü bu dönemlerde rekabetin veri üzerinden yapıldığı, kar ve satış rakamlarını artırmaya yönelik veya yeni ürün tasarımında, ürün önermede ve müşteri beğenisini analiz etmede, analitikler ve onlarla ilişki diğer çalışma alanları giderek önem kazanmaktadır (Gürsakal, 2017).

Bu tez çalışmasında, kredili sistem uygulanan üniversitelerde, öğrencilere ders seçiminde kullanılacak öneri sistemi oluşturulması amaçlanmıştır. Bu amaçla bu tez

çalışmasında öğrencilerin dönemlik ders seçimlerinde ya da sistem tarafından ders atamalarında bir öneri geliştirilmesi veya otomatik danışman modeli oluşturulmasında ders öneri sistemi önerilecektir. Bu amaçla, Yükseköğretimde kredili sistemde eğitim gören öğrenciler, ders seçiminde daha başarılı olacağı bir öğrenim dönemi geçirmeleri için öneri sistemi oluşturulması amacıyla başarı notları tahmin edilmesi amaçlanmaktadır.

Tezin ilerleyen bölümlerinde, Yükseköğretim öğrencileri için ders seçiminde yardımcı olacak başarı notlarının tahmin edilmesinde yöntem işbirliğine dayalı filtreleme tekniklerinden yararlanılmıştır. Öneri sistemlerinde İşbirliğine dayalı filtreleme teknikleri arasından ders-ders benzerliğine dayalı teknikleri ile öğrenci-öğrenci benzerliğine dayalı kümeleme tabanlı k-Ortalamalar (k-Means), En yakın Komşu (KNN) ve boyut indirgeme ve gizil yapıyı belirlemeye yarayan Tekil değer ayrışımı (SVD) algoritmaları ile ders başarı tahmin sonuçları karşılaştırılmıştır.

3.3. Öneri Sistemleri

Yıllar içinde, insanların daha doğru ve ilgili bilgiye ulaşması için “Bilgi Filtreleme Sistemleri” kavramı ortaya çıkmıştır. Bilgi Filtreleme Sistemlerinin amacı; kullanıcıya otomatik olarak işlenmiş, gereksiz ve istenmeyen verilerden arındırılmış saf bilgiyi sunmaktır. Verileri topladıktan sonra, bilgi filtrelemede kullanarak tahminlerde bulunmak için filtrelemenin iki temel yolu vardır.

Bunlar Pasif filtreleme ve aktif filtrelemedir. Pasif filtreleme, en temel yöntemdir. Pasif filtrelemede tahminlerde bulunmak için veri toplamalarını ortalamalarını (örneğin, bir ögenin ortalama derecesi gibi) kullanılabilir. Örneğin, en çok izlenen ve beğenilen film önerisi pasif filtrelemedir. Bilgi filtrelemede kullanılan ikinci yöntem ise Aktif filtrelemedir. Aktif filtreleme, tahmin yapmak için kullanıcı geçmişindeki kalıpları kullanmaktır. Bunun bir örneği, mevcut kullanıcıya benzer kullanıcıları bulmak ve bir derecelendirmeyi tahmin etmek için geçmiş beğenilerini ya da notlarını kullanabilir.

Bu iki yöntem arasındaki fark incedir, ancak temelde önerilerin kullanıcıya özel olup olmamasına bağlıdır. Pasif filtrelemede, her kullanıcıya belirli bir ürün için aynı öngörülerde bulunulur. Aktif filtrelemede, bir öneride bulunmak için sistem kullanıcının

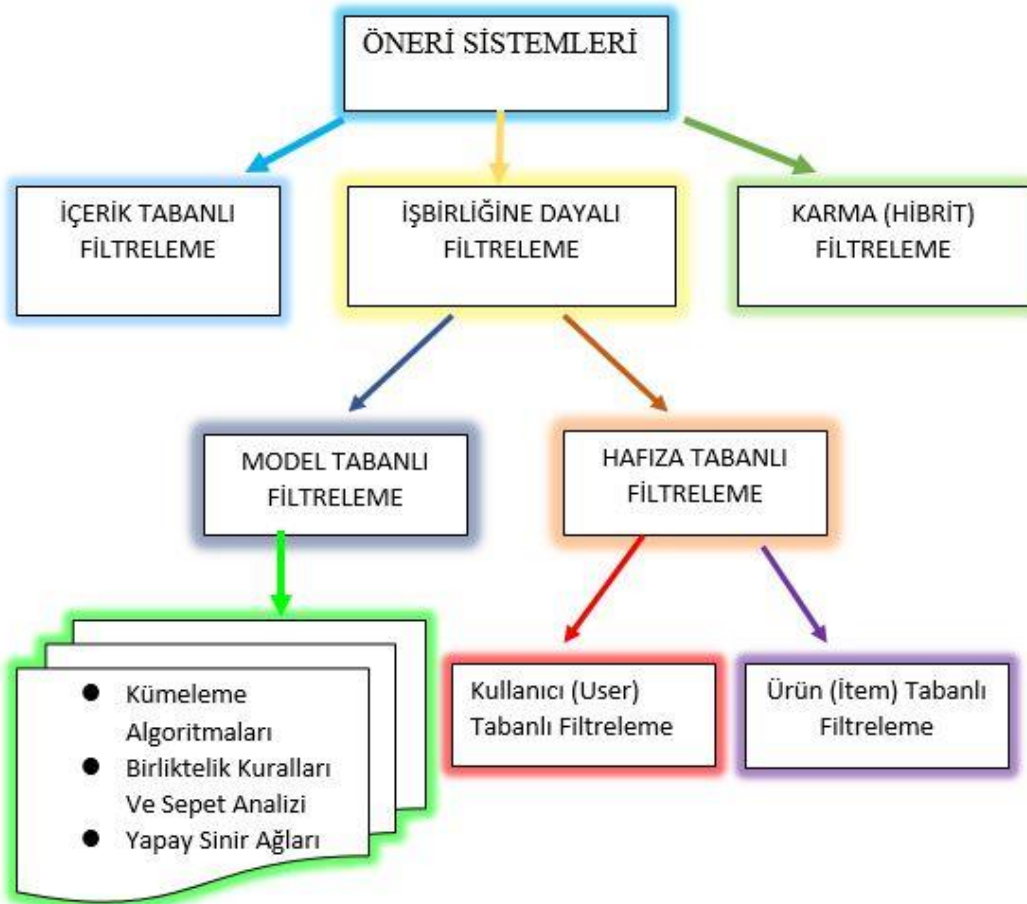
geçmiş verileri ile diğer kullanıcı benzerliklerini dikkate alır. Sonuçta, aktif filtreleme, çoğu kişinin işbirlikçi filtrelemeden bahsettiklerinde kastettiği şeydir. Pasif filtreleme çok kullanışlı ve pratik uygulamalara sahip olsa da, kişisel bir öneri sistemi sadece aktif filtreleme kullanılarak uygulanabilir. Kişiselleştirilmiş öneri sistemleri aktif filtrelemedir.

Bu işlemi yaparken de kullanıcı profilinin kıyaslanması ve benzer karakteristik özelliklere sahip olan kullanıcıları referans alarak benzer bilgileri sunmak veya önermek temel alınmıştır. İşte bu noktada “Öneri Sistemleri” de ortaya çıkmıştır. Bilgi Filtreleme, Web üzerindeki (over abundance) bolluk içindeki veriyi elemanın bir yoludur. Haber grupları katlanarak arttıkça, USENET veri tabanı yöneticileri, e-yığılmayı azaltma yolunda çaba harcamışlardır. Öneri yapan sistemler 1990’ların ortalarında İşbirliğine dayalı filtreleme (Collaborative filtering) makalelerini yayınladığında dikkatleri üzerinde toplamıştır. Öneri sistemlerinde bilgi filtreleme tekniği olarak “İşbirliğine dayalı filtreleme” aynı zamanda aktif filtreleme tekniği uygulamasıdır. 2006 yılının Ekim ayında başlayan Netflix Ödülü yarışması işbirliğine dayalı filtreleme alanında çok yeni bir ilerleme körüklemiştir. İlk kez, bilim adamları, öğrenciler, mühendisler ve meraklıları binlerce araştırma topluluğu 100 milyon sinema filmine ait derecelendirme-beğeni puanından oluşan büyük ölçekli veri setine eriştiler. Araştırmacılar bu veri seti üzerinde, rekabetin doğası gereği yeni nesil teknikleri ile yenilikçi tahmin ve tahmin doğruluğu geliştirmek için çaba sarf etmişlerdir. Böylece Netflix yarışması Öneri sistemlerinin hızlı gelişimini teşvik etmiştir (Greene.K, 2006).

Öneri sisteminde ilk formülasyonlar daha gelişmiş ve yaygın istatistiksel uygulamalar ve makine öğrenme (machine learning) literatüründen daha çok korelasyon istatistikleri ve tahmine dayalı modelleme üzerine kurulmuştur. Başlangıçta, öneri sistemi, akıllı ajan sistemleri bir tür kullanıcılarıyla ilgili öğeleri bilgi filtreleme sistemi yardımı ile öneren sistemlerdir. Buna ek olarak, önerilen öğeler filmler, Dersler ve haberler vb. gibi çeşitli seçenekler var (Adomavicius ve Tuzhilin, 2005). Bu durumda, madde ve ürün yaygın öneri sistemleri girdilerini tanımlamak için kullanılan soyut terimlerdir (Ricci, vd., 2010). Öneri sistemi daha resmi tanımı daha önce benzer ürünlere ait beğenisini belirten veya satın alan kullanıcılara benzer ürünleri alan kullanıcılara henüz tanışmadıkları ancak beğeneceği öngörülen ürünleri önermede kullanan bir kullanıcı modeli olarak tanımlanabilir (Mahmood ve Ricci, 2009). Ürünler hakkında bilgi tanımına göre, kullanıcı modeli belirli bir kullanıcı

için tercih değerlerini vermek için kullanılan bir terimdir. Öneri sistemleri, kullanıcının beğenilerini toplarken, hem bilgi topladığı kullanıcıya hem de diğer bir kullanıcıya öneri sunmada kullanır. Bu işlemde önerisi sisteminin kalitesini artırabilir (Ricci, vd., 2010).

Öneri sistemleri (Recommender System) son yıllarda internetin gelişmesiyle beraber çok yaygın olarak kullanılmaya başlanmıştır. İşbirliğine dayalı filtreleme (Collaborative Filtering) algoritması Öneri sistemleri için en yaygın olarak kullanılan algoritmadır. İşbirliğine dayalı filtreleme algoritması ilk olarak Tapestry sistem tarafından öneri üretmek amacıyla kullanılmıştır. CF (İşbirliğine dayalı filtreleme) algoritmaları e-ticaret, finans, sosyal ağ, eğitim gibi birçok alanlarda kullanılmaktadır (Goldberg vd., 1992) Ancak en yaygın olarak e-ticaret sitelerinde kullanılmaktadır. Amazon, ebay, Netflix gibi popüler e-ticaret siteleri kullanıcılara milyonlarca ürün arasından seçim yapmaya yardımcı olmak amacıyla öneri sistemlerini kullanmaktadırlar (Ricci, vd., 2010).



Şekil 3.3. Öneri sistemleri yöntemleri

Bu dönemde önerisi sistemleri hemen hemen her alanda kullanılabilir. Dersler, haber, müzik bu başlıkların en önde gelenlerindedir. Öneri sistemlerinde temel amaç, insanların beğenisini tahmin edip onlara bu beğenileri doğrultusunda önerilerde bulunmaktır. Öneri sistemlerinde öneri sunulacak alana, elde edilen veri yapısına ve amacına göre başvurulan yöntemler birkaç ana temel üzerine inşa edilmiştir. Bu yöntemlerin ve algoritmaların türleri ve birbiriyle olan bağlantılarını gösteren, öneri sistemleri yapısı literatürden derlenerek akış diyagramı halinde Şekil 3.3’de verilmiştir.

Öneri sistemlerinde istenilen amaca ulaşmak kullanılan yöntemler;

1. İşbirliğine dayalı filtreleme yöntemi (CF – Kullanıcı profillerindeki ve Ürün derecelendirme notu benzerliklere dayanır)
2. İçeriğe dayalı öneri üretmek İçerik Bazlı Yöntem (önerisi yapılacak öğenin ya da ürünün temel içerik bilgilerine dayanarak)
3. Hibrid Yöntemler (“İçerik Bazlı Filtreleme” ve “İşbirliğine Dayalı Filtreleme” yöntemlerinin birleşimine dayanarak)
4. Demografik Yöntemler
5. Faydaya dayalı öneri
6. Bilgiye dayalı öneri sistemleri bulunmaktadır.

Öneri sistemleri (Recommender Systems), İşbirliğine dayalı filtreleme temel varsayımına dayanır ve "...diğerlerinin fikirlerine dayanarak kişilerin tercih yapması" olarak tanımlanır. Zaman içinde "öneri sistemi" kavramı "işbirliğine dayalı filtrelemenin" önüne geçmiştir. Bu tür bir sistem, "insanların önerilerini girdi olarak sağladıkları ve sistemin bunları toplayıp uygun kullanıcılara yönlendirdiği sistem" olarak tanımlanmıştır. Sonuç olarak öneri sistemleri bugün geçerli olan anlamıyla, "...büyük bir olası seçimler uzayındaki ilginç veya faydalı nesnelere kişisel bir şekilde yönlendirme etkisi taşıyan, çıktı olarak kişiselleştirilmiş öneriler üreten herhangi bir sistem" olarak tanımlanmıştır (Bennet.J, 2007).

İşbirliğine dayalı (Collaborative) filtrelemede bir kullanıcının eylemleri ve belirli bir bilgi kırıntısına yönelik analizleri daha büyük bir kitlenin çıkarı için kayıt altına alınmaktadır. Bu topluluğun üyeleri yeni bilgiyi tüketip tüketmeme kararı vermeden önce başkalarının deneyimlerinden yararlanabilirler. Özünde, işbirliğine dayalı filtreleme sistemleri ‘ağızdan-ağıza’ tavsiyeleri otomatik hale getirmektedir. Otomatik işbirliğine

dayalı filtreleme işbirliğine dayalı filtreleme kavramını alır ve otomasyonu, ölçeği ve anonim ligi ilave eder. GroupLens Usenet filtreleme sistemi otomatik işbirliğine dayalı filtreleme yapabilen ilk sistemdir. O günden bu yana, Minnesota Üniversitesindeki GroupLens Araştırma Projesi işbirliğine dayalı filtreleme sistemleri üzerinde, işbirliğine dayalı filtreleme için matris depolama yöntemleri, tahminleme kalitesinin geliştirilmesi için oldukça çalışma yapmıştır. Bu çalışmalardan bazıları içerik-tabanlı filterbots bütünleştirilmesi, işbirliğine dayalı filtreleme ile bağımsız filtreleme ajanlarını birleştirme, tahminleme algoritmalarının işbirliğine dayalı filtreleme deneysel analizleri ile bağımsız filtreleme ajanlarının birleştirilmesi, SVD ile boyutun azaltılması ve işbirliğine dayalı filtreleme tavsiyelerinin açıklanması gibi alanları da kapsayan ileri araştırmalar gerçekleştirmiştir (Good, vd., 1999). GroupLens Usenet sistemiyle aynı zamanlarda, Ringo müzik tavsiye sistemi ve Bellcore Video Tavsiyecisi gibi pek çok benzer sistem geliştirilmiştir. Bu üç araştırma sisteminden ikisi ticari ürün olarak kullanılmaya başlanmıştır, komşu-tabanlı tahminleme algoritması olarak adlandırılan teknikleri kullanmıştır. Hızlarına, esnekliklerine ve anlaşılır olmasına bağlı olarak, komşu-tabanlı tahminleme algoritmaları hâlihazırda işbirliğine dayalı filtrelemede tahminlerin hesaplanmasında en etkin yollardan biridir (Goldberg, vd., 1992).

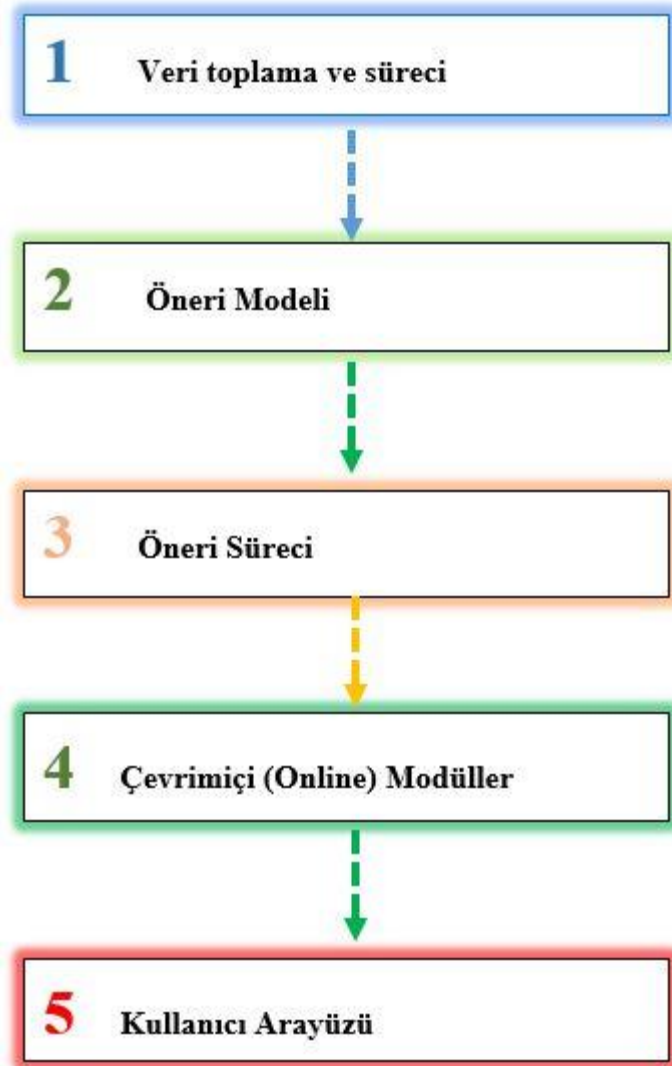
Öneri sistemleri, kullanıcı tarafından daha önce benzer ürünlere ait beğenisini belirten veya satın alan ancak henüz tanışmadıkları ve beğeneceği öngörülen ürünleri önermede kullanılan bir kullanıcı modeli olarak tanımlanabilir. Ayrıca Çizelge 3,1’de Öneri sistemleri ve İşbirliğine dayalı filtrelemenin literatüre dayalı olarak yıllara göre tanımları verilmiştir (Manouselis, vd., 2013).

Çizelge 3.1 Öneri sistemleri ve İşbirliğine dayalı filtrelemenin literatüre göre tanımları

İşbirliğine dayalı filtreleme en basit anlamıyla insanların, bir başkasının filtreleme sürecine, onun okuduklarına gösterdiği tepkileri kaydederek yardımcı olmak üzere işbirliği yapmasıdır (Goldberg, vd. 1992).
Toplumsal bilgi filtrelemesi özünde ‘ağızdan ağıza’ öneri sürecini otomatik hale getirir: mallar bir kullanıcıya kendisiyle benzer beğenilere sahip başka insanlarca belirlenmiş değerlere bağlı olarak öneri edilir (Shardanand ve Maes 1995).
Tipik bir öneri sisteminde insanlar girdi olarak önerilerini sunarlar ve sonrasında sistem bunları toplar ve uygun alıcılara yönlendirir (Resnick ve Varian 1997).
Öneri sistemleri, genellikle beğenecekleri ürünleri konumlandırma gibi bunaltıcı bir işte tüketicilere rehberlik etmek için kullanılır ki burada bilgi uzmanların sunduğu manuel olarak kodlanmış bilgi olabileceği gibi tüketicilerin davranışlarından veri madenciliği ile üretilen bilgi de olabilir (Schafer, 2001).
‘İşbirliğine dayalı filtreleme’ terimi bir grubun bilinen tercihlerinin yeni bir kullanıcının tercihlerinin tahmin edilmesi amacıyla kullanıldığı bir tekniği ifade eder; yeni kullanıcılara sunulan öneriler de bu tahmine dayanır. Önerilen diğer terimler ise ‘toplumsal bilgi filtrelemesi’ ve ‘öneri sistemleridir’ (Pennock ve Horvitz 1999; Goldberg, 2001).
Öneri sistemleri topluluk bireylerinin görüşlerini kullanarak o topluluktaki diğer bireylere ihtiyaçlarına veya ilgilerine en uygun bilgi veya ürünleri sunan sistemlerdir (Konstan, 2004).
Öneri sistemleri –belli bir kullanıcının belli bir maddeyi beğenip beğenmeyeceğini ön görmek için kullanılan kişiselleştirilmiş bilgi filtreleme (tahmin problem) veya belli bir kullanıcının ilgilenebileceği N sayıda madde kümesini tanımlamak (top-N öneri sorunu) için kullanılan sistemdir (Desphande ve Karypis, 2004).
Öneri sistemleri, o topluluktaki bireylerin potansiyel olarak yorucu seçeneğe sahip ili içeriğini daha etkin bir şekilde seçmelerine yardımcı olmak üzere o topluluktaki kullanıcıların görüşlerini sunar (Herlocker vd., 2004).
Kişiselleştirilmiş öneri sistemi tüketicinin geçmiş davranışları ve diğer kullanıcıların benzer tercihlerinden çıkarım yoluyla tüketiciye bire bir hizmet sunabilir. Kişiselleştirmenin amacı kişiye tam olarak ne istediğini sormaksızın öneri sunmak ve kişiler arası etkileşimin toplumsal bileşenini ele geçirmektir (Hung, 2005).
Öneri sistemleri kullanıcılara açık ve örtük tercihleri, başka kullanıcıların tercihleri ve kullanıcı ve madde özellikleri temelinde ilgi duyması muhtemel maddeyi önermektedir (Schein vd., 2005).
...çıkıtı olarak bireysel öneriler üreten veya kullanıcıyı olası seçenekler içinde faydalı nesneye kişisel bir şekilde yönlendirme etkisine sahip herhangi bir sistemdir (Burke 2002; Lops vd., 2011).
Öneri sistemleri öneride bulunan kişiselleştirilmiş bilgi kaynaklarıdır: öneriler muhtemelen bir kullanıcının yararına olacaktır... Öneri sistemi, bilgiyi ele geçirme sisteminden, kullanıcıların kendi aralarındaki etkileşimlerinin semantik (göstergebilimsel) yapısıyla farklılaşır (Burke, 2007).
... aktif kullanıcının tercihlerini makul şekilde tahmin edebilmek için başka kullanıcıların görüşleri seçilebilir ve birleştirilebilir (Ekstrand, 2010).

3.4. Öneri Sistemleri Bileşenleri

Bir öneri sistemi genel olarak 5 temel bileşenden oluşur. Ancak ticari işletmeler için öneri sistemlerinde bu bileşenler ekiplerine ve hedeflerine göre değişiklik gösterebilir (Isinkaye, vd., 2015). Öneri sisteminin temel 5 bileşeni izleyen kısımlarda açıklanmıştır. Şekil (3.4)'de öneri sisteminin ana temel 5 bileşeni şematik olarak verilmiştir.



Şekil 3.4. Ders öneri sistemi temel bileşenleri

3.4.1. Öneri sistemlerinde veri toplama süreci

Öneri sistemlerinde veri toplama süreci önemli bir süreçtir. Üretilen öneri kalitesi veri kalitesi ile sınırlıdır. Dolayısıyla toplanan veriler gereksiz bilgi barındırmamalı, hedefe yönelik gerekli ölçümler belirlendikten sonra veri toplanmalıdır. Bu nedenle ölçme ve ölçüm araçları önemlidir. Veri toplama işlemi sonrasında verinin bazı özelliklere göre ayıklanması “veri temizleme” ya da klasik veri madenciliği tanımına göre gürültü azaltma sürecidir. Ek olarak, farklı karakteristik yapılara sahip değişken sonuçlarını birbirlerine benzer hale getirebilmek için normalleştirme işlemi yapılarak, tüm değişken sonuçları [0,1] temsil edilirler. Sosyal ağlar, İnternet siteleri, mobil cihazlar gibi çok farklı kaynaklardan toplanan veri ve bilgilerin güvenli bir biçimde nerede ve nasıl depolanacağı sorunu yavaş kendisini göstermektedir. Toplanan bu veri ve bilgileri elinde bulunduranlar, artık veri veya bilgiyi piyasada alınıp-satılan bir ürün haline getirmişlerdir.

Öneri sistemleri, bilgi tanımına göre kullanıcının ürünler hakkında verdiği tercih ya da beğeni değerleridir. Öneri sistemlerinde kullanıcıdan toplanan bilgi, ürünler için beğenilerini yıldız verme veya 1-5 arasında ölçek puanlaması şeklindedir. Öneri sistemi toplanan bilgiyi hem bilgi topladığı kullanıcı için hem de diğer bir kullanıcı için kullanır. Bu işlem de öneri sisteminin kalitesini artırabilir. Öneri sistemleri daha çok film, Dersler, haber, makale, uygulama sistemleri ve Web siteleri için oluşturulabilir. Tipik bir öneri sistemi, kullanıcı bilgisinden yararlanarak, kullanıcının daha önceden incelemeyeceği ancak kullanıcının beğeneceği düşünülen ürünleri önerir. Öneri sistemleri, verileri açık ya da kapalı biçimde ele alır (Adomavicius ve Tuzhilin, 2005).

Sistem, normalde kullanıcıdan, sistem arabirimi aracılığıyla, modelini oluşturmak ve geliştirmek için öğeler için derecelendirmeler yapmasını ister. Önerinin doğruluğu, kullanıcı tarafından sağlanan oyların sayısına bağlıdır. Bu yöntemin tek eksikliği, kullanıcılar için çaba gerektirmesi ve kullanıcıların daima yeterli bilgi vermeye hazır olmamasıdır. Açık bildirimlerin kullanıcılardan daha fazla çaba harcaması gerekmesine rağmen, hâlâ daha güvenilir veriler sağladığı ve öneri sürecinde şeffaflık sağlayarak, biraz daha yüksek algılanan öneri kalitesiyle sonuçlandığı görülüyor. Açık veri için izleyen örnekler sıralanabilir (Vialardi, vd., 2009). Kullanıcının ürünler için belirli bir ölçek kullanılarak (ikili veya daha fazla seçenek) oy (beğeni derecesi) vermesi istenir.

- Kullanıcının ürünleri beğenisine sıralaması istenir. Örneğin, kullanıcı ürünleri en çok beğendiği öğeden, en az beğendiğine doğru sıralar.
- Kullanıcı iki farklı öğeden (ürünlerden) en sevdiğini seçer.
- Kullanıcı ürün seçeneklerinden, en beğendiği ürünlere ait bir liste yapar.

Sistem, satın alma geçmişi, web gezinme geçmişi ve bazı web sayfalarında harcanan süre, kullanıcı tarafından takip edilen bağlantılar, e-postanın içeriği ve diğerleri arasında düğme tıklaması gibi kullanıcıların farklı eylemlerini izleyerek kullanıcının tercihlerini otomatik olarak değerlendirir (Gadanh ve Lhuillier, 2007). Örtülü geribildirim, kullanıcının tercihlerini sistemle olan davranışlarından çıkararak kullanıcı yükünü azaltır. Bu yöntem kullanıcı çabası gerektirmiyor, ancak daha doğruluğu daha azdır. Aynı zamanda, örtük tercih verilerinin, kullanıcıların toplumsal açıdan istenen bir şekilde yanıt vermesinden kaynaklanan bir önyargı bulunmadığı ve kişisel çekinceler duymadığından dolayı, daha objektif olabileceği ileri sürülmüştür (Fisher, 1993). Kapalı veri için izleyen örnekler sıralanabilir,

- Kullanıcının web sitesindeki aktif süresi, ziyaret sıklığı, ürün(sayfa) tıklama sayıları incelenir.
- Çevrim içi mağazalarda kullanıcının incelediği ürünlerin belirlenir.
- Kullanıcının öğeyi görüntüleme süresi tespit edilir.
- Kullanıcının çevrim içi sayfalarda satın aldığı öğelerin kayıtları tutulabilir.
- Kullanıcının sosyal ağları incelenerek, ürünler hakkındaki beğenileri tespit edilir.
- Öneri sistemlerinin dikkate aldığı kullanıcı özellikleri yaş, eğitim, cinsiyet, meslek ve yaşadıkları şehir demografik bilgiler gibi statik olabilirken; kullanıcıların sistemle etkileşimleri, geçmiş alışverişleri, inceledikleri ürünler, gezindikleri Web siteleri, arama motorlarındaki aranılan anahtar kelimeler veya sosyal ağlardaki beğenileri gibi dinamik veriler de olabilir.

Öneri sistemlerinde, toplanan verinin türü öneri sisteminde uygulanacak algoritmaların ve tekniklerin seçilmesinde sınıflamasında önemlidir. Öneri sistemleri için

veri türü on madde halinde sınıflanabilir. (Goldberg vd., 1992; Adomavicius ve Tuzhilin, 2005).

- 1.) Öğeler × kullanıcılar: Skaler bir rakam, puan, Evet-hayır şeklinde ikili veriler veya Tag verileriyle günlükler (derecelendirmeler, satın almalar, göz atma ...). Metin verileri ile öğelerin kullanıcılara ait değerlendirme anketleri ya da forumlar.
- 2.) Kullanıcılar × kullanıcının sos yo-demografik verileri: İstemci dosyaları. Coğrafi pazarlama dosyaları.
- 3.) Kullanıcılar × öğe meta verileri: Kullanıcıların öğe özellikleri veya etiketleri üzerindeki tercihleri bildirildi.
- 4.) Kullanıcılar × kullanıcılar: Sosyal ağda arkadaş grupları. Burada sosyal ağlardaki node (düğümler).
- 5.) Öğeler × öğeler: HTML sayfaları arasındaki bağlantılar. "İlgili arama" cihazıyla bağlantılı öğeler.
- 6.) öğeler × meta veriler: Ürün kataloğu. Metin uygulamaları için belgeler x anahtar kelimeler. Etiketli öğeler.
- 7.) öğeler × kullanıcının sos yo-demografik verileri: Örneğin bir kampanya için belirli pazarlama segmentasyonu (bölünmesi).
- 8.) Meta veriler × kullanıcının sos yo-demografik verileri: Örneğin, bir kampanya için genel pazarlama segmentasyonu (bölünmesi).
- 9.) Öğe meta verileri × öğe meta verileri: Tür, türler, anahtar kelimeler arasında bağlantı veren eş anlamlı kelimeler.
- 10.) Kullanıcının sos yo-demografik verileri × kullanıcının sos yo-demografik verileri: Örneğin, bir toplantı sitesi için bir öncelik bölümlenme. Özellikle pazarlama segmentasyonu (bölünmesi), sos yo demografik sınıflama özellikleri.

3.4.2. Öneri Modeli

Öneri sistemlerinin çekirdek bileşenidir. Kullanıcılara, önerisi yapılacak ürün bilgisi, kullanıcı tercihleri, kullanıcılara önerisi sunulacak ürünler için kullanıcının geçmiş beğeni puanları gibi verileri almaktan sorumludur. Bu modelde öneri sistemi elde ettiği veriye bağlı olarak amaca yönelik kullanacağı algoritmaların ya da hangi yöntemlerden

yararlanıldığı veya öğeler (ürünler) için kestirimlerin yapıldığı ve önerilerin oluşturulduğu süreçtir (Resnick ve Varian, 1997).

3.4.3. Öneri Süreci

Öneri modeli tarafından oluşturulan öneriler, genellikle kullanıcılara gösterilmeden önce bazı ek düzenlemeler (post-processing) gerektirir. Öneri sürecinde, düzenleme önerinin doğruluğunun belirlenmesinden sorumludur. Kullanıcının ihtiyacı olan ürün alternatiflerini genişleterek çeşitlilik yaratan bir süreçtir. Bu bileşen sistemin gereksinimine bağlı olarak gerçek zamanlı çevrimiçi olabileceği gibi, çevrimdışı da olabilmektedir.

3.4.4. Çevrimiçi (Online) Modüller

Öneri modeli bileşeninde üretilen önerilerin, düzenlenmesinden sonra, öneri performansı, kullanım raporları ve günlüklerin saklanması gibi öneri modeli bileşenini oluşturur. Kısaca kullanıcı ile öneri sistemi arasındaki etkileşimin olduğu bileşendir.

3.4.5. Kullanıcı Ara yüzü

Öneriler oluşturulduktan sonra, bu önerileri kullanıcılara aktarmak gerekir. Bu aktarma için kullanılan araç ara yüz (arabirimi) bileşenidir. Bu bileşen öneri sistemlerinde kullanıcılara sunulan öneriler ile etkileşimi tanımlar. Öneri sisteminin kullanılabilirliği ile kullanıcı ara yüzünün kullanılabilirliği arasında büyük bir ilişki olduğu açıktır.

3.5. Öneri Sistemlerinde Kullanılan Yöntemler

Öneri sistemleri son yıllarda internetin gelişmesiyle beraber çok yaygın olarak kullanılmaya başlanmıştır. Öneri sisteminde ilk uygulamalar, gelişmiş istatistikî teknik kullanımı ve makine öğrenmeden daha çok korelasyon istatistikleri ve kestirime dayalı modelleme üzerine kurulmuştur. Öneri sistemi yöntemlerini anlatırken göz önünde bulundurulması gereken unsur, öneri sistemlerinin temelinde bilgi filtrelemesini bulunmasıdır. Genel olarak bilgi filtreleme yaklaşımları işbirliğine dayalı filtreleme, içerik filtreleme ve her iki tipin bir karması olan Melez (hibrid) filtreleme olmak üzere üç ana gruba ayrılırlar. Bunların yanı sıra Demografik, Faydaya dayalı ve Bilgiye yöntemler dayalı olarak

sıralanırlar. Bir işbirliğine dayalı öneri sistemi için iki ana yaklaşımlardan biri olabilir: kullanıcı tabanlı algoritmalar (ya da sosyal filtreleme) ve madde-tabanlı algoritmalar (Ricci, vd., 2010). İşbirliğine dayalı filtrelemede, içerik tabanlı sistemlerin aksine, tavsiye edilen kalemlerin içeriğinden bağımsızdır ve bu nedenle kendi aralarında farklılık gösteren öğeleri tavsiye edebilir (Su ve Khoshgoftaar, 2009).

3.5.1. İşbirliğine dayalı filtreleme yöntemi

İşbirliğine dayalı filtreleme, tıpkı bir arkadaş tavsiyesi niteliğindedir. Her gün insanlar birbirlerine film, Dersler, tatil, müzik öneriyorsa, işbirliğine dayalı filtrelemede de öneriler benzer biçimde “madem bu kitabı beğendiniz, o halde sana önereceğim şu kitabı da beğenirsiniz” yaklaşımını içerir. İnsanlar film izlemek için öneri arayışına girdiklerinde, kendilerine benzer film zevkleri olan, benzer Dersleri daha önce alan, bu derse ilişkin hoşlanan veya başaran arkadaşlarına sorarlar veya onların önerilerini benimserler.

İşbirliğine dayalı filtrelemenin amacı, aktif kullanıcının tercihlerini, bir grup kullanıcının tercihlerine dayanarak kestirmektir. Örneğin, aktif kullanıcının ve diğer kullanıcıların filmler için verdikleri beğen puanlarını veri olarak ele alırsak, sistem aktif kullanıcının görmediği filmleri nasıl derecelendireceğini kestirebilir. İşbirliğine dayalı filtrelemede temel yaklaşım, kestirimlerin veya önerilerin benzer kullanıcıların görüşlerine dayandırılmasıdır.

3.5.2. İçerik tabanlı filtreleme

İçerik tabanlı filtreleme (Content Based Filtering), öneri sistemlerinin diğer bir önemli kategorisidir. İşbirliğine dayalı filtrelemeden temel farkı kullanıcı benzerliğine değil, ürünlerin benzerlik özelliklerine odaklanılmasıdır (Adomavicius ve Tuzhilin, 2005). Burada amaç, kullanıcının tercih ettiği ürünlere benzer özelliklere sahip ürünlerin önerilmesidir. Bu öneri sisteminde, kişiler arasında benzerlik araştırmadığı için hiç kimsenin tercih etmediği ürünler de aktif kullanıcıya önerilebilir. Oysa işbirliğine dayalı filtrelemede bu mümkün değildir. İçerik Tabanlı Filtreleme, içeriğin belirli bir özellik seti altında gruplanması ve kullanıcı profilini temsil eden belirli özelliklerin listesi ile uyum sağlaması sonucunda bazı içeriklerin filtrelemesidir. İçerik tabanlı filtrelemede kullanıcıların daha önce beğeni puanı

verdikleri ürünleri ya da içerikleri kaydetmek suretiyle bir kullanıcı profili oluşturulur (Herlocker, vd., 2004). Daha sonra bu profile bakarak kullanıcı için en uygun, en benzer özellikteki içerik ya da ürün öneri olarak sunulur.

İçerik tabanlı filtrelemenin bazı dezavantajlı yönleri de bulunmaktadır. İçerik her zaman kolayca analiz edilebilecek bir biçimde olmayabilir. İçerik tabanlı filtrelemede, resim, müzik ve video türlerini içeren çoklu ortam içeriklerinin analizi ve sınıflandırılması ya tam anlamıyla mümkün olamamakta, ya da elle sınıflandırılması gerekmektedir. Bunun sonucunda ise, içeriklerin oldukça kısıtlı bir özellik seti üzerinden değerlendirilmesi gerekebilir.

İçerik tabanlı filtrelemede karşılaşılan bir diğer kısıtlama da kullanıcı profillerinin oluşmasından sonra, bu profile uymayan diğer içeriklerin, ne kadar ilginç ya da güzel olurlarsa olsunlar, öneri edilme imkânlarının kalmamasıdır. Örneğin bir Müzik öneri sisteminde, kullanıcı daha önce Türk Sanat Müziği dinlememiş ise Türk Sanat Müziğinin en popüler şarkıları dahi önerilmeyecektir. Bu durumda kullanıcının bu içerikten haberi olmayacaktır.

3.5.3. Hibrit öneri sistemleri

Bu yöntemdeki temel amaç, öneri sisteminin kalitesini arttırmak için işbirliğine dayalı filtreleme ve içerik tabanlı filtrelemeyi birleştirerek, Melez (hibrit) bir yaklaşıma başvurulmasıdır. Böylece her bir yöntemin ayrı ayrı dezavantajlarını bertaraf etmek veya avantajlarından yararlanmak mümkün olabilir (Su ve Khoshgoftaar, 2009).

İşbirliğine dayalı öneri sistemleri ile içerik tabanlı filtreleme sistemlerinin birlikte kullanılmasıdır. İçerik tabanlı filtreleme yöntemleri, kullanıcılar için öneri yapılmak istenen öğeler, ürünler (şarkı, film, Dersler vs.) ile ilgili bilgiler kullanılarak benzer ürünlerin önerilmesidir. İşbirliğine dayalı filtreleme yönteminin çalışma prensibi ise diğer kullanıcı ve/veya kullanıcı gruplarının öneri ve öngörülerine dayanır.

Bir kullanıcıya tavsiyede bulunabilmek için diğer kullanıcıların benzer içeriklere verdiği değerler kullanılır. Kullanıcılar arasındaki benzerliklere göre diğer kullanıcıya

benzer ürün önerilerinde bulunulur. Bu yöntem Hibrit (melez) filtreleme yöntemi olarak adlandırılır. Bu yöntemde İşbirliğine dayalı ve İçerik tabanlı filtreleme sonucunda elde edilen öneri sonuçları birleştirilir. Daha sonra ürün bilgileri ve öneriler tek bir gösterimde kullanıcıya sunulur.

3.5.4. Demografik yöntemler

Bu yöntemde; benzer davranış gösteren kişileri araştıran işbirliğine dayalı filtreleme ile benzerlik göstermektedir. Bu yöntemde temel olarak kullanıcıların demografik özelliklerinin benzerliği dikkate alınır (Su ve Khoshgoftaar, 2009). Demografik bilgi kullanıcının sisteme kayıt olurken verdikleri bilgilerdir. Dolayısıyla öneri sistemleri için ilk dikkate alınacak veri bu bilgilerdir.

Ayrıca günümüzde sosyal paylaşım sitelerinin de yaygınlaşması ile neredeyse kullanıcılar için gizlilik ortadan kalkmış, bunun sonucunda kullanıcıların birçok kişisel ve demografik bilgileri açık hale gelmiştir. Ancak demografik yöntemlerin en büyük sorunlarından birisi, kullanıcıların demografik bilgileri toplanırken karşılaşılan zorluklardır.

Bir diğer zorluk ise, demografik bilgilerin genellikle sabit kalması, insanların zevk ve tercihlerinin zaman içerisinde daha hızlı değişmesidir. Dolayısıyla bazı durumlarda demografik bilginin çok genel kalması ve de kişilerin gerçek yaşları ile hissettikleri yaşlarının farklı olması durumu ortaya çıkabilir. Demografik özelliklerin değişim hızının günlük tüketim hayatındaki değişim hızıyla paralellik göstermemesinden dolayı öneri sistemlerinde benzerlik kurulması zor olduğundan konuya olan araştırmacı ilgisi azalmıştır (Towle ve Clark, 2000). Fakat bazı durumlarda, kullanıcılara bir öneride bulunulabilmesi sadece demografik bilgiler sayesinde yapılabilir. Örneğin, bir kullanıcı sisteme yeni kayıt olduğunda, geçmiş davranış bilgisi bulunmadığından, demografik bir benzerlik kurmak faydalı olabilmektedir.

3.6. İşbirliğine Dayalı Filtreleme İşlem Adımları

İşbirliğine dayalı filtreleme, film ve müzik gibi meta verilerle kolayca ve yeterince tanımlanamayan içerik için içerikten bağımsız bir tahmin tekniğidir. İşbirliğine dayalı

filtreleme, kullanıcılar ve ürünler arasındaki bağlantıları kuran bir veri tabanını kullanır (Sarwar, vd., 2000). Kullanıcı için öneri üretilmesi istendiğinde sistem, benzer tercihlere sahip diğer kullanıcıları araştırır. Benzer kullanıcının bulunması sisteme ve veri miktarına bağlıdır. İşbirliğine dayalı filtrelemenin temel varsayımı; A ve B kullanıcıları n tane ürünü/hizmeti benzer şekilde sıralıyorlarsa veya satın almada, dinlemede, izlemede benzer davranışları gösteriyorlarsa, diğer ürünleri/hizmetleri de benzer şekilde sıralarlar olarak ifade edilebilir. İşbirliğine dayalı filtreleme kullanıcıların ürünler hakkındaki derecelendirmelerini kullanarak, oylanmamış ürünler hakkında öneriler üretir (Ricci, vd., 2010). Bu işlemi gerçekleştirmek için n sayıda kullanıcıdan m sayıda ürün (daha önce satın aldığı veya değerlendirme yapabilecek kadar bir fikre sahip olduğu ürünler) hakkında bilgi (örneğin, puan) toplar. Bir öneri sisteminde İşbirliğine dayalı filtreleme yöntemi, problem tanımlandıktan sonra 3 temel adımda kestirim yapabilmekte ve öneri oluşturmaktadır. Bu adımlar;

1. Problem tanımı,
2. Benzerlik hesaplama,
3. Komşuları belirleme,
4. Öneri hesaplama, beğeni puanı kestirimi biçiminde sıralanabilir.

İşbirliğine dayalı filtrelemede hedef kullanıcıya kişiselleştirilmiş öneri sunar. Bir ürün için beğeni öneri tahmini hesaplandıktan sonra en yüksek beğeni puanı olan bir Üst-N öneri listesi sunulur.

İşbirliğine dayalı filtreleme tekniği, kullanıcıların öğeler için tercih ettiği bir veri tabanı (kullanıcı ögesi matrisi) oluşturarak çalışır. Ardından, önerilerde bulunmak için kullanıcı profilleri arasındaki benzerlikleri hesaplayarak kullanıcılar ile alakalı ve tercihlerle eşleştirir. İşbirliğine dayalı filtreleme sisteminde, $m \times n$ boyutlu bir kullanıcı-ürün matrisi oluşturulur. Bu matris içinde, n kullanıcı ve m ürün sayısıdır. Matris çeşitli ürünler hakkında kullanıcıların tercihlerini içerir. Bu tercih matrisinde, q hedef ögesi için bir tahmin elde etmeyi hedefleyen a gibi bir aktif kullanıcı sisteme oy/puan bilgilerini tanımlar. Bu aktif kullanıcı için q ürünü hakkındaki tahmin (paq), birbirine en çok benzeyen kullanıcılar seçilerek elde edilir. İşbirliğine dayalı filtreleme sistemi, tüketici/kullanıcı tercihlerinin korelasyon içinde olduğu varsayımına bağlıdır. Burada temel fikir, aktif kullanıcının onun düşüncelerine benzer düşüncelere sahip kişilerin tercih ettiği ürünleri tercih edeceğidir.

İşbirliğine dayalı filtreleme için en yaygın yaklaşım, kullanıcı odaklı en yakın komşuluk tabanlı yaklaşımdır.

3.7. İşbirliğine Dayalı Filtrelemede Problem Tanımı

İşbirliğine dayalı filtreleme, n tane kullanıcının, m tane ürün için beğeni puanlarını kestirme problemidir. Bu problemin çözümü, İşbirliğine dayalı filtreleme ile kullanıcının elde mevcut bilgilerinden yararlanılarak, örneğin; daha önce izlediği filmlere vermiş olduğu beğeni puanları ile henüz izlemediği ancak sistemde var olan diğer filmler için beğeni puanını kestirerek çözüm elde edilir. Beğeni puanı herhangi bir kullanıcının, izlediği film, TV Programı, okunan Dersler, dinlenen müzik veya satın alınan ürün için, (1-5) aralığında sıralanmış (1= Hiç beğenmedim, 5= çok beğendim) olmak üzere rakam, yıldız verme şeklindedir. Veya beğendi (0) –beğenmedi (1) şeklindeki ya da satın almadı (0) -satın aldı (1) şeklindeki 0 ve ya 1 değeri verilen ikili puanlardan oluşurlar.

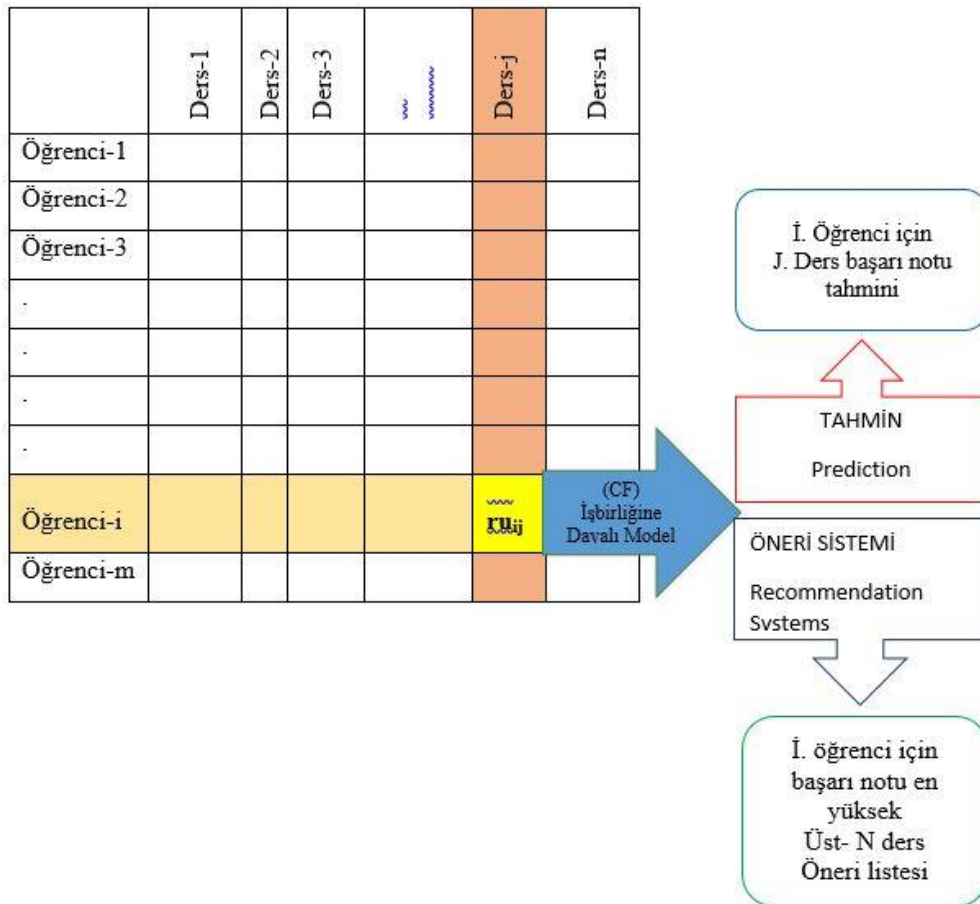
Öneri sitemlerinde kullanıcıya veya ürüne ait beğeni puanları bir vektör veya dizi ile ifade edilmektedir. İşbirliğine dayalı filtrelemede hedef kullanıcı için, sistemdeki m tane ürün içinde olan ancak kullanıcının henüz karşılaşmadığı ürünler için beğeni puanları kestirilir. Kestirim yapılabilmesi için hedef kullanıcının ve diğer kullanıcıların geçmiş bilgilerinden (beğeni puanlarından) yararlanır. Bu kullanıcılar komşu adı verilen bir grup oluştururlar. Bir kullanıcı, daha önce oy kullanmadığı, ancak komşuluğundaki kullanıcılar tarafından olumlu değerlendirilen öğeler için öneri üretilir. İşbirliğine dayalı filtreleme tarafından üretilen öneriler ya öngörü ya da tavsiye olabilir. Tahmin, kullanıcı i için j maddesinin öngörülen skorunu ifade eden sayısal bir değer R_{ij} 'dir, Kestirim sayısal bir değerdir ve örneğin beğeni derecelendirme 1-5 arasındaysa, belirli bir ürünün belirli bir kullanıcı için beğeni kestirimi aynı aralıktaki bir rakam ile belirlenir. Kestirilen bu beğeni puanı değerine göre öneri yapmak olanaklı olur. Öneri, aktif kullanıcının en fazla beğeneceği kestirilen n tane ürünün listesidir. Kestirim yapılan n tane madde için öneri listesi oluşturur. Öneri listesinde tüm ürünleri önermek kullanıcı için bir fayda sağlamayacağından, beğeni puanı (ya da öneri puanı) en büyük olan ürünlerden oluşan Üst- N öneri listesi oluşturulur. İşbirliğine dayalı filtrelemede problem (U, I, R) üçlüsü ile tanımlanır (Sarwar, vd., 2002).

Burada;

U: Kullanıcı; U_i , i'inci kullanıcı olmak üzere, $i=\{1,2,\dots,n\}$, n kullanıcı sayısını belirtir.

I: Ürün; I_j , j'inci ürün olmak üzere, $j=1=\{1,2,\dots,m\}$ ve m ürün sayısını belirtir.

R kestirilen derecelendirme değeridir; (0, 1) aralığında veya (1, 2, 3, 4, 5) aralığında kesikli ya da sürekli değerler alabilir. Bu tez çalışmasına 0-4 arasında sürekli bir değer alan başarı puanları aynı ölçme düzeyinde kestirim yapılacaktır. (U, I, R)'in bir işlevi (u, i, r) ile gösterilir. (u, i, r); u kullanıcısının i ürünü için r ile derecelendirmesidir. Burada amaç R(U, I) geçmiş derecelendirme değerlerinden R'yi kestirmektir. İşbirliğine dayalı filtreleme Kullanıcı-benzerliği ve Madde benzerliğine dayanan iki temel yaklaşımda da önce kullanıcı tabanlı yaklaşımda kullanıcılar arasındaki benzerlik, madde tabanlı da ise maddeler(öğeler-ürünler) arasındaki benzerliğin hesaplanması gerekir (Isinkaye, vd., 2015).



Şekil 3.5. Ders seçme öneri sistemi için problemin şematik gösterimi

Bu tezde problem olarak; Yükseköğretimde öneri sistemleri tekniklerinden İşbirliğine dayalı filtreleme tekniğinde yararlanarak, tahmin modeli (öneri motoru) oluşturmaktır. İşbirliğine dayalı filtreleme tekniğinin problem tanımı şematik olarak Şekil 3.5.'de gösterilmiştir.

3.8. Benzerlik Ve Uzaklık Hesaplamada Kullanılan Ölçütler

Öneri sistemlerinde, öncelikle problem tanımından sonra benzerliklerin hesaplanması gerekmektedir. Bu benzerlik düzeyine göre kullanıcı (user) veya ürün (item) kendilerine en benzer olan diğerleri ile gruplara etiketlenir. Böylece veri matrisinden yararlanarak kullanıcı veya ürün kümeleri oluşturulur. Küme, birbirlerine yakın bireylerin çok boyutlu uzayda oluşturdukları birlik olarak ifade edilebilir. Bu durumda küme kavramı, "benzerlik" ve "uzaklık" kavramlarını ifade etmektedir. Kümeleme analizi öncesinde benzerliklerin hesaplanmasında uzaklık ölçütlerinden yararlanır (Su ve Khoshgoftaar, 2009).

Verilerin geometrik olarak gösteriminde ikiden fazla boyut olduğunda noktaları ve birbirine olan uzaklıklarını çok boyutlu uzayda hesaplamak ve göstermek gerekmektedir. P tane değişkene ilişkin birimler ya da değişkenler arasındaki uzaklıkları hesaplamak için en sık kullanılan ölçütleri:

1. Minkowski Uzaklığı,
2. Öklid (Euclide) Uzaklığı
 - A. Ölçekli Öklid Uzaklığı
 - B. Binary Öklid Uzaklığı
3. Pearson Uzaklığı,
4. Manhattan (City-Blok) Uzaklığı,
5. Mahalanobis Uzaklığı,
6. Hotelling T2 Uzaklığı,
7. Canberra Uzaklığı,
8. Cosine, olarak sıralanabilir.

Temel olarak birim ya da deęişkenler arası benzerlik ya da uzaklıklar hesaplanırken geometrik yaklaşımlardan yararlanır. İşbirliğine dayalı filtrelemede en yaygın kullanılan Pearson korelasyon ve Kosinüs benzerlikleri açıklanacaktır. Geometride koordinat sisteminde yer alan iki nokta arasındaki en yakın uzaklık Pisagor bağıntısına göre bulunabilir. Dolayısıyla koordinat sisteminde yer alan A ve B noktaları arasındaki doğrusal uzaklık, A'nın koordinat $A(X_1, Y_1)$ değerleri ve B'nin koordinat değerleri $B(X_2, Y_2)$ olmak üzere denklem (3,1) deki Pisagor bağıntısına göre;

$$d(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3,1)$$

Olacaktır.

Verilerin geometrik olarak gösteriminde ikiden fazla boyut olması durumunda, noktaların birbirine olan uzaklıklarını çok boyutlu uzayda hesaplamak ve göstermek gerekmektedir. P tane deęişkene ilişkin birimler ya da deęişkenler arasındaki uzaklıkları hesaplamak için en sık kullanılan arasındaki bazı ölçütler Öklid (Euclidean) Uzaklığı, Pearson Uzaklığı ve Kosinüs uzaklığı olarak sıralanabilir (Chen, vd., 2009).

Temel olarak bir uzaklık fonksiyonunun özellikleri için izleyen durumlar söylenebilir.

$$d(i, j) \geq 0 ; \text{Uzaklık negatif olamaz.}$$

$d(i, j) = 0$; (i=j) olduğunda, her birimin kendisine olan uzaklığı sıfırdır.

$$d(i, j) = d(j, i) ; \text{Uzaklık fonksiyonu simetriktir.}$$

$d(i, j) \leq d(i, h) + d(h, j)$; İki birimin arasındaki uzaklık bu iki birimin üçüncü bir birime olan uzaklıkları toplamından küçük olamaz (üçgen eşitsizliği).

3.8.1. Öklid (euclidean) uzaklığı

En çok kullanılan uzaklık ölçü birimidir. Öklid uzaklık ölçüsü Öklid uzayında noktalar arasındaki mesafeyi düz çizgiler ile ölçen yöntemdir. Eğer iki nokta arasındaki Öklid uzaklığı bu noktaların birbirlerine olan uzaklıkları ve asgari mesafe üçgende hipotenüs

uzunluđuna denk gelmektedir. Öklid uzaklık ölçüsü kullanılarak iki nesne arasındaki uzaklık izleyen eşitlik yardımıyla hesaplanır. Uzayda iki nokta arasındaki Öklid uzaklığı hesaplamak için kullanılan, aralarındaki Öklid uzaklığı denklem (3,2) de verilmiştir (Field, 2009).

$$d(x, y) = \sqrt{(\sum (x_i - y_i)^2)} \quad (3,2)$$

Denklem (3.2) de görülebileceđi gibi, Öklid mesafe düzlemde iki nokta arasındaki mesafedir. Daha küçük mesafe, benzerlik derecesi: Öklid ile mesafe genellikle dönüşüm için denklem (3,3) ile verilen aşağıdaki formül kullanılarak benzerlik gösterilir (Field, 2009).

$$sim(x, y) = \frac{1}{1 + d(x, y)} \quad (3,3)$$

3.8.2. Pearson korelasyon katsayısı

Pearson korelasyon katsayısı genellikle yakından [-1, +1] arasındaki değeri derecesine bađlı iki verilen deđişkenler arasındaki mesafeyi denklem (3,4) de verilen eşitlik yardımıyla hesaplanır (Field, 2009).

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (3,4)$$

Pearson Uzaklığı: Pearson uzaklık ölçüsü kullanılarak iki birim arasındaki uzaklık denklem (3,5) de verilen eşitlik yardımıyla hesaplanır.

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 / S_1^2 + (x_{i2} - x_{j2})^2 / S_2^2 + \dots + (x_{ip} - x_{jp})^2 / S_p^2} \quad (3,5)$$

Eşitlikte kullanılan s_p , uzaklığın hesaplandığı deđişkene ait standart sapmadır. Bununla birlikte farklı gruplar hakkında önceden bilgi sahibi olunmadığı için, uzaklık hesaplanmasında s değerinin kullanılması doğru olmaz. Bu nedenle, Pearson uzaklık ölçüsü yerine genellikle Öklid uzaklık ölçüsü tercih edilir. Özellikle benzer kullanıcıları tespit etmek amacıyla kullanılacak kümeleme analizinde, ilgili deđişkenler belirli önem derecelerine göre ağırlık verilmiş ise, Pearson uzaklık ölçüsü denklem (3,6) da verilen eşitlik yardımıyla çözülür (Hartigan, 1975). W_i , ağırlıklandırma puanları olmak üzere;

$$d(i, j) = \sqrt{w_1(x_{i1} - x_{j1})^2 / S_1^2 + w_2(x_{i2} - x_{j2})^2 / S_2^2 + \dots + w_p(x_{ip} - x_{jp})^2 / S_p^2}$$

(3,6)

Pearson uzaklık ölçüsü, “Karesel Pearson uzaklık” ya da “standartlaştırılmış Öklid uzaklığı” olarak da adlandırılır.

3.8.3. Kosinüs (cosine) Ölçüsü

Kosinüs benzerliği yaygın belge verilerinin benzerliği hesaplamak için kullanılır: Özellikle belge ve çoklu ortam nesnelerinin kıyaslanması ile metin madenciliğinde kullanılmaktadır. \mathbf{X} ve \mathbf{Y} birer vektör olmak üzere, bu iki vektör arasındaki açının kosinüsü, açısız benzerliği, izleyen denklem (3,7) da verilen eşitlik yardımıyla hesaplanır (Tabachnick ve Fidell, 2001)

$$S_{xy} = \text{Cos}\theta = \frac{x \cdot y}{\|x\| \|y\|} = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (3,7)$$

Eşitlikte yer alan “x.y” ifadesi x ve y vektörlerinin nokta (skaler) çarpımını ifade etmektedir. Açısız benzerlik ölçütü [-1,1] aralığında değerler alır. Örneğin tamamen aynı yönü gösteren, tamamen birbiri ile aynı olan iki vektör için kosinüs değeri 1 olacaktır. Tamamen birbiri ile ilişkisiz olan vektörler için ise kosinüs değeri 0 olacaktır. Bu vektörlere dik vektörler denir. Açısız benzerlik iki vektör arasındaki açısız farklılığı temel aldığından dolayı vektörlerin matematiksel büyüklükleri üzerinden bir bilgi sağlamayacaktır (Su ve Khoshgoftaar, 2009).

Ayrıca Jaccard katsayısı denilen Tanimoto katsayısı, Kosinüs benzerliği de benzerlik hesaplama belge verileri için kullanılır. Jaccard katsayısı aşağıdaki denklem (3,8) da verilen eşitlik yardımıyla hesaplanır.

$$T(x, y) = \frac{x \bullet y}{\|x\|^2 + \|y\|^2 - x \bullet y} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} + \sqrt{\sum y_i^2} - \sum x_i y_i} \quad (3,8)$$

3.9. İşbirliğine Dayalı Filtreleme Algoritmaları

İşbirliğine dayalı filtreleme algoritmaları, e-ticaret, finans, sosyal ağ, eğitim gibi birçok alanda kullanılmaktadır. Özellikle e-ticaret sitelerinde yaygın kullanımı söz konusudur. Amazon, Ebay gibi popüler e-ticaret siteleri kullanıcılara milyonlarca ürün arasından seçim yapmaya yardımcı olmak amacıyla öneri sistemlerini kullanmaktadır. Öneri sistemlerinde, İşbirliğine dayalı filtreleme algoritmaları temel olarak 3 kategoride incelenebilir. Bunlar Bellek-tabanlı (memory-based), model-tabanlı (model-based) ve hibrit (hybrid) algoritmalarıdır (Su ve Khoshgoftaar, 2009). Algoritmalar arasındaki büyük ayrımlardan biri, bellek tabanlı algoritmalar ve model tabanlı algoritmalarıdır. Temel fark, bellek tabanlı algoritmaların tahmin yapmak için her zaman tüm verileri kullanmasıdır. Bu, bellek tabanlı algoritmaların genellikle bellekteki tüm verilere sahip olması gerektiği anlamına gelmektedir.

Çizelge 3.2 İşbirliğine dayalı Filtreleme Algoritmalarının Olumlu Ve Olumsuz Yönleri

CF Kategoriler	Örnek teknikleri	Avantajları	Eksiklikleri
Bellek tabanlı CF	Komşu tabanlı CF (Pearson / vektör kosinüs korelasyon ile madde-tabanlı / kullanıcı tabanlı CF algoritmaları) Madde-tabanlı / kullanıcı tabanlı üst N önerileri	Kolay uygulama, Yeni verileri kolayca ve aşamalı eklenebilir, Öğelerin içeriğini dikkate gerekmez.	İnsan yorumuna bağlıdır. Seyrek veri yapısında performans azalması görülebilir. Yeni kullanıcılar ve öğeler için tavsiye edemez. Büyük veri setleri için sınırlı ölçeklenebilirlik
Model tabanlı CF	Bayesian CF ağlar, Kümelenme CF, Gizli anlamsal CF, Faktör analizi CF, SVD ve PCA kullanır.	Kıtlık (sparsity), ölçeklenebilirlik ve diğer sorunların çözümünde. Öngörü performansını artırmak, Önerileri için sezgisel bir gerekçe vermek	Pahalı modeli oluşturma, Tahmin performans ve ölçeklenebilirlik arasındaki değiş tokuş vardır. Boyut indirgeme teknikleri için yararlı bilgiler kaybedersiniz
Hibrid Öneri Sistemleri	İçerik tabanlı ve İşbirliğine dayalı filtreleme tekniklerini birleştirerek, her ikisinin ayrı ayrı olumsuz yönlerini gidermeye çalışır.	Öngörü performansını artırır, Sparsity (kıtlık) ve gri koyun gibi problemlerin üstesinden gelir.	Uygulanması karmaşık ve zamanı alıcıdır.

Model tabanlı algoritmalar, daha sonra tahmin yapmak için kullanılacak bir modeli öğrenmek / eğitmek için verileri kullanır. Model tabanlı, orijinalden daha az veri kullanarak hızlı öngörülerde bulunabilir. Bu algoritmaların olumlu ve olumsuz yönleri Çizelge 3,2’de karşılaştırmalı olarak verilmiştir.

3.9.1. Bellek-Tabanlı Algoritma

Bellek-tabanlı filtreleme algoritmaları tahmin üretmek için sistemdeki bütün verileri kullanır. Bellek-tabanlı işbirliğine dayalı filtreleme algoritmaları çevrimiçi çalışmaktadır. Bu algortmada önerisi sistemi tahmin, (kullanıcı –ürün) matrisi üzerinde hesaplanan benzerlikler yardımıyla hesaplanır.

Madde-tabanlı algoritmaları, öneri sisteminin farklı öğeleri arasındaki benzerlik derecesine göre yapılır. Bir kullanıcı önceden görmüş olduğu öğelere benzer öğeleri tercih fikrine dayanmaktadır (Almazro, vd., 2010). Ana örneklerinden biri Amazon veri tabanındaki diğer kullanıcıların yüksek memnuniyet derecelendirme sahip kullanıcılar tarafından satın alınan ürünlerle ilgili diğer kullanıcılara memnuniyeti yüksek ürünlere dayalı olarak öneriler yapar. "Ürün-Ürün İşbirliğine dayalı Filtreleme" olarak adlandırılan Amazon algoritması, milyonlarca ürün içinde kullanıcılara öneriler sunmada, başa çıkmak için etkili ve uygun olduğu görülmüştür (Hernández-del-Olmo ve Gaudio, 2008).

Madde Tabanlı İşbirliğine dayalı Filtreleme Yöntemi; İşbirliğine dayalı filtrelemede madde (öge-ürün) benzerliğine dayanan bu yöntemde, maddeler (öğeler-ürünler) arasında benzerlik veya uzaklık ölçüleri hesaplanarak, maddeler arasındaki uyum belirlenir. Daha sonra öneri sunulacak aktif kullanıcılar ile benzer maddeleri beğendikleri, satın aldıkları ürünlerden aktif kullanıcının henüz karşılaşmadığı ürünler için öneri listesi oluşturularak sunulur. Bir kullanıcının önceden deneyim sahibi olduğu maddeler (öğeler-ürünler) ile benzer ürünleri tercih edeceği temel fikrinden faydalanılır. Madde tabanlı, madde benzerliğine dayalı işbirliğine dayalı filtrelemede temel yaklaşım için örnek Çizelge 3.3’de verilmiştir.

Çizelge 3.3 Madde Tabanlı (User-İtem) Matrisi

Dersler	Ders-1	Ders-2	Ders-3	Ders-4	Ders -5
Öğrenciler					
Öğrenci-A	✓	?	✓	?	?
Öğrenci-B		✓			✓
Öğrenci-C	✓		✓		✓
Öğrenci-D	✓			✓	✓
Öğrenci-E		✓		✓	

Öneri sistemlerinde içerik tabanlı filtreleme yönteminde temel yaklaşım madde (ürün) için daha önce izlenme, satın alma veya beğeni durumlarına göre en çok izlenen, satın alınan veya beğenilen ürün önerisi yapılır. Çizelge3.3’de verilen örneğe kullanıcı-ürün matrisine göre, A Öğrencisi için yeni bir Ders önerisi yapmak istenmektedir. Bu Amaçla A Öğrenci için daha önce aldığı Ders-1, C ve D Öğrencileri tarafından da alınmıştır. Bu bilgiden hareketle A’nın almadığı dersler arasında Ders-5 olduğu tespit edilmiştir. O halde C öğrencisi ile D öğrencisi tarafından alınan Ders-5, Öğrenci-A’ya alması için önerilir.

3.9.2. Model-tabanlı algoritma

Model tabanlı algoritmalarda (kullanıcı-ürün) matrisindeki bütün verilerin kullanılması yerine (kullanıcı-ürün) matrisini temsil edecek bir model oluşturulur. Oluşturulan bu modele bağlı olarak kullanıcı-kullanıcı veya ürün-ürün tabanlı işbirliğine dayalı filtreleme teknikleri kullanılarak, benzerlikler çevrimdışı hesaplanır. Benzerliklere göre kümeleme analizi yapılarak, kullanıcılar ya da ürünler kümelere ayrıştırılır

(Adomavicius ve Tuzhilin, 2005). Kümeleme tabanlı yaklaşımda kullanıcılar benzerliklerine göre kümelere ayrılırlar. Aktif (çevrimiçi) kullanıcı sistemden bir öneri talep ettiğinde, kullanıcının veya öneri talep ettiği ürünün kümesi belirlenir. Küme tespitinden sonra, kümede yer alan diğer kullanıcıların verilerinden yararlanılarak öneri hesaplanır. Model-tabanlı algoritmalar (kullanıcı-ürün) matrisinin tamamını kullanmadığından, veri kaybı problemi nedeniyle, öneri kalitesinde düşüşler görülebilir. Model-tabanlı işbirliğine dayalı filtrelemede birçok kümeleme tekniğinden yararlanır. En yaygın olarak kullanılan Bellek-tabanlı algoritma KNN (k en yakın komşuluk) algoritmasıdır (Su ve Khoshgoftaar, 2009).

Kullanıcı tabanlı öneri sistemi algoritmaları, farklı kullanıcılar arasındaki benzerlik derecesine göre yapılır. Belirli bir kullanıcının benzer zevklere sahip kullanıcılara tanımlar ve bu kullanıcıların beğendiği öğeleri önerir (Ricci, vd., 2010). Algoritma kullanıcıların ürünler için verdikleri bir dizi ürün değerlendirmeleri bir veri tabanında toplar ve daha sonra, bir kullanıcının henüz görmediği öğeler hakkında tercihlerini tahmin etmek için o kullanıcının tercihlerine benzer diğer kullanıcılardan alınan derecelendirmelerini tavsiye kararları olarak kullanır (Ricci, vd., 2010). Kullanıcı-Tabanlı İşbirliğine dayalı Filtreleme Yöntemi; İşbirliğine dayalı filtrelemede (user-user) Kullanıcı benzerliğine dayanan bu yöntemde, kullanıcılar arasında benzerlik ve uzaklık ölçüleri hesaplanarak, kullanıcılar arasında benzerlikler (komşuluklar) belirlenir (Koren, vd., 2009). Daha sonra, öneri sunulacak aktif kullanıcı ile benzer kullanıcıların beğenileri, satın aldıkları ürünler arasından aktif kullanıcının henüz karşılaşmadığı ürünler için öneri listesi oluşturularak sunulur.

Çizelge 3.4 Kullanıcı Tabanlı (User-İtem) Matrisi

Dersler	Ders-1	Ders-2	Ders-3	Ders-4	Ders-5
Kişiler					
Öğrenci-A	✓	?	✓	?	?
Öğrenci-B		✓		✓	✓
Öğrenci-C	✓		✓		✓
Öğrenci-D		✓		✓	✓
Öğrenci-E				✓	✓

✓ =Öğrencilerin daha önce satın aldıkları kitapları göstermektedir.

Bir kullanıcının ürünler için beğenilerinin, başka kullanıcıların beğenileriyle Kullanıcı tabanlı, kullanıcı benzerliğine dayalı işbirliğine dayalı filtreleme temel yaklaşım için bir user-item matrisi Çizelge 3.4’de verilmiştir.

Öneri sistemlerinde kullanıcı tabanlı yöntem olan İşbirliğine dayalı filtrelemede temel yaklaşım kullanıcıya en yakın komşuyu belirlemektir. Çizelge 3.4’de verilen kullanıcı-ürün matrisine göre, A Öğrencisi için yeni bir ders önerisi yapmak istenmektedir. Bu Amaçla A Öğrenci için daha önce benzer dersleri alan Öğrencinin belirlenmesi gerekir. Çizelge 3.4’de incelendiğinde A Öğrencisine en yakın C Öğrencisidir. Çizelge 3.4’deki bilgilere göre A ve C Öğrencileri birbirlerine benzerlerdir. A ve C Öğrencileri aynı kümede yer alırlar. C Öğrencisinin aldığı dersler ile A Öğrencisinin aldığı dersler benzerlik göstermektedir. O halde benzer dersleri alan ve aynı kümede yer A ve C Öğrencileri için Öğrenci-C’nin aldığı Ders-4, Öğrenci-A’ya önerilir.

3.9.3. Hibrit (melez) algoritma

İşbirliğine dayalı algoritmaların kısıtlamalarını ortadan kaldırmak için Bellek ve Model tabanlı yaklaşımların avantajlarını bu model de birleştirilir. Hibrit algoritmaların öneri performansı bellek ya da model tabanlı algoritmaların performansları arasında yer alırken, bazı model ve bellek tabanlı işbirliğine dayalı filtreleme algoritmalarından daha iyi öneri performansı gösterdiği düşünülmektedir (Su ve Khoshgoftaar, 2009).

3.10. İşbirliğine Dayalı Filtrelemede Öneri Hesaplama Temel Yaklaşımları

İşbirliğine dayalı filtrelemede temel yaklaşım, hedef kullanıcı için yapılacak ürün önerisinde kullanıcının geçmiş bilgilerine dayanarak en yakın komşu ya da komşuların oluşturduğu küme veya kümelerin belirlenmesidir (Vialardi, vd., 2009). Küme, birbirlerine yakın bireylerin çok boyutlu uzayda oluşturdukları birlik olarak ifade edilebilir. İşbirliğine dayalı filtrelemede benzerlik, hedef kullanıcının beğeni puanları ile diğer kullanıcıların beğeni puanları benzerliğine (komşuluk) dayanan veya madde (ürün) benzerliklerine bağlı olarak iki şekilde kestirim yapılabilir (Jahrer, vd., 2010). Bunlar Kullanıcı tabanlı filtreleme ve madde tabanlı filtreleme olmak üzere iki yöntem içerir.

3.11. İşbirliğine Dayalı Filtrelemede Kullanılan İstatistiksel Teknikler

Netflix Ödüllü yarışma rekabet için uygulanan iyi bilinen bir yaklaşımdır. Bu yarışma sonucunda birçok şekilde çeşitli algoritmalar tavsiye sistemleri ve kullanılabilir olduğunu göstermiştir. Elde edilen sonuçlar, belirli bir sorun çözümünde tasarlanmış algoritmalarıdır. Bu algoritmalar verilerin mevcut ilişkiye bağlı olarak değişebilir. Bu algoritmaların birçoğu öğrenme, tahmin ve karar verme algoritmaları ile uğraşan, yapay zekâ ve bir alt uygulaması makina öğrenme bağlamında geliştirilmiştir (Jahrer, vd., 2010).

Öneri sitemlerinde kullanıcıya veya ürüne ait beğeni puanları bir vektör veya dizi ile ifade edilmektedir. İşbirliğine dayalı filtrelemede hedef kullanıcı için, sistemdeki m tane ürün içinde olan ancak kullanıcının henüz karşılaşmadığı ürünler için beğeni puanları kestirilir. Kestirim yapılabilmesi için hedef kullanıcının ve diğer kullanıcıların geçmiş bilgilerinden (beğeni puanlarından) yararlanır. Bu kullanıcılar komşu adı verilen bir grup

oluştururlar. Bir kullanıcı, daha önce oy kullanmadığı, ancak komşuluğundaki kullanıcılar tarafından olumlu değerlendirilen öğeler için öneri üretilir. İşbirliğine dayalı filtreleme tarafından üretilen öneriler ya öngörü ya da tavsiye olabilir. Tahmin, kullanıcı i için j maddesinin öngörülen skorunu ifade eden sayısal bir değer R_{ij} 'dir (Hernández-del-Olmo ve Gaudio, 2008), Kestirim sayısal bir değerdir ve örneğin beğeni derecelendirme 1-5 arasındaysa, belirli bir ürünün belirli bir kullanıcı için beğeni kestirimi aynı aralıktaki bir rakam ile belirlenir. Kestirilen bu beğeni puanı değerine göre öneri yapmak olanaklı olur. Öneri, aktif kullanıcının en fazla beğeneceği kestirilen n tane ürünün listesidir. Kestirim yapılan n tane madde için öneri listesi oluşturur. Öneri listesinde tüm ürünleri önermek kullanıcı için bir fayda sağlamayacağından, beğeni puanı (ya da öneri puanı) en büyük olan ürünlerden oluşan Üst- N öneri listesi oluşturulur (Resnick ve Varian, 1997).

3.11.1. Kümeleme analizi

Kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine göre belirli gruplar içinde toplayarak sınıflandırma yapmayı, birimlerin ortak özelliklerini ortaya koymayı ve bu sınıflar ile ilgili genel tanımlar yapmayı sağlama bir yöntemdir (Kaufman ve Rousseeuw, 1990). Kümeleme, birincil amacı özelliklerine göre birimleri gruplamak olan çok değişkenli istatistik tekniğinden biridir. Kümeleme analizinin genel amacı, gruplanmamış verileri benzerliklerine göre sınıflandırmak ve araştırmacıya uygun, işe yarar özetleyici bilgiler elde etmede yardımcı olmaktır. Kümeleme analizinde amaç, benzer özellikleri gösteren birimleri aynı kümede toplayarak, özet bilgi elde etmektir. Kısa bir tanımla kümeleme analizi doğal sınıflamaları hakkında kesin bir bilgi bulunmayan durumlarda, kütleliyi oluşturan birimlere ilişkin tahminlerin yapılmasında kullanılan teknikler topluluğudur (Hartigan, 1975). Birimlerin benzerliğini ortaya koymak için çeşitli uzaklık matrisleri kullanılır. Kümelemede birimlerin gruplanması önceden belirlenmiş bir ölçüte göre gerçekleştirilir. Kümeleme sonuçları, kümeler içinde yüksek derecede homojenlik, kümeler arasında ise yüksek derecede heterojenlik göstermelidir. Kümelemenin odak noktası rastlantı değişkeninin tahmini değil, rastlantı değişkenine dayalı olarak birimlerin karşılaştırılmasıdır. Ancak, ikinci bir kullanım alanı olarak da, kümeleme analizinin değişkenlerin kümelenebilmesi için kullanılabileceği bilinmektedir (Johnson ve Wichern, 2007). Kümeleme yapısı içerisinde rastlantı değişkeni kavramının önemli bir yeri vardır. Ancak, rastlantı değişkeninin kümeleme yapısı içindeki yeri diğer çok değişkenli tekniklerde

olduğundan tamamen farklıdır. Küme rastlantı değişkeni, kümelemede birimleri karşılaştırmak için kullanılan özellikleri simgeleyen değişkenler kümesidir (Theodoridis ve Koutroumbas, 2008). Kümeleme analizinde verilerin normal dağılımlı olması gerektiği varsayımı olmakla birlikte normallik varsayımı prensipte kalmakta, uzaklık değerlerinin normalliği yeterli görünmektedir. Ayrıca kovaryans matrisine ilişkin herhangi bir varsayım bulunmamaktadır (Tatlidil, 2002) . Kümeleme analizinde gruplandırma benzerlik ve farklılıklara göre yapılır. Girdiler benzerlik ölçülerini veya veriye hangi benzerliklerin hesaplanabilmelerinin gerekliliğidir (Johnson ve Wichern, 2007). Kullanıcının amacına ve kullanım alanına göre kümeleme analizinin amaçları aşağıdaki gibidir (Field, 2009).

1. Doğru türlerin belirlenmesi
2. Model oluşturmak
3. Gruplara dayalı tahmin
4. Hipotez testi
5. Veri araştırma ve aykırı değerleri belirleme
6. Hipotez oluşturma
7. Veri indirgeme

Kümeleme Algoritmaları, bir tür rastgele (veya işaretli) veri saflarının yapısını belirlemek için "kendiliğinden öğrenme" (denetimsiz öğrenme) algoritmalarıdır. Genel olarak, bu tür bir algoritma özelliği kişiler aralarındaki mesafeyi hesaplayarak maddeler arasındaki benzerlikleri belirlemeye dayanmaktadır. Bağımsız birimlerin (niteliklerin) özellik sayısı küme alanı boyutunu tanımlar. Birimler birbirlerine "yakınlarsa", bunlar tek bir küme halinde kombine edilebilir. Kümeleme için birçok algoritma vardır. Bunların en basit algoritma k ortalama (k: küme üzerinde öğelerini ayıran), k küme anlamına gelmektedir. Başlangıçta, elemanları rasgele sırayla bu kümelere dağıtılır. Sonra küme merkezine her küme üyesi mesafesini kontrol eder. Bu onay sonucunda birim sonuçları diğer kümeye üyelerine daha yakın ise, birim diğer kümeye taşınır. Küme merkezleri tüm üyelerine tüm mesafeleri kontrol ettikten sonra yeniden hesaplanır. Küme merkezleri hareket etmiyor ve kararlı duruma ulaştıktan sonra kümeler uygun şekilde oluşturulur ve algoritma durur (Field, 2009).

İki nesne arasındaki mesafenin hesaplanması görselleştirmek zor olabilir. Bu sorunu çözmek yaygın bir yöntem çok boyutlu bir vektör olarak kümenin her üyesi için hesaplamada Öklid mesafe adaptif rezonans teorisi (Adaptif Rezonans Teorisi), bulanık kümeleme yöntemi C-ortalama (Bulanık C), EM-algoritma (Beklenti-Maksimizasyon) ve olasılıklı kümelene dâhil kümelene ve diğer birçok varyasyonu vardır (Johnson ve Wichern, 2007).

3.11.2. K-ortalamlar algoritması

Komşuluğa dayalı k-ortalamlar algoritması, hata fonksiyonunu minimum yapılmasına dayalı olarak verideki en uygun parçalanmayı bulmayı amaçlayan adımsal optimizasyon yöntemidir. Kümeleme analizinde k -ortalamlar algoritması en yaygın olarak kullanılan algoritmadır (Theodoridis ve Koutroumbas, 2008). 1967 yılında Mac Queen tarafından bulunan k-means (ortalamlar) algoritması, kümeleme problemini çözen en basit denetimsiz öğrenme (herhangi bir öğrenme olmaksızın) algoritmalarından biridir. İnıflayıcı kümeleme tekniklerinden biri olan k-means, bilimsel ve endüstriyel uygulamalarda en yaygın olarak kullanılan kümeleme algoritmaları arasında yer almaktadır. Kümeler, benzer özellik gösteren veri elemanlarının kendi aralarında gruplara ayrılmasıdır.

K-ortalamlar algoritmasının genel mantığı n adet veri nesnesinden oluşan bir veri kümesini(X), giriş parametresi olarak verilen k($k \leq n$) adet kümeyle bölümlenmektedir. K-ortalama kümelemesinde, bu kümelerin tanımlanma şekli, her grup için bir (merkez) centroid oluşturmaktır. Centroidler kümenin kalbi gibidir, kendilerine en yakın noktaları “yakalar” ve kümeyle eklerler. Amaç, gerçekleştirilen bölümlenme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olmasını sağlamaktır (Everitt, vd., 2011). K-ortalama kümelemesi işlem adımları aşağıdaki gibi tanımlanır;

Giriş verisi:

Veri tabanı, $X = \{x_1, x_2, x_3, \dots, x_n\}$

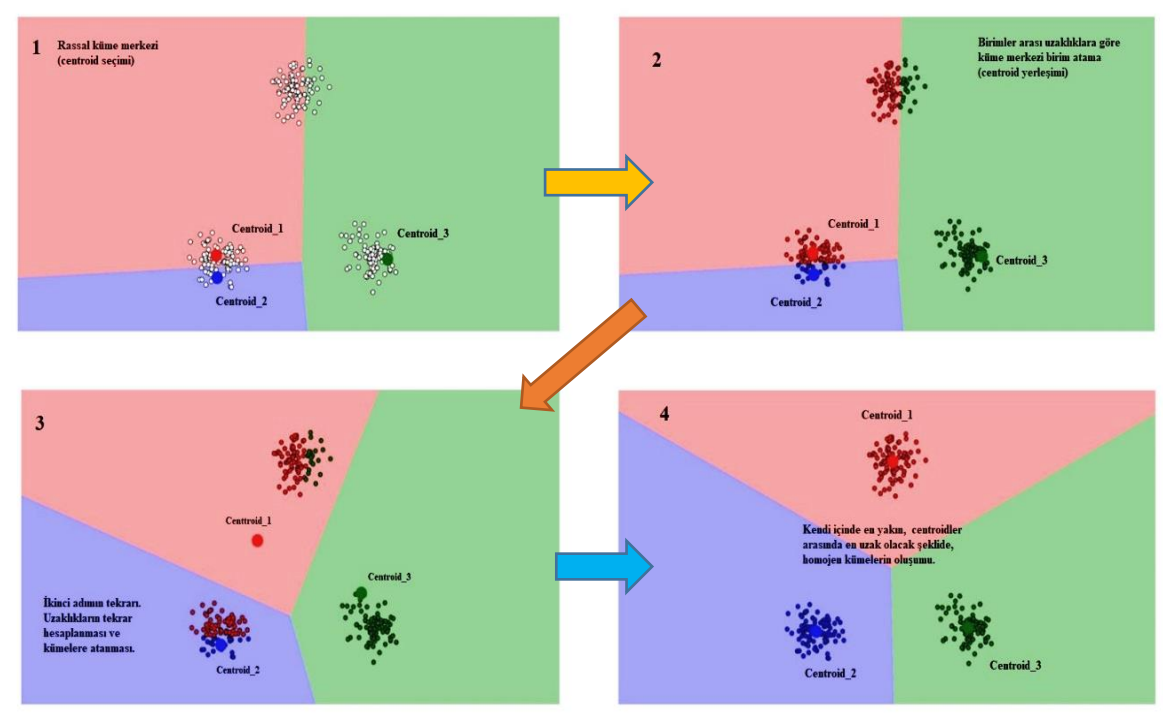
n: veri nesnelere.

Cıktı:

1. N veri setleri kaynaktan k nesnelere rasgele seçilmiştir ve küme merkezi vardır ($m_1, m_2, m_3, \dots, m_k$),
k: k-centroid yada k küme sayısı olarak tanımlanır.
2. Her nesne ve küme merkezi arasındaki mesafe hesaplanır ve ardından her nesne en yakın kümeye atanır.

K-ortalama algoritmasında, verinin kümelenebilmesi küme sayısı k 'nın belirlenmesi ile başlar. Algoritmanın ikinci adımında küme başlangıç merkezleri seçilir ve birimlerin küme merkezlerine olan uzaklıkları hesaplanır. Genellikle küme merkezlerinin belirlenmesinde aritmetik ortalama, birimlerin küme merkezlerine olan uzaklıklarının hesaplanmasında Öklid uzaklığı kullanılmaktadır (Gan, vd., 2007). Küme merkezlerine yakınlığa göre birimler kümelere atanır ve yeni oluşan kümelerin merkezleri belirlenir. Belirlenen küme merkezlerine olan uzaklıklar hesaplanarak tekrar atamalar yapılır. Bu süreç küme üyelikleri sabitleninceye kadar veya hata fonksiyonunun değeri değişmeyinceye kadar devam eder (Everitt, vd., 2011). K-ortalama küme algoritmasının çalışma prensibinin görselleştirilmesi Şekil 3.6'da verilmiştir

Düzlemdeki her nokta en yakın olduğu merkeze (centroid) göre atanır. Centroid'lerin (mavi, kırmızı ve yeşil büyük noktalar) rastgele başladığını ve ardından ilgili kümelerini yakalamak için hızlıca ayarlandığını göstermektedir.



Şekil 3.6. K-ortalamlar algoritması kümeleme adımları görselleştirilmesi.

3.11.3. En yakın komşuluk (KNN) algoritması

K-en yakın komşu algoritması (KNN) eğitici ve örnek tabanlı bir sınıflandırma algoritmasıdır. K en yakın komşu algoritması, sınıflandırma problemini çözen denetimli öğrenme (sınıflandırma için öğrenme kümesi kullanır) algoritmalarından biridir. Sınıflandırma, yeni bir nesnenin özelliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır (Everitt, vd., 2011). Burada önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır. K en yakın komşu yönteminde; sınıflandırma yapılacak verilerin öğrenme kümesindeki normal davranış verilerine benzerlikleri hesaplanarak; en yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları yapılır. Diğer yöntemlere göre daha kolay uygulanır. Bu tip algoritmalarda eğitime ihtiyaç yoktur. KNN, veri madenciliği, bilgi güvenliğinin sağlanmasında saldırı tespit sistemlerinde, coğrafi bilgi sistemlerinde, genetik ve biyoinformatik alanında, örüntü tanıma sistemleri gibi birçok benzeri sistemde kullanılmaktadır (Vialardi, vd., 2009).

KNN’de bir vektörün sınıflandırılması, sınıfı bilinen vektörler kullanılarak yapılmaktadır. Test edilecek örnek, eğitim kümesindeki her bir örnek ile tek tek işleme alınır. Test edilecek örneğin sınıfını belirlemek için eğitim kümesindeki o örneğe en yakın K adet örnek seçilir. Seçilen örneklerden oluşan küme içerisinde hangi sınıfa ait en çok örnek varsa test edilecek olan örnek bu sınıfa aittir denilir. Örnekler arası uzaklıklar Öklid uzaklığı ile bulunur. Öklid uzaklığı kullanılarak hesaplanan tüm uzaklık değerleri sıralanır. Sıralı değerler arasından K sayısına bağlı olarak en küçük K tanesi belirlenir. Test edilecek örneğe en yakın K tane komşu örnek belirlenmiş olur. Test edilecek örneğin sınıflandırılması için bulunan K tane komşunun sınıf etiketleri kullanılır (Hartigan, 1975). Sınıf etiketlerinin “1” ve “0” olarak belirlendiği varsayılırsa; test edilecek örnek ile eğitim örneklerinin arasındaki uzaklık değerleri hesaplandıktan sonra, K tane en yakın örneğin sınıf etiketlerine bakılır. Sınıf etiketi “1” olanlar “0” olanlardan fazla ise test örneğinin sınıfı “1” dir, “0” olanlar “1” olanlardan fazla ise “0” olarak sınıflandırılır. Testin sınıfına karar verilmesi esnasında K değerini seçmek için iki durum yaşanabilir. Birinci durumda K değeri tek sayı seçilerek “+1” ve “0” örneklerin sayısının eşit değerde çıkması önlenir. K değeri çift sayı seçilirse de K tane örnek için her bir sınıfa ait örnekler kendi aralarında toplanır ve ortalamaları bulunur. En küçük ortalamaya sahip olan sınıf, test edilecek örneğe daha yakın olacağı için test örneğinin sınıfı en küçük ortalamaya sahip olan sınıf olacaktır. Bu algoritma için sınıf sayısında bir kısıtlama yoktur. İstenilen sayıda sınıf belirlenerek sınıflandırma işlemi yapılabilir (Koren, vd., 2009).

Kümeleme, nesnelerin gruplarını o şekilde bulur ki, bir grup içindeki nesneler birbirlerine benzerken, başka gruptaki nesnelere farklıdır. Kümeleme en önemli denetimsiz (unsupervised) öğrenme tekniğidir. Kümelene ve aşırı uç tespiti için kritik noktaları, arasındaki yoğunluk ve mesafe kavramları çok boyutlu uzaylarda daha anlamlı hale gelir. Bu olarak bilinir Boyutluluk Sorununu. Orijinal yüksek boyutlu uzayı boyut indirgeme teknikleri ile düşük boyutluluk haline dönüştürerek bu sorunu aşmak yardımcı olur.

K en yakın komşu yönteminde; sınıflandırma yapılacak verilerin öğrenme kümesinde ki normal davranış verilerine benzerlikleri hesaplanarak; en yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları yapılır (Theodoridis ve Koutroumbas, 2008).

Kümeleme Analizinde, küme sayısının belirlenmesi konusunda son yıllarda çalışmalar yapılmaktadır, Ancak küme sayısının belirlenmesinde en pratik yöntem n: küme sayısı olmak üzere; denklem (3,9) deki gibi hesaplanır.

$$k = \sqrt{\frac{n}{2}} \quad (3.9)$$

KNN algoritmanın performansını etkileyen kriterler, aşağıdaki şekilde sıralanmaktadır.

- k en yakın komsu sayısı (benzerlik ölçümü için seçilecek komsu sayısı: k),
- Esik değeri (etiketleme işlemi için verinin k en yakın komşuya olan benzerliklerinden hesaplanan ortalama değerinin kıyaslanmasında kullanılır).
- Benzerlik ölçümü
- Öğrenme kümesindeki normal davranışların yeterli sayıda olması öğrenme kümesi yeterli çeşitlilikte ve sayıda normal davranış verisi içermiyorsa, test kümesinde yer alan yeni normal davranış verileri anormal olarak algılanabilir.

3.11.4. Matris ayrışımı algoritmaları

Verinin sahip olduğu boyut sayısı arttıkça, genellikle çok az sayıda boyut doğrudan kümelerle ilgili olur ancak ilgisiz boyutlardaki veri, çok fazla gürültüye sebep olabilir ve bu da keşfedilecek kümelerin gizlenmesine sebep olur. Bundan daha önemlisi boyut arttıkça, verinin seyrekleşmesi sorundur. Sabit sayıda veri noktasının bulunduğu bir veri seti boyut sayısı arttıkça üstel artan bir biçimde seyrekleşir. Bu yüzden verinin boyutu arttıkça kümeleme işlemi zorlaşacağından ve doğru sonucun bulunma ihtimali azalacağından veri setinin boyutunun azaltılması gerekmektedir. Bu nedenle verinin analizinde özellikle veri madenciliğinde çeşitli uygulamalarda boyut azaltma tekniklerine ihtiyaç duyulmaktadır. Son zamanlarda, matris ayrıştırma modelleri, onların çekici doğruluk ve ölçeklenebilirlik sayesinde popülerlik kazanmıştır (Koren, vd., 2009). Matris ayrışımı karmaşık bir matrisi daha basit matrislerin çarpımına dönüştüren bir yöntemdir. Pratikte boyut indirgeme, gerek maliyetleri düşürmek gerek içerik temsiliyi yalınlaştırarak başarıyı artırmak açısından önemlidir (Golub ve Loan, 1996).

Bunun için literatürde bulunan birçok ayrıştırma yöntemi veri boyutunu küçültüp önemli bileşenleri ortaya çıkarmaya çalışır. Matrisin ayrışımında en çok kullanılan algoritmalar ise SVD, QR ve LU ayrışımı algoritmalarıdır (Golub ve Loan, 1996). Ayrıca temel bileşenler analizi “TBA” (principal component analysis, PCA) gösterilebilir (Jolliffe, 1986). Bu tür ayrıştırma yöntemlerinin en bilinenleri “TDA” (singular value decomposition, SVD) olarak tekil değer ayrıştırmasıdır (Baker, 2005).

3.11.5. Temel bileşenler analizi

Çok değişkenli istatistiksel analizlerde n tane bireye (birime) ait p tane değişken incelenmektedir. Bu p tane değişkenin birbiriyle bağımlı olması ve değişken sayısının da p tane olması analizlerde sorun yaratmaktadır (Tabachnick ve Fidell, 2001).

Bir \mathbf{X} veri matrisi için; n birey(gözlem, birim) için p tane değişkeni göstermek üzere, p boyutlu bir uzayda her birey bir noktayı göstermek üzere çok sayıda noktadan oluşan bir topluluk oluşturur. Buradaki p tane değişkenin tam bağımsız olmadıkları ve oluşan geometrik şeklin eksenlerinin bir birine dik olmamasından dolayı tanımı da yapılamayacaktır (Stewens, 1996).

Bu geometrik şekil eksenleri birbirine dik bir elipsoidi içine almak daha ayrıntılı ve açıklayıcı bilgi verecektir. Bu amaçla yapılan dönüştürmede noktaların ilk eksenler boyunca sahip oldukları toplam varyans değişmediği gibi yeni eksenler birbirine dik olmaktadır. Bu da bize analizlerde kolaylık ve birçok değişkenin aslında genel bir değişkenin ölçüm değerleri olduğunu ve bu değişkenlerin genel bir değişken altında sınıflandırılması (boyut) ve hangi değişkenler olduğu bilgisini vermektedir. Bu bağlamda Temel bileşenler analizinin (PCA) ana hedefi gözlemlerin boyutun (değişken sayısı) azaltmaktır (Jolliffe, 1986).

PCA, Pearson (1901) tarafından tanıtılan bu teknik daha sonra Harold Hotelling(1933) tarafından çok değişkenli veri setinin bir biriyle ilişkisiz bir veri setine dönüşü olarak tanımlanmıştır (Tabachnick ve Fidell, 2001). Çok değişkenli veri setinde p tane değişkenin ölçüm düzeyleri farklılık gösterebilir. Bu durumda $n \times p$ ham veri seti yerine standartlaştırılmış bir veri seti kullanılarak işlemler PCA işlemleri uygulanır. Ancak $n \times p$

boyutlu ham veri seti de kullanılabilir ancak pratikte bu çok nadir karşılaşılan bir durumdur (Theodoridis ve Koutroumbas, 2008).

- I. Genel olarak n birimin p tane değişken için bir birleriyle bağımlı (korelasyon) $n \times p$ boyutlu \mathbf{X} ham veri matrisi x_1, x_2, \dots, x_p şeklindedir.
- II. Temel bileşenler analizi sonucunda biri biriyle ilişkisiz p tane yeni dik değişken (temel bileşenler ya da öz vektörler) elde edilir.
- III. Elde edilen temel bileşenlerin birimi yoktur, p değişkenin taşıdığı bilginin k tane genel değişken/değişkenlerin (temel bileşen) ile açıklanması ise temel bileşenler analizinin ana hedefi olan boyut azaltma işlemi gerçekleşmiş olur.

PCA sonucunda elde edilen veri seti artık $n \times p$ değil k tane bileşenden oluşan bir vektördür. Burada n tane bireyin p değişkene ilişkin ölçüm değerlerinden oluşan veri seti \mathbf{X} ile bu değişkenlerin varyansları yardımıyla p değişken ($k \leq p$) tane temel bileşen (öz vektör) elde edilmiştir (Stewens, 1996).

En büyük öz değer, birinci öz vektör, en küçük öz değer ise en son öz vektöre aittir. Böylece pratikte uygulanan \mathbf{X} veri matrisi $n \times p$ temel bileşenler analizi ile bir faktörler matrisi olarak dönüştürülmüştür. (veri sonunda standart ve ortalanmış olacaktır). P boyutlu uzaydaki varyans, öz değerler toplamına eşittir. X_1, X_2, X_p vektörlerinin standartlaştırılmış hali olan Z_1, Z_2, \dots, Z_p vektörlerinin p tane doğrusal birleşimi, ya da temel bileşeni olarak adlandırılır (Wall, vd., 2003) (Koren, vd., 2009).

3.11.6. Tekil değer ayrışımı

Tekil değer ayrışımı (TDA), 3 farklı ve birbiriyle bağlantılı veri durum için kullanılmaktadır. İlk olarak, ilişkili verileri, ilişkisiz veri kümesine dönüştürmesi amacıyla kullanımıdır (Wall, vd., 2003). İlişkisiz kümelerle dönüştürme işlemi; diğer bir ifade ile korelasyonlu verileri, orijinal veri parçaları arasındaki çeşitli ilişkileri daha iyi ortaya çıkaran

korelasyonuz veri kümesine dönüştürme yöntemi olarak tanımlanmaktadır (Wall vd., 2003). İkincisi, TDA hangi verilerin en çok değişim gösterdiği boyutu tanımlama ve sıralama yöntemi olarak kullanılabilir. Üçüncü olarak da, Tekil değer ayrışımı veri madenciliği yapılacak, gereksiz yüzlerce değişkenden oluşan veri kümesinin boyutunu indirgemedede sıkça kullanılan bir yöntemdir. Kullanılmasıyla daha az boyut ile orijinal veri noktalarının en iyi tahmininin yapılması mümkündür (Baker, 2005).

Bir matrisin, pozitif tanımlı bir matris olabilmesi için, A tekil olmayan herhangi bir matris ile matrisin transpozununun A^T çarpımları AA^T şeklinde ifade edilebilmesi gerekir. Matematikte, öz değer ve öz vektör lineer cebir alanında birbiriyle ilişkili kavramlardır. Lineer cebir, vektörler üzerine uygulanan matrisler şeklinde temsil edilen doğrusal dönüşümleri araştırır. Öz değerler ve öz vektörler bir matrisin özellikleridir. Bunlar matris hakkında önemli bilgiler içerir. Bir özvektör, ona doğrusal bir dönüşüm uygulandığında yönü değişmeden kalan bir vektördür.

$A = AA^T$ veya $A^T A$ için tekil olmayan bir A matrisi varsa, A matrisi pozitif tanımlıdır denir. Pozitif tanımlı bir matrisin özdeğerleri (λ , özdeğer skaler bir değerdir.) her zaman 0'dan büyüktür ve karekökleri gerçektir sayılardır.

Tekil değer ayrıştırması, $A \in \mathbb{R}^{m \times n}$ şeklindeki bir A matrisini, iki ortogonal matrise $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ ve bir köşegen matrise $\Sigma \in \mathbb{R}^{m \times n}$ ayırır. Böylece A matrisi, ortogonal matrisler olan U, Σ ve V şeklinde ayrıştırılabilir. Bunun sonucunda ise $A = U\Sigma V^T$ dir. Σ 'nin asıl köşegen değerleri A'nın tekil değerleri olarak adlandırılır. Burada U matrisi sütunları A'nın sol tekil vektörleri ve V matrisi ise sütunları A'nın sağ tekil vektörleri olarak adlandırılır. Bu ayrışma A matrisinin "Tekil Değer Ayrışımı" SVD olarak bilinir. Bu yöntemde, veri matrisi $A_{m \times n}$ Σ , U, V olarak adlandırılan daha küçük boyuttaki üç matrisin çarpımına eşittir (G. Strang, 2009).

$$[A]_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Σ , U , V matrislerinin hesaplanama aşamaları aşağıdaki gibi sıralanır (Baker, 2005).

I. U Matrisinin hesaplanama aşamaları aşağıdaki gibidir.

Aşama 1: AA^T hesaplanır, böylece $A_{m \times n}$ tipindeki A matrisi, $A^T A_{n \times n}$ tipinde karesel matrise dönüşür.

Aşama 2: AA^T matrisinin öz-değerler hesaplanır.

Aşama 3: Hesaplanan öz-değerler yardımıyla öz vektörler hesaplanır.

Aşama 4: Bu öz vektörler bir matrise yerleştirilir. Böylece, U matrisi elde edilir.

II. V Matrisinin hesaplanama aşamaları aşağıdaki gibidir.

Adım 1: $A^T A$ hesaplanır

Adım 2: $A^T A$ matrisinin öz-değerleri hesaplanır.

Aşama 3: Hesaplanan öz-değerler yardımıyla öz vektörler hesaplanır.

Aşama 4: Bu öz vektörler bir matrise yerleştirilir. Böylece, V elde edilir.

III. Σ Matrisin hesaplanama aşamaları aşağıdaki gibidir.

Adım 1: $A^T A$ veya AA^T öz-değerlerinin karekökü hesaplanır

Adım 2: Bu değerleri asıl köşegen olarak yerleştirilir. Diğer değerler sıfır olarak yazılarak, böylece (Σ) matrisi elde edilir.

U = sol orthogonal SVD matris

U_k = k boyutunda kesilmiş sol orthogonal SVD matris

V = sağ orthogonal SVD matris

V_k = k boyutunda kesilmiş sağ orthogonal matris

σ = tekil değer

Σ = tekil değer köşegen matris

Böylece A matrisi üç parçaya ayrılır. A matrisi üç matrisin çarpımı sonucuna eşittir. Denklem (3.10)'da verilen eşitlik elde edilir. Tekil değer ayrışımı (SVD), tekil değerleri azalan düzende düzenleme, $A_{m \times n}$ veri matrisinin sıralı (bileşenli) parçalarını verir (G. Strang, 2009).

$$[A]_{m \times n} = [U]_{m \times r} [\Sigma]_{r \times r} [V^T]_{r \times n} \quad \dots \quad (3.10)$$

Eşitlik (3.10) genişleme, yeniden düzenlenebilir.

$$U = [u_1 \quad \dots \quad u_r]_{m \times r} \quad (3.11)$$

$$V = [v_1 \quad \dots \quad v_r]_{n \times r} \quad (3.12)$$

$$V = \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}_{r \times n}$$

$$(3.13)$$

$$\Sigma = [\sigma_1 \quad \dots \quad \sigma_r] \text{ ve } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 \quad (3.14)$$

Böylece, Tekil değer ayrışımının matris ve vektör şeklindeki gösterimi eşitlik (3.15)' de verilmiştir.

$$[A]_{m \times n} = [U]_{m \times r} [\Sigma]_{r \times r} [V^T]_{r \times n} = [u_1 \quad \dots \quad u_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix} \quad (3.15)$$

Tekil değer ayrışımı ile iki tekil vektör elde edilir. Bu tekil vektörler sırasıyla u ve v matrisleridir. Her iki matris de simetrik olduğundan onların öz vektörleri ortonormaldir.

Tekil değerleri azalan düzende, veri matrisinin önem sırasına eşitlik (3.16)' da gösterildiği gibi göre sıralı (bileşenli) parçalarını verir.

$$A \sum_{j=1}^r u_j \sigma_j v_j^T = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \dots + u_r \sigma_r v_r^T \quad (3.16)$$

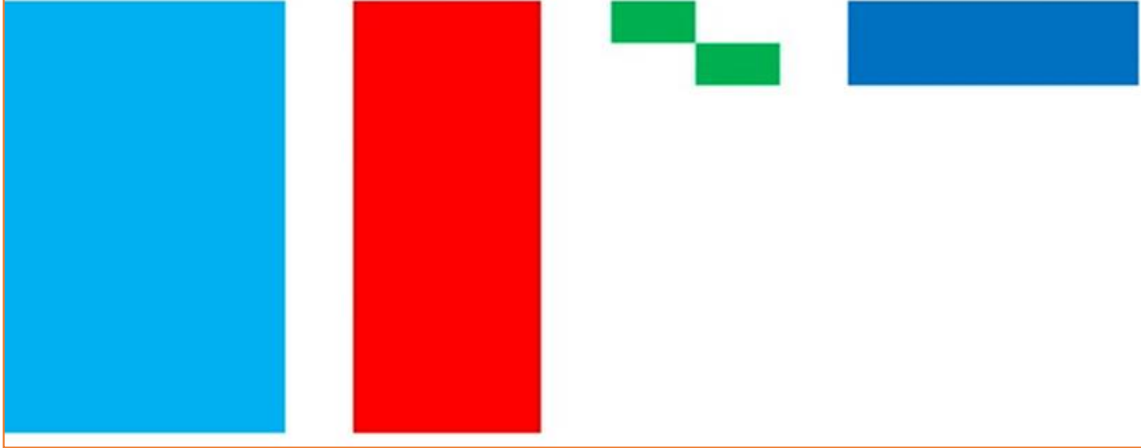
Bir $m \times n$ boyutlu A matrisinin tekil olmayan en büyük boyutlu kare alt matrisinin $r \times r$ boyutuna A matrisinin rankı denir. ($r \leq m$, $r \leq n$) $\text{Rank}A=r$ şeklinde gösterilir. Tekil değer ayrışımına (3.16) verilen eşitlik, matrisin determinantı sıfırdan farklı olan en büyük karesel matris sayısı r boyutu için hesaplanmıştır. Ancak uygulamada r kadar matris için hesaplamak çok zahmetli ve uzun bir işlem süresi gerektirir. Bu nedenle $k < r$ olmak üzere k boyutta kesilmiş matris ile işlem yapmak ve modellemek daha kolaydır. Ancak kesilmiş (truncated) bir matrisle çalışmak özellikle resim ve dijital görüntü üzerinde ise, boyut indirgmeden dolayı ciddi bir nitelik kaybı olabilir (Wall, vd., 2003). Bu nedenle çok küçük k değerleri ile çalışmak veya büyük k değerleri ile çalışmak yapılan araştırmanın türüne göre değişir.

Kesilmiş (truncated) k boyutta işlemde tekil değer ayrışımı işlemi sonucu eşitlik (3.17)'de verildiği gibidir.

$$[A]_{m \times k} = [U]_{m \times k} [\Sigma]_{k \times k} [V^T]_{k \times n} = [u_1 \quad \dots \quad u_k] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix} \quad (3.17)$$

Tekil değer ayrışımı ile sahip olduğumuz orijinal veri matrisi içerisinde k gizil temanın (ya da ilişkinin) olduğunu varsayımı ile k boyutlu $[A]_{m \times k}$ matrisinin tekil değer ayrışımı sonucu eşitlik (3,18)'deki gibidir.

$$[A]_{m \times k} = A \sum_{j=1}^k u_j \sigma_j v_j^T = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \dots + u_k \sigma_k v_k^T \quad (3.18)$$

$$\begin{array}{l}
 A_{(m \times n)} = U_{(m \times k)} \times \Sigma_{(k \times k)} \times V^T_{(k \times n)} \\
 A_{(10 \times 4)} = U_{(10 \times 2)} \times \Sigma_{(2 \times 2)} \times V^T_{(2 \times 4)}
 \end{array}$$


Şekil 3.7. Tekil değer ayrışımının matris ve vektör şeklindeki gösterimi

Eşitlik (3,18)'de k değeri için, daha küçük boyutları düşünürsek daha genel temalar elde edebiliriz, daha fazla boyut düşünürsek daha detay temalar oluşturabiliriz. Ancak k değerini çok fazla büyütürsek gürültülü boyutlarla sonuçlanacaktır (Sarwar, vd., 2000). Tıpkı bir yağlı boya tabloyu ne kadar büyüterek bakarsak fırça izlerini daha net görürüz. Çünkü orijinal boyut değişirse de gürültü artacaktır. Bu nedenle TDA'a da k değerinin seçilmesi önemlidir. Bunun için kestirim için elde değerler ile gerçek değerler arasındaki hata minimum olduğunda veya hata değeri stabil olduğu noktayı optimum k değeri olarak almak gerekir (Koren, vd., 2009).

3.12. Öneri Kalitesi

Öneri sistemleri son yirmi yılda geliştirilip iyileştirilmiş geniş bir algoritma çeşitliliği ile öneri üretebilecek bir noktaya gelmiştir. Her bir algoritma yaklaşımının belirli amaçlar için üstün olduğunu savunan yandaşları vardır. Açıkça belirli bir amaç için en iyi algoritmayı belirlemek önemli bir sorundur (Hernández-del-Olmo ve Gaudioso, 2008). Çünkü bu alanda çalışan araştırmacılar söz konusu amaçlar için hangi özelliklerin ölçülmesi ve bu özellikler için hangi ölçülerin kullanılması konusunda anlaşmazlık içindedirler.

Literatürde birçok nicel ölçü ve ek olarak nitel değerlendirme teknikleri bulunmaktadır. Bu nedenle öneri sistemlerini ve bunların kullandığı algoritmaları bu aşamada değerlendirmek zordur. Bu zorlukların başlıca nedenlerinden birisi farklı veri kümelerinde farklı algoritmaların iyi veya kötü sonuçlar vermesidir. Birçok işbirliğine dayalı filtreleme algoritması kullanıcı sayısının madde (ürün) sayısından fazla olması için tasarlanmıştır. Örneğin, MovieLens'in en son 2016 veri setinde 33 bin film için 580 bin etiket, 240 bin kullanıcının 22 milyon reyting (beğeni) puanı vardır. Bu tür algoritmalar madde(ürün) sayısı kullanıcı sayısından fazla olduğu durumlar için uygun değildir. Bir diğeri, önerinin yapılmasındaki amaçların farklılığından kaynaklanmaktadır. Bir öneri sisteminde, ister kullanıcı veya ürün benzerliğine dayalı, isterse içerik tabanlı filtrelemede kişinin beğeneceği bir ürün (Ders, Dersler, film, Web sayfası) önerildiğinde kişinin tercihlerinin de kesişmesi ile iyi bir öneri sistemi sağlanmış olur (Ricci, vd., 2010).

Öneri Sistemi hatalarının iki ayrı tipi vardır. Kullanıcı tarafından beğenilecek ya da satın alınmak istenen ürün önerisinin yapılmaması yanlış negatif olarak adlandırılır. Yanlış negatif hatası e-ticarette müşteri kaybına neden olur. Burada satın alınma olasılığı yüksek ürün önerilmediğinden, ürün satışı yapılamaz. Kullanıcının beğenmeyeceği bir ürünün önerilmesi yanlış pozitif olarak adlandırılır. Özellikle e- ticarete yanlış pozitif önemli hataların başında gelmektedir. Çünkü Müşteriye satın almayacağı bir ürün önerisi yapılmaktadır. Bu kızgın müşterilere yol açar, sisteme olan güveni azaltır.

Bu bölümde, bilgi filtreleme sistemlerini değerlendirmeden önce kullanılan kuramsal ve deneysel metrikleri inceleyeceğiz. Araştırma alan yazınında İşbirliğine dayalı filtreleme algoritmaları 1994'ten beri değerlendirilmektedir. Bu zaman zarfında pek çok farklı İşbirliğine dayalı sistemi geliştirilmiştir. Birkaç istisna dışında, o günden beri yayımlanan çalışmada her bir sistemin değerlendirilmesi için farklı bir metrik kullanılmıştır. Kullanılan her bir ölçütün güçlü ve zayıf noktalarına vurgu yaparak inceleyeceğiz. Öneri sistemleri değerlendirilmesinde dikkate alınacak temel özellikler ve açıklaması maddeler halinde verilmiştir (Hernández-del-Olmo ve Gaudioso, 2008).

- 1) **Güven/itimat;** Güven, sistemin kendi tavsiyelerine güvenini/itimadını ifade etmektedir. Kullanıcılar bu tür bilgilerden istifade edebilir; örneğin sistem aynı

değerli iki madde önerdiğinde ve bu maddelerin güven aralıkları farklı ise, kullanıcı bir maddeyi doğrudan sepete atıp diğeri hakkında daha özellikli bilgileri okuyabilir.

- 2) **Güven/İnanç;** Bu özellik kullanıcının tavsiyelere güvenini ifade etmektedir. Örneğin kullanıcıya onun hâlihazırda beğendiği maddeleri göstermek yararlı olabilir. Bu türden tavsiyeler ek bir değer yaratmasa da kullanıcının güvenini ve ilgisini artırır.
- 3) **Yenilik;** Bu özellik sistemin tercih değişimlerine uyum sağlayabilme yeteneğiyle ilgilidir. Örneğin Bilim Kurgu filmlerini beğenen bir kullanıcının ilgisi komedi filmlerine kayabilir. Bu niceliği ölçmenin yolu puanlar kümesini sabit bir tarihle iki alt kümeye ayırmak ve o tarihten sonraki puanlamaların tamamını ve o tarihten önceki kimi puanlamaları saklamaktır/gizlemektir. Sonrasında öğrenilenleri uygulamak ve sistemi ikiye ayırma tarihinden önce beğenilen ürünler tavsiye etmemem konusunda cezalandırarak değerlendirmektir.
- 4) **Tesadüfilik;** Tesadüfilik başarılı tavsiyenin ne kadar şaşırtıcı, beklenmedik olduğunu ifade eder. Kullanıcıya uygun ilave bilgilerin miktarını tanımlar. Bu faktörde daha iyi sonuçlar elde edebilmek için kişi madde alanında hâlihazırda sıralanmış olanlardan ‘uzakta’ kalan tavsiyelerde bulunan sistemleri ödüllendirebilir. Uzaklık ölçütü olarak farklı özelliklerin sayısı (yönetmen, tür vb.) kullanılabilir.
- 5) **Sağlamlık/Robustness:** Sağlamlık, sistemin sahte bilgilerin farkına varmasıdır. Araştırmaların da gösterdiği üzere, her geçen gün daha fazla insan İnternetteki bilgilere güvenmektedir. Bu nedenle Halkla İlişkiler ve pazarlama şirketleri kendi ürünlerini tutundurarak tavsiyeler konusunda önyargı oluşturmaya çalışırlar. Saldırıları bağımsız bir sistem oluşturmak neredeyse imkânsız olmasına karşın güvenilmez bilgilerden sakınmak üzere bazı yöntemler kullanılabilir. Aynı IP numarasından yapılan pek çok puanlama veya aynı zamanda birkaç ürünü puanlayan pek çok benzer kullanıcılar gib, ne yazık ki, sistem yalnızca belirli saldırılara karşı korunabileceğinden bu özelliğin ölçülmesi ve değerlendirilmesi güçtür.

- 6) **Mahremiyet;** Belirli bir kullanıcıya özel şahsi bir bilginin ifşa edilmesi uygunsuz bir durumdur. Bu tür bilgilerin ifşasından kaçınabilmek için, kullanıcı tercihlerinin kamuya (veya benzer tercihleri olan kullanıcılara) sızmasına dayalı kuramsal senaryolar kontrol edilmeli ve derhal bloklanmalıdır.
- 7) **Hız ve Ölçeklenebilirlik / scalability;** Uygulamaların büyük bir yoğunluğunda, kapsamlı veri setlerinin ölçeklenebilmesi önemli bir özelliktir. Online tavsiyeler üreten sistemler için hız hayati önemdedir. Hesaplamaların karşılaştırılabilirliği özelliği de uygulamalarda dikkate alınan bir unsurdur.
- 8) **Etkileşim kolaylığı;** Tavsiye sistemleri insanlar için tasarlandığından, kullanımlarının kolay olması beklenir. Birkaç soru sormak büyük yetenekler sağlayan bilgiler verebilir ancak bu durum kullanıcıyı rahatsız etmemelidir. Bu özellik oldukça öznel olduğundan kullanıcı araştırmalarıyla ölçülür.
- 9) **Tavsiye edene faydaları;** Bütün özellikler arasında bir dengenin çıkarılması veya bir özellik üzerinde yoğunlaşmanın yaygın bir yolu fayda fonksiyonudur.
- 10) **Kullanıcı faydası ve riski;** Modellenmesi zor olsa da kullanıcı faydasını hesaplamayı deneyebiliriz. Eğer performans fonksiyonu anlaşılacak olunursa risk de hesaplanabilir ve zararlı öneriler cezalandırılabilir veya ödüllendirilebilir.

3.13. Öneri Sistemleri Değerlendirme Ölçümleri

Öneri sistemleri filtreleme ve geri çağırma teknolojilerine alternatif bir ara yüz sunmaktadır. Öneri sistemlerinin tek bir odağı vardır: bir kullanıcının hangi bilgileri ilginç veya yararlı bulacağını tahmin edilmesi. Bir tavsiye sisteminde üretilen tahminler her bir kullanıcının genellikle önceden görülmüş ürünlere (maddelere) ilgileri ve ileride görecekları ürünlere ilgileri tahmini arasındaki farklılığın ölçülmesi gerekir. Bu nedenle öneri sistemi değerlendirme ölçütleri ilerleyen bölümlerde açıklanmıştır.

3.13.1. Ortalama Mutlak Hata (MAE)

Öneri sistemlerinde, kullanıcıların derecelendirme tahmin sisteminin doğruluğunu değerlendirmek için, yaygın olarak Ortalama Mutlak Hata (MAE) metriği kullanılır. MAE, kullanıcının gerçek değerlendirmesi ile tahmin değerlendirmesi arasındaki ortalama mutlak sapma değeridir (Herlocker, vd., 2004). Mutlak Ortalama Hata; burada n , test seti için seçilen kullanıcı sayısı, s_i bir öge için tahmin değeri ve r_i gerçek derece değeridir.

$$r_{u,i} = \hat{r}_u + \frac{\sum_{v \in \text{Komşular}} (r_{v,i} - \hat{r}_u) \times \text{Sim.}(u,v)}{\sum_{v \in \text{Komşular}} |\text{Sim.}(u,v)|} \quad (3.11)$$

$$\text{MAE} = \frac{\sum_{i=1}^N |r_{u,i} - \hat{r}_{u,i}|}{N} \quad (3.12)$$

Ortalama Mutlak Hata (Mean Average Error: (MAE) değeri ne kadar düşük ise, yapılan öneri o kadar tutarlıdır. Ayrıca İşbirliğine dayalı filtreleme algoritması ile doğru başarı puanını tahminin çok düşük hata ile yapıldığının bir göstergesidir.

$$\text{MAE} = \frac{1}{|K|} \sum_{(i,j) \in K} |r_{ij} - \hat{r}_{ij}| \dots \quad K < N \text{ olmak üzere,} \quad (3.13)$$

K = Kullanıcı ve ürün çiftleri sayısını gösterir. Tahmin hatası için, kullanıcı-ürün çiftleri gerçek değeri ile kullanıcı-ürün tahmin değeri arasındaki farkı hesaplamak için belirlenen sayıdır.

3.13.2. Kök Ortalama Mutlak Hata (RMSE)

Kök Ortalama Hata (Root Mean Squared Error: (RMSE) değeri ne kadar düşük ise, yapılan öneri o kadar tutarlıdır. Ayrıca İşbirliğine dayalı filtreleme algoritması ile doğru başarı puanını tahminin çok düşük hata ile yapıldığının bir göstergesidir. RMSE, daha büyük hataları daha fazla küçülterek cezalandırır ve küçük tahmin hatalarında küçültmek daha az önemsizleştirir. Eğer tahmin değeri için çok küçük hataları önemsizmeyecek ise; RSME kullanımı daha uygundur (Ricci, vd., 2010).

$$\text{RMSE} = \sqrt{\frac{\sum_{(i,j) \in K} (r_{ij} - \hat{r}_{ij})^2}{|K|}} \quad (3.14)$$

3.14. Öneri Sistemleri Ve Karşılaşılan Problemler

İşbirliğine dayalı filtreleme, veri toplama olanakları büyük ölçüde “kalabalıkların bilgeliği” kavramını basitleştirilmiş şekliyle internet üzerinden sağlar. Öte yandan, kalabalığın bilgeliği kavramıyla örtüşmeyen durumlar söz konusu olabilir. Bazı insanların davranışlarını modelleme oldukça kolay iken, bazı kullanıcılar diğerleri gibi tipik davranışı göstermezler.

Bu tip kullanıcıların mevcudiyeti tavsiye sistemi düşük verimlilikte bir kaymaya ve sonuçlara neden olabilir. Rakipleri tarafından bir ürün için ve negatif değerlendirmeler veya yorumlar gönderilebilir. İyi öneri sistemi bu sorunlarla başa çıkmak zorundadır.

Öneri sistemlerinde doğasında bir başka sorun ölçeklenebilirlik ile ilgilidir. Geleneksel algoritmalar veri nispeten küçük miktarlarda iyi çalışır, ancak geleneksel algoritmalar ile aynı kalite seviyesinde sonuçları elde etmek büyük veri setleri için sorun olabilir. Geleneksel algoritmalarda offline (çevrim-dışı) işlem durumunda pek bir sorun olmayabilir, ama bir gerçek zamanlı Online durumlarda daha uzmanlaşmış yaklaşımlar gerektirir.

Son olarak, bazı sorunlar gizlilik ihtiyacını yaratır. İnsanların özeline ait bilgiyi tahmin etme sonucu bunları desenleyen öneri sistemi problemler doğurmuştur. Böyle bir durumun son örnekleri, hedefli reklam amacıyla tüketici tercihlerine göre dizin gebelik tahmini hesaplamak mümkün olan büyük bir şirket tarafından yaşanmıştır. Şirket tahmin sistemi tarafından anne adaylarına gelecekte alacakları ürünleri doğru tahmin etmiştir. Hedefli reklamlar sonrası ürün önerisi sunulan genç kızının babası tarafından kızının gebeliğini öğrenmek sürpriz olmuştur (Gürsaka, 2017).

İşbirliğine dayalı filtreleme sisteminin avantajlarının yanında kendi karşılaştığı bazı sorunlar vardır. Öneri sistemlerinin kalitesini etkileyen başlıca problemlerden bazıları veri kıtlığı (sparsity), ölçeklenebilirlik scalability, doğruluk (accuracy), gizlilik (privacy) ve şilin (shilling) ataklarıdır (Schafer vd., 2007; Su ve Khoshgoftaar, 2009).

3.14.1. Doğruluk problemi

Öneri sistemlerindeki en önemli problemlerin başında kaliteli ve yüksek doğrulukta öneri üretme problemi gelir. Kullanıcının bu önerilere göre hareket etmesi beklendiği için, önerilerin doğru olması önemlidir. Eğer sistem kullanıcıları yanlış yönlendirirse, sistemin güvenilirliği azalır (Herlocker, vd., 2004).

3.14.2. Ölçeklenebilirlik problemi

Öneri sistemlerinde İşbirliğine dayalı filtrelemeye yönelik bir diğer temel sorun, “Ölçeklenebilirlik” dir (Hernández-del-Olmo ve Gaudioso, 2008). Etkili bir şekilde filtreleme için çok sayıda kullanıcıdan elde edilen veriye ihtiyaç duyulmaktadır. Ayrıca önerilerini, her kullanıcı için, kullanıcıların belirttikleri öğelere sınırlandırdıklarında, büyük miktarda veri gerektirir. E-ticaret sitelerinde kullanıcı ve öğe sayısında dinamik bir artış olabilir. Öneri sisteminde, sistem çok büyük sayıda kullanıcı veya öğe, üründen oluştuğunda, geleneksel işbirliğine dayalı filtreleme algoritmaları ciddi ölçeklenebilirlik sorunları yaşayacaktır. Ölçeklenebilirlik sorununu çözmek için, model tabanlı öneri algoritmaları genellikle boyut indirgeme teknikleri, tekil değer ayrışımı (SVD), kümeleme, ya da madde (ürün, öğe) tabanlı algoritmalar olarak ortaya çıkmaktadır (Koren, vd., 2009).

3.14.3. Nadirlik (sparsity) problemi

Bu gibi sistemlerde aktif kullanıcılar sadece ürünlerin % 1'inden azını satın alacakları için, kullanılan matrisin hücrelerinin çoğu boştur ve bu problem nadirlik (sparsity) veya veri kıtlığı şeklinde adlandırılır (Adomavicius ve Tuzhilin, 2005). Ayrıca, Nadirlik nedeniyle önerilerin doğruluğu zayıf olabilir. Bu sorun, kullanıcı sayısı sınırlı olduğu zamanlarda karşımıza çıkmaktadır. Kullanıcı sayısının sınırlı olduğu zaman, öneri zayıf öneri ile sonuçlanır (Sarwar, vd., 2000). Çünkü her kullanıcı için komşular bulmak zordur, bu sorun genellikle en yakın komşu algoritmalarının yoğun kullanıldığı işbirliğine dayalı yöntemlerde daha çok karşımıza çıkar. Nadirlik veya yetersiz veri, İşbirliğine dayalı filtreleme algoritmalarının karşılaştığı önemli problemlerden birisidir (Su ve Khoshgoftaar, 2009). Bu problemin üstesinden gelmek için kullanılan yöntemlerden biri de yetersiz veriye

sahip sitelerin veri setlerini birleştirerek bütünleşmiş veri setleri üzerinden öneriler üretmektir (Huang, vd., 2004).

3.14.4. Soğuk başlangıç problemi

Soğuk başlangıç sorunu, işbirliğine dayalı filtreleme de öneri sistemleri için yaygın olarak karşılaşılan bir sorundur. Bir kullanıcı veya bir öge (ürün) için önce beğeni puanı elde edilemediği sürece, doğru ve güvenilir bir ürün tavsiye edilemez. Yeni bir ögenin (ürün) sisteme girmesi ya da sistemde var olan ancak hiçbir beğeni puanı bulunmayan öge (soğuk ürün) bulunması durumudur. Bu şekilde problemler ile karşılaşıldığında, öneri sürecinde bir benzerlik hesaplanamaz veya komşuları bilinemez (Chen, vd., 2009). Örneğin öneri sisteminde ilk kez yer alan bir ürün veya kullanıcı için daha önceden tespit edilmiş herhangi bir bilgisi yok ve oluşturulamıyor ise sistemin doğru öneri verme şansı çok azalır. Farklı durumlara dayanarak, soğuk başlangıç problemini “yeni sistem-soğuk-başlangıç problemi” ve “yeni-kullanıcı-soğuk-başlangıç-problem” olarak ikiye ayırabiliriz. Bu durumda, işbirliğine dayalı filtreleme sisteminde, önerebilecekleri herhangi bir öngörüselle bilgiye sahip değildir. Bu nedenle, kötü performans gösterirler (Middleton, vd., 2002). Yeni-kullanıcı-soğuk-başlangıç probleminde, öneri sisteminde yeni bir kullanıcının sınırlı da olsa hiçbir derecelendirme bilgisine sahip olmadığı durumdur. Bu sorunda ise, sisteme yeni bir kullanıcı girmesidir. Bu kullanıcıya ilişkin sistemde bir bilgi yoktur. Yeni katılan bir kullanıcının tercihleri konusunda sistemin bir bilgisi olmadığı için, öneri benzerliği için kurulacak komşu kullanıcıların da kimler olduğu hesaplanamaz. Bu nedenle yeni katılan kullanıcı kendisini sisteme tanıtmaya yetecek sayıda içeriği izlemeden veya oylamadan önce, sistemin tutarlı ve doyurucu önermelerde bulunması beklenmez. Bu yeni kullanıcı probleminde hedef kullanıcılara öneri yapamaz.

Bu problem, sınırlı sayıda kullanıcı profiline sahip sistemler için hala geçerlidir (Middleton, vd., 2002). Öneri Sistemde yeni bir ürün ortaya çıktığında, bilgi elde edilene kadar bir kullanıcıya tavsiye edilebilir. Genellikle bu durum “erken-rant problemi” olarak adlandırılır (Towle ve Clark, 2000). Bu sorun kullanılan İşbirliğine dayalı filtreleme algoritmalarına göre iki ana kategoride ele alınabilir.

3.14.5. Olağandışı kullanıcı problemi

İşbirliğine dayalı filtrelemenin doyurucu önerilerde bulunmakta zorlandığı kullanıcı gurubu, tercihleri kullanıcıların büyük bölümüyle örtüşmeyen kullanıcılardır. Bunlardan ilki, aynı zamanda Gri Koyun problemi olarak da bilinir (Su ve Khoshgoftaar, 2009). Bu sorun bazı kullanıcıların sürekli olumsuz ya da sürekli olumlu görüş belirtmesidir. Böyle kullanıcıların tercihleri diğer kullanıcılarla örtüşmediğinden, bir benzerlik ilişkisi kolay kolay kurulamaz. Dolayısıyla bu kullanıcı bilgileri, hem kendisine, hem de diğer kullanıcılara öneride bulunmak için öneri sisteminin yararlanacağı benzerlik kuracağı bir bilgi uzayı oluşturmaz. Bir diğer Olağandışı kullanıcı problemi ise; Siyah koyun (Black Sheep) olarak adlandırılır (Marques. ve Dario, 1994). Bu guruba kendine özgü beğenileri karşısında öneri yapılması neredeyse imkânsızdır. Bu kullanıcılar için doğru tahminler yapılamaz. Bu kullanıcı görüşlerinin, beğenilerinin ister kullanıcı benzerliğine dayalı önerilerde, ister ki ürün benzerliğine dayalı önerilerde öneri sistemleri için fayda sağlamazlar.

3.14.6. Sınırlı içerik analizi problemi

Bu sorun, İçerik tabanlı sistemler ile ilgilidir. İçerik tabanlı sistemlerde öneri, öğelerin özellikleri ile sınırlıdır. Bu nedenle, içerik tabanlı sistemlerde içerik bilgisinin öneri sunacak yeterlilikte olması gerekir (Parameswaran, vd., 2011). Bu sorunun aşılmasında içeriklerin ya bir bilgisayar tarafından otomatik olarak çözümlenebilir bir yapı da ya da ürün özelliklerinin elle sınıflandırılarak kodlanması ile yeterli içerik sağlanabilir.

3.14.7. Kapsam geçerliliği problemi

Öneri sistemleri, ister kullanıcı tabanlı ister madde tabanlı filtreleme yaklaşımları sonucunda oluşturulan öneri listesini kullanıcıya öneri olarak sunar. Ancak her iki yaklaşımda da popüler ürünler daha öncelikli olarak öneri listesine alınır. Böylece popüler ürünler daha çok önerilirken, sistemde var olan ancak kullanıcının belki de çok beğeneceği ürün/ürünler listeye girmemektedir (Sarwar, vd., 2002). Böylece popüler olamayan ürünler sistemde yetersiz çeşitlik veya kapsam geçerliliği problemi olarak karşımıza çıkar. Kapsama, sistemin bir tahmin üretebileceği maddeleri basit bir yüzdesidir.

Belirli durumlarda işbirliğine dayalı filtreleme sistemi, verinin bulunmaması veya verinin aşırı kısıtlayıcı eşikten/threshold dolayı bir tahmin üretememektedir. Bu nedenle değerlendiricilerin kendi sistemlerinin gerçekleştirdiği kapsama düzeyini rapor etmesi önemlidir. Çoğu durumda sistemlerin yüksek güvenilirlikte iyi bir tahmin üretemeyeceği durumları (ör; Bir maddeye ilişkin çok az miktarda veri olduğunda) tespit etmeleri mümkündür.

Sistemlerin bu türden maddelere ilişkin tahminler üretememesiyle doğruluk değerlerini şişirmesi olasıdır. Belirli sıralama-tabanlı görevlerde bu durum kabul edilebilir. Ancak, veri tabanından bulunan herhangi bir maddeye ilişkin bir tahmin talebi olduğunda, bir tahmin üretilmiyor olması genellikle pek de uygun değildir. Hangi durumda olursa olsun, kapsama rapor edilmeli ve sistem doğruluğu, her iki sistemin de tahminde bulunabildiği maddeler üzerinden kıyaslanmalıdır (Jannach, vd., 2012).

3.15. Yükseköğretimde Öneri Sistemlerine Dayalı Ders Seçme Modeli

Yükseköğretimde öğrencilerin doğru dersi seçmeleri, geleceği bu karara bağlı olduğu için çok önemlidir. Öğrenci, oldukça fazla zorunlu ve seçmeli ders arasından doğru dersi seçme kararında zorlanır. Yanlış derslerin seçilmesi, öğrenci yeteneği, yetenek ve kişisel ilgi arasındaki uyumsuzluk anlamına gelir. Bu zorluğun üstesinden gelmek için Fakülte yönetimi tarafından danışman öğretim elemanı görevlendirilir. Ancak söz konusu danışman öğretim elemanları, öğrenci yeteneğini ve ilgisini bilemediği gibi herhangi bir dersten daha başarılı olacağını da tahmin edemez. Bu nedenle, genel olarak öğrenciyi en uygun yöne doğru yönlendirebilen başka güvenilir kaynak bulunmadığından, doğru bir ders seçmede ona rehberlik etmesi için ders seçme öneri sistemi modeli geliştirilmiştir.

Geliştirilen öneri sistemi, öğrencilere “tavsiye” niteliğinde yol gösterecek olmasından dolayı öneri sistemine “yol gösterici (pathfinder)” ismi tez danışman hocam ile kararlaştırılarak verilmiştir. Yol gösterici olarak isimlendirdiğimiz öneri sistemi, öğrencilerin geçmiş ders puanlarından gelecekteki seçeceği bir ders bulmaya yardımcı olan ve istatistik tekniklerine göre tahminleme yapan bir sistemdir.

Bir istatistiksel tahmin modeli, verileri oluşturulan karmaşık olayların bir temsilidir. Genel bir tanım olarak, rassal değişkenler ve parametreler arasındaki ilişkileri tanımlayan matematiksel formülasyonlara sahiptir. Bu istatistiksel tahmin sürecinde rassal değişkenler ve bazen de parametreler hakkında varsayımlarda bulunur (Tabachnick ve Fidell, 2001; Johnson ve Wichern, 2007).

Genel bir tahmin formu: Veri = model + artıklar (hata terimi) olarak ifade edilir. Bu ifade de yer alan, modeli verilerdeki değişkenliğin çoğunu açıklamalıdır. Artıklar, uyumsuzluğun, model tarafından açıklanamayan verilerin bir kısmının temsilidir (Stewens, 1996).

Tahmin, verilere uygun modele dayanarak anakütle parametresini öğrenme yollarını veya öğrenme sürecini temsil etmektedir. Tahmin anakütlenin bir parametresi için önerilen tahmin edicinin değeri şeklindedir (Brooks ve Thompson, 2017).

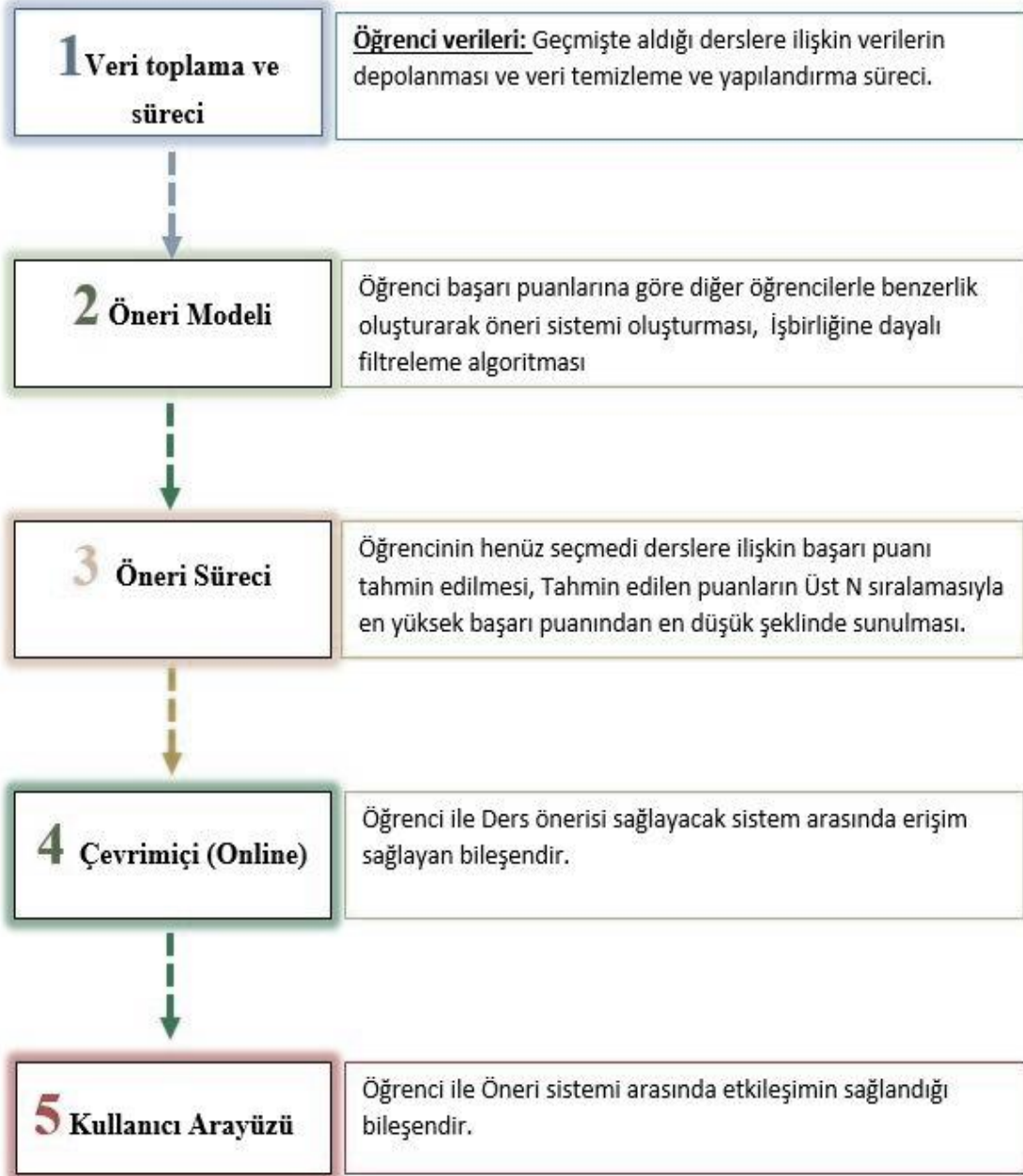
Örnek istatistikten anakütle parametresi hakkında bilgi edinmenin, nokta tahmini ve aralık tahmini ve hipotez testleri olmak üzere üç temel yolu vardır. Nokta tahmini: parametreyi tahmin eden örnekten hesaplanan tek değerlerdir. Güven Aralıkları: parametre için bir değer aralığı verir. Aralık tahminleri: parametrenin belirli bir güven derecesi ile gerçekleşmesi beklenen değer aralıklarıdır. Hipotez testleri: parametrenin belirli bir değeri için uygulanan testlerdir (Tabachnick ve Fidell, 2001).

Bu tezde, tasarlanan öneri modeli öğrencinin ders seçimi için gelecekteki zorunlu seçmesi gereken ya da seçmeli olarak ders seçiminde başarı puanı istatistik tahmin modeli önermektedir. Önerilen istatistik tahmin modeli, eksik gözlemlerin elde edilmesi probleminde, her bir öğrencinin her bir desisi için puan tahminidir.

3.16. Yükseköğretimde öneri sistemlerine dayalı ders seçme modeli bileşenleri

Öneri sistemleri temel 5 bileşen olduğu daha önceki bölümde detaylı olarak açıklanmıştır. Yükseköğretimde öneri sistemlerde aynı 5 temel bileşenden oluşur. Bu beş bileşen öğrenci verilerin toplanması, öneri modelinin geliştirilmesi, kişiselleştirilmiş öneri sunma süreci, daha hızlı işlem için çevrimiçi erişim ortamı ve öğrenci ile öneri sistemi

arasında etkileşimin sağlanacağı ara yüz tasarımı olarak sıralanmaktadır. Ders öneri sistemi temel bileşenleri şematik olarak Şekil (3,8)'de verilmiştir.



Şekil 3.8. Yükseköğretimde öneri sistemi temel bileşenleri

3.16.1. Ders öneri modeli veri toplama süreci

Yükseköğretimde bir üniversiteye ilk kaydolan her öğrenciye kayıta bir öğrenci ID tanımlanır. Bu ID, T.C kimlik numarası ve ya öğrenci numarası olabilir. Üniversiteye ilk kaydolan her öğrenciye ilişkin bazı bilgileri; demografik bilgiler (doğum tarihi, cinsiyet, medeni durum, yaşadığı yerleşim yeri, mezun olduğu lise türü vb.) merkezi yerleştirme sınavı sonucunda ÖSYM tarafından, kayıt yaptıracığı Üniversite veritabanına aktarılmaktadır. Bu bilgiler, Üniversitenin bilgi işlem merkezi birimi tarafından ilgili öğrenci işleri birimine ve fakülte ya da yüksekokul öğrenci bilgi sistemine aktarılır. Bunun haricinde öğrenci sınav, staj ve bilgiler fakülte öğrenci ileri tarafından takip edilir, girilir ve düzenlenir. Öğrencinin derslerine ait notlar ilgili öğretim elemanı tarafından not sistemine girilerek onaylanır. Öğrencilerden ilgili derse ilişkin değerlendirme anketleri uygulanır. Dolayısıyla, bir öğrenciye ait birçok kanaldan veri girişi yapılmaktadır. Artık öğrencinin bütün bilgileri öğrenci ID ile eşleştirilerek birleştirilir. Öneri sistemi uygulama aşamasında Öğr_ID ile eşlenik veriler kullanılır. Bu bilgiler, öğrencinin sınıfı, alacağı zorunlu dersler, almış olduğu derslere ilişkin puanları, sınav takvimi, ders ön koşulları vb. bilgileridir. Öneri sisteminde, başvurulacak yaklaşıma göre veri seti analiz yapılacak ortama aktarılır. Demografik tabanlı yaklaşımda öğrencinin demografik bilgileri veri olarak kullanılır. İçerik tabanlı bir filtreleme sisteminde ders içerikleri, öğrenci ilgi alanları, güçlü ve zayıf yönleri gibi sözel verilerden yararlanır. İşbirliğine dayalı filtreleme sistemi, öğrenci geçmiş ders puanlarıdır. Karma (Hibrit) filtreleme tekniğinde, İşbirliğine dayalı ve içerik tabanlı yaklaşımda kullanılan sözel veriler ile başarı puanları birlikte kullanılır. Hibrit filtreleme tekniğinde, içerik bilgilerinin kullanıldığı durumlarda pearson korelasyon yerine kosinüs(cosine) benzerlik ölçütünden yararlanır (Towle ve Clark, 2000).

3.16.2. Ders öneri modeli

Ders öneri modeli, öğrencinin seçebileceği derslere ait başarı puanlarının tahminlerinin yapıldığı bileşendir. Bu bileşende öğrenci geçmiş verilerinden yararlanarak, henüz seçmediği derslere ilişkin başarı notu tahmin yapılır. Tahmin yapılırken kullanılacak öneri sistemi burada belirlenir. Eldeki veri yapısına göre; İçerik tabanlı filtreleme, işbirliğine dayalı filtreleme veya hibrit filtreleme yaklaşımlarından biri veya birden fazla uygulanabilir.

İşbirliğine dayalı filtreleme tekniğinde, hafıza(memory) tabanlı yaklaşımda öğrenci-ders veri matrisindeki öğrenci başarı puanları ile dersler arasındaki benzerliğine dayalı olarak hesaplanır. Hafıza tabanlı yaklaşımda ders-ders benzerliğine dayalı olarak Pearson korelasyon değeri hesaplanır, ders çiftleri arasında en yüksek benzerlikler bulunur. Söz konusu derse ilişkin öğrenci başarı puanları ile hesaplanan korelasyon değerleri çarpılır. Elde edilen değer korelasyon değerleri toplamına bölünerek, tahmin edilecek ders notu hesaplanır. Hafıza tabanlı yaklaşımda bütün veri setini kullandığı için bu işlem süresi, çok uzun olduğundan çok rağbet görmemiştir (Sarwar, vd., 2002).

Model tabanlı yaklaşımlar ise, öğrenciler arasında benzerliklere göre kümeleme algoritmalarından yararlanır. Model tabanlı yaklaşım diğer bir ifade ile kümeleme algoritmaları kullanır. Makina öğrenme algoritmalarından birisi en çok uygulanan K en yakın komşuluk algoritması aynı zamanda, en iyi performans veren kümeleme algoritmalarından biridir (Jannach, vd., 2012). KNN, tüm veri kümesindeki en yakın birime göre birimleri sınıflandıran bir yöntemdir. Ders önerisi sistemi durumunda, aynı dersleri seçmiş ve benzer notları alan tüm öğrencileri tanımlamak ve sınıflandırmak için KNN kullanılır. Bu sınıflama için önce her öğrenci çifti için uzaklık hesaplanır, böylece 2 öğrenci arasındaki benzerlik veya uzaklık ölçülür. Aynı kümede yer alan öğrenciler için, uzaklık değerinden yararlanarak her öğrenciye bir ağırlık atanır. Her bir öğrenci için en yakın komşu öğrencilerin ders başarı puanları ve uzaklık değeri ile öğrencinin ağırlıklı ortalaması hesaplanır, bu ağırlık ortalaması değeri bir öğrenci için belirli bir derse ait tahmin edilen notu verir.

Ayrıca model tabanlı yaklaşımda, matris tamamlama teknikleri arasında yer alan matris ayrıştırma (matrix decomposition) tekniklerinde yararlanır. Öneri sistemleri uygulamalarında en çok uygulanan ve başarılı sonuç veren matris ayrıştırma tekniği ise, tekil değer ayrışımı (TDA) tekniğidir. Tekil değer ayrışımına göre veri matrisi, u ve v iki ortogonal matris ve orijinal matristen elde edilen öz değerlerin çarpımı ile işlem yapar. Orijinal veri matrisinden k boyutunda kesilmiş (truncated) matris ile $(U_k \times \sigma \times V_k)$ şeklinde kesilmiş matrislerin yardımıyla hesaplar. TDA, orijinal veri matrisi yerine daha küçük bir veri seti ile orijinal veri matrisini modellediği için işlem süresi olarak oldukça hızlıdır.

3.16.3. Ders Öneri Süreci

Ders öneri sürecinde, tahmin modelinde elde edilen ders başarı notu tahmin değerleri veri matrisine yazılır. Herhangi bir öğrenci ders seçiminde öneri olarak tahmin değerlerini sunmak amacıyla büyükten küçüğe sıralanır. Bir öğrencinin seçebileceği ders havuzunda yüzlerce ders varsa hepsini sıralayarak vermek anlamlı değildir. Bunun için Üst N $N=10,20$ gibi daha az sayıda bir ders önerisi yapılır. Eğer öğrenci, bu dersler dışında daha fazla ders için öneri isterse N değerini kendisi tanımlayarak veya tüm dersler için hesaplanan öneri listesini isteyebilir.

3.16.4. Çevrimiçi: Ders seçimi öneri yönetim sistemi

Öneri sistemlerinde kullanıcı ile sistem arasındaki erişimin sağlandığı bileşendir. Yükseköğretimde hâlihazırda öğrenciler derslerini çevrimiçi ortamlardan, bilgisayar, tablet, mobil araçlardan yararlanarak yapabilmektedirler. Öğrenci ders seçimi sonrası seçmiş olduğu dersleri danışman öğretim elemanına çıktısını alarak onaylatmaktadır. Danışman burada ders çakışmalarına veya ön koşul bilgisi varsa ona dikkat etmektedir.

Bu tezde, Yükseköğretimde ders seçiminde öneri sistemi modelinde, “yol gösterici” olarak tanımladığımız e-danışman (çevrimiçi danışman) mimarisine erişim sağladığı alandır. Öğrenci siteme Öğr-ID ile girdiğinde, sitem öğrenciye ders seçme süreciyle ilgili bütün bilgilere erişim sağlandığı alandır. Bu alanda kayıt yenileme, ders seçimi ve onayı, ders seçiminde öneri (Üst N), ders ücreti ödeme ve fakülte tarafından görevlendirilen danışman öğretim elemanı onayı ve önerisinin yer aldığı yerdir.

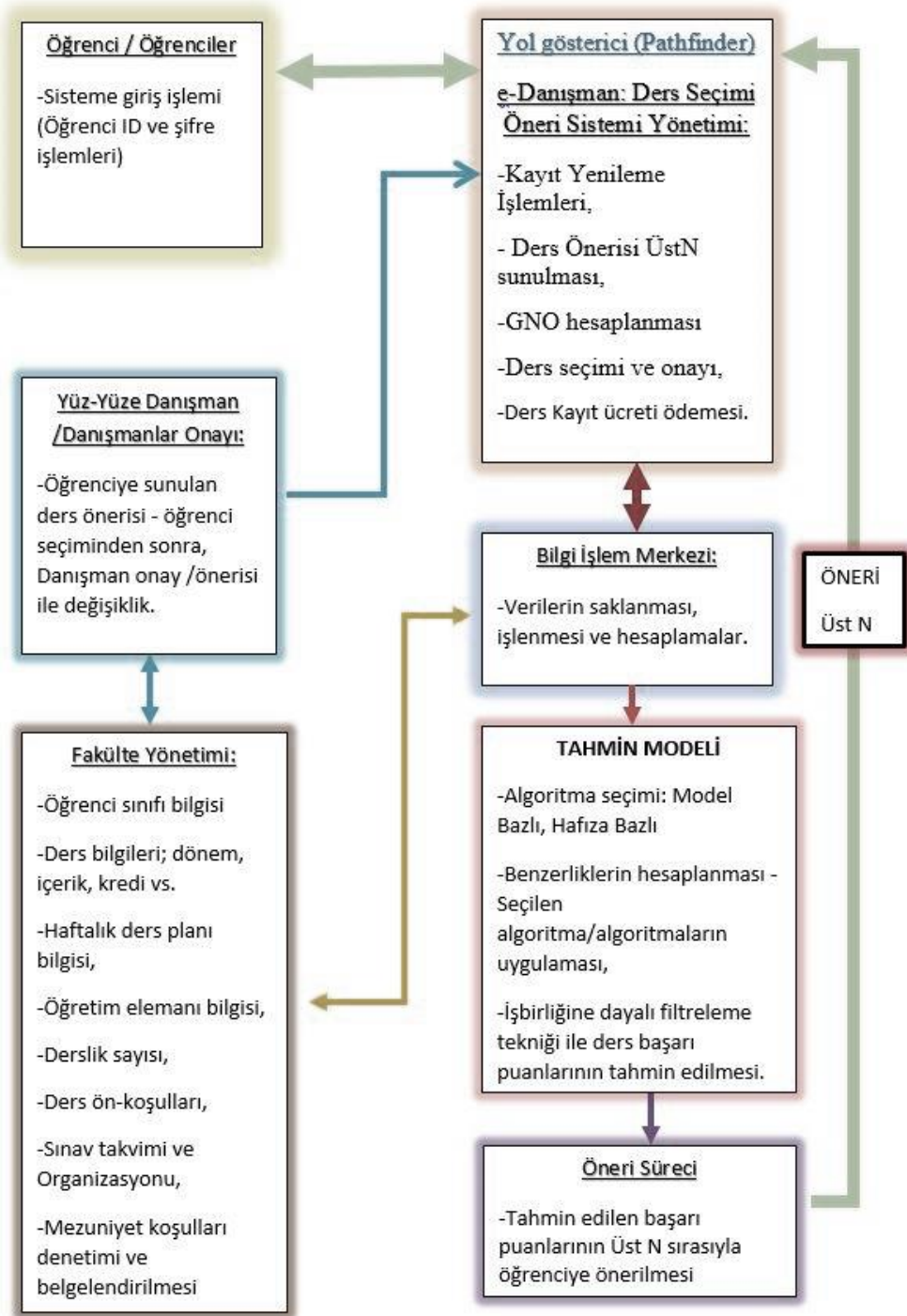
3.16.5. Ders seçimi öneri yönetim sistemi: Ara yüzü

Ders öneri sistemi ara yüzü; öğrenin sistemler etkileşim kurduğu alandır. Öğrencinin sisteme giriş yaptığı andan, çıkışına kadar işlem yaptığı web alanıdır. Bu alanın tasarımı önemlidir. Web sayfası tasarımı, kolay erişim ve etkileşim, destek, anket, forum, gibi alanlardan oluşur. Bu alan öğrencinin sisteme girdiğinde kendine göre tasarlayacağı esnek yapıda düşünülebilir. İlk tasarım sonrası, tasarıma ilişkin bir değerlendirme anketi ile öğrenci görüşleri alınarak, ara yüz tasarımı yenileme ve iyileştirmeleri sağlanabilir.

3.17. Yükseköğretimde işbirliğine dayalı ders seçme öneri modeli

Bir öneri sisteminde, öğrencilere uygun öneri yapmak için sınıflama ve puan tahmin etmede istatistiksel teknikleri, makine öğrenme ve veri madenciliği, karar ağaçları, Bayesian sınıflandırıcı teknikleri uygulanabilir. Bizim önerdiğimiz öneri sisteminde İşbirliğine dayalı filtreleme algoritmalarından yararlanılmıştır. Bu algoritmalar, K-Ortalamlar ve K en yakın komşu algoritmaları ile Tekil değer ayrışımı algoritmalarıdır. Kümeleme algoritmasında veri içerisindeki yapıları ve ilişkileri bulmak için kullanılmıştır. Kümeleme tekniğinde öğrencileri daha önce sahip oldukları başarı puanları değişkenlerine göre, benzerlik hesaplayarak gruplandırır. Benzerlik ölçüsü olarak pearson korelasyon hesaplanmıştır. Korelasyon iki değişken arasındaki ilişkiyi gösteren bir ölçüdür ve -1 ile +1 arasında yer almaktadır (Tatlidil, 2002). Benzerlik hesaplamada her bir çift öğrenci arasındaki benzerlik pearson korelasyon katsayıları hesaplanarak belirlenmiştir. Kümeleme analizinin öneri sistemlerinde, uygulanmasının ardındaki neden aynı grupta yer öğrencilerin benzer başarı elde edeceği varsayımına dayanmaktadır. Kümeleme analizde hesaplanan benzerlik ölçütü pearson korelasyon katsayısı mutlaka nedensellik belirtmez. Örneğin, sonbahar da havaların soğuması ile yaprakların sararıp dökülmesi arasında, sıcaklık ile dökülen yaprak sayısı arasında korelasyon pozitif ve yüksek olabilir. Ancak bu korelasyon yaprakların sararıp dökülmesinin nedeni değildir. Yaprakların sararıp dökülmesindeki temel neden güneş ışınlarının sonbaharda daha yatay bir şekilde gelmesidir. Bu nedenle geliştirdiğimiz öneri sisteminde, veri setinde mevcut araştırılan öğrencilerin rassal değişkenleri olan ders başarı puanları arasındaki korelasyon değerleri her zaman nedenselliği açıklamamaktadır. Bu amaçla rassal değişken olarak değerlendirilen, başarı puanları arasındaki nedenselliği keşfedici ve gizil yapıyı belirlemek amacıyla matris ayrıştırma tekniği “tekil değer ayrışımı” yaklaşımı uygulanmıştır. Ayrıca, Tekil değer ayrışımı tekniğine başvurulmasındaki temel nedenlerinden birisi, veri matrisinde %10’unda başarı puanlarından oluşan ve yaklaşık %90’nın boş olup tahmin edilecektir. TDA’ya başvurulmasının ikinci temel neden ise, orijinal veri setindeki rassal değişkenleri arasında ilişkili yapıyı daha küçük boyutta ilişkisiz değişkenlerle tahmin edecek olmasıdır. Boyut küçültme işlem süresini kısalttığı için öneri sistemlerinde oldukça başvurulan bir tekniktir. TDA’ya başvurulmasının üçüncü temel neden ise, rassal değişkenler arasındaki gizil yapının belirlenmesi ve karşılaştırılan orijinal değerler ile doğruluk derecesi yüksek, tahmin hata değeri daha küçük olması hedeflendiği için başvurulmuştur.

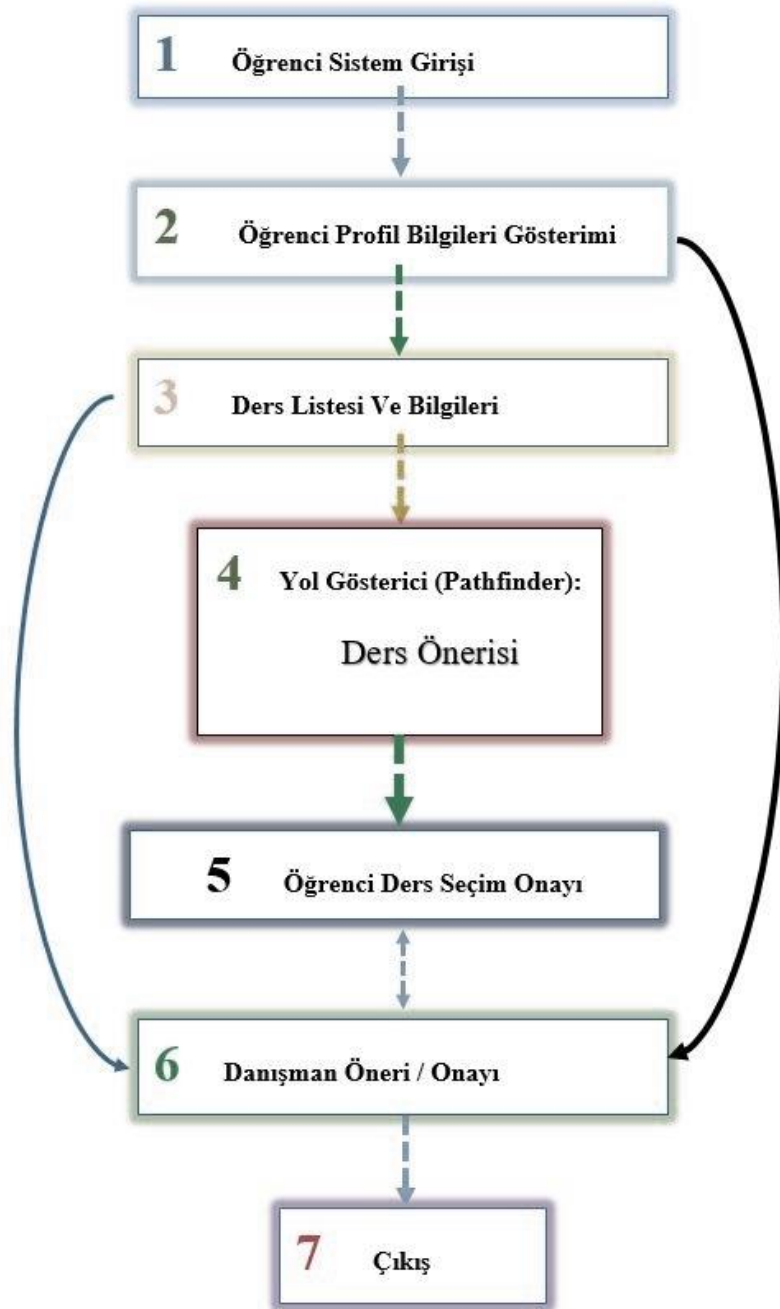
Öğrenci siteme girişini, ders öneri sistemi (yol gösterici) ara yüzü ile yapmaktadır. Sisteme girdikten sonra ders seçimi işlemini seçtiğinde, öneri modeli tarafından hesaplanan tahmin başarı puanları üstN olarak öğrenciye sunulacaktır. Öğrenci sunulan derslerden birinin üstüne geldiğinde, derse ait öğretim elemanı bilgisi, tahmini başarı notu bilgisi, kaçınıcı dönemde yer aldığı gibi bilgileri görmesi sağlanmalıdır.



Şekil 3.9.. Yükseköğretimde İşbirliğine Dayalı Ders Seçme Öneri Modeli

Bu öneri sisteminde üç büyük elemanın yer aldığı teknolojisi ilişkilendirilmiştir. İlk insan-bilgisayar ara yüzü teknolojisi gerekli bilgileri toplamaktır. İkincisi, bu amaçla, bilgi alma teknikleri ile toplanan verilerden istatistiksel tahmin ve makine öğrenme ile ulaşılan bilgileri önerilere dönüştürmektir. Üçüncüsü öneri sisteminde gerekli bilgileri toplamaya, veri tabanı üzerinde işlemek, hesaplamak ve dağıtım ağ ile ilgili bir teknolojidir. Şekil 3.9'de Yükseköğretimde işbirliğine dayalı ders seçme modelinde söz konusu üç elemanın ilişkileri şematik olarak gösterilmiştir. Şekil 3.9. incelendiğinde ilk olarak öğrenci ile öneri sistemi etkileşimim olduğu, “yol gösterici” e-danışman sistem etkileşimini göstermektedir. Burada, Bilgi işlem merkezinde kayıtlı veri tabanı ile Fakülte yönetimi tarafından girilen bilgilerin yer aldığı bilgi toplama ve toplanan verilerin düzenlendiği bölümdür. İkincisi toplanan verilerden öneri modeli veya tahmin modelinin uygulandığı, eldeki verilere algoritmaların uygulandığı teknolojidir.

Üçüncü olarak, öneri süreci ve veri ve bilgi akışının sürekliliğini böylece uygulanan algoritmalar sonucunda öğrenci tekrar sisteme ders seçimi için giriş yaptığında yeni tahminlerin üretileceği teknolojidir.



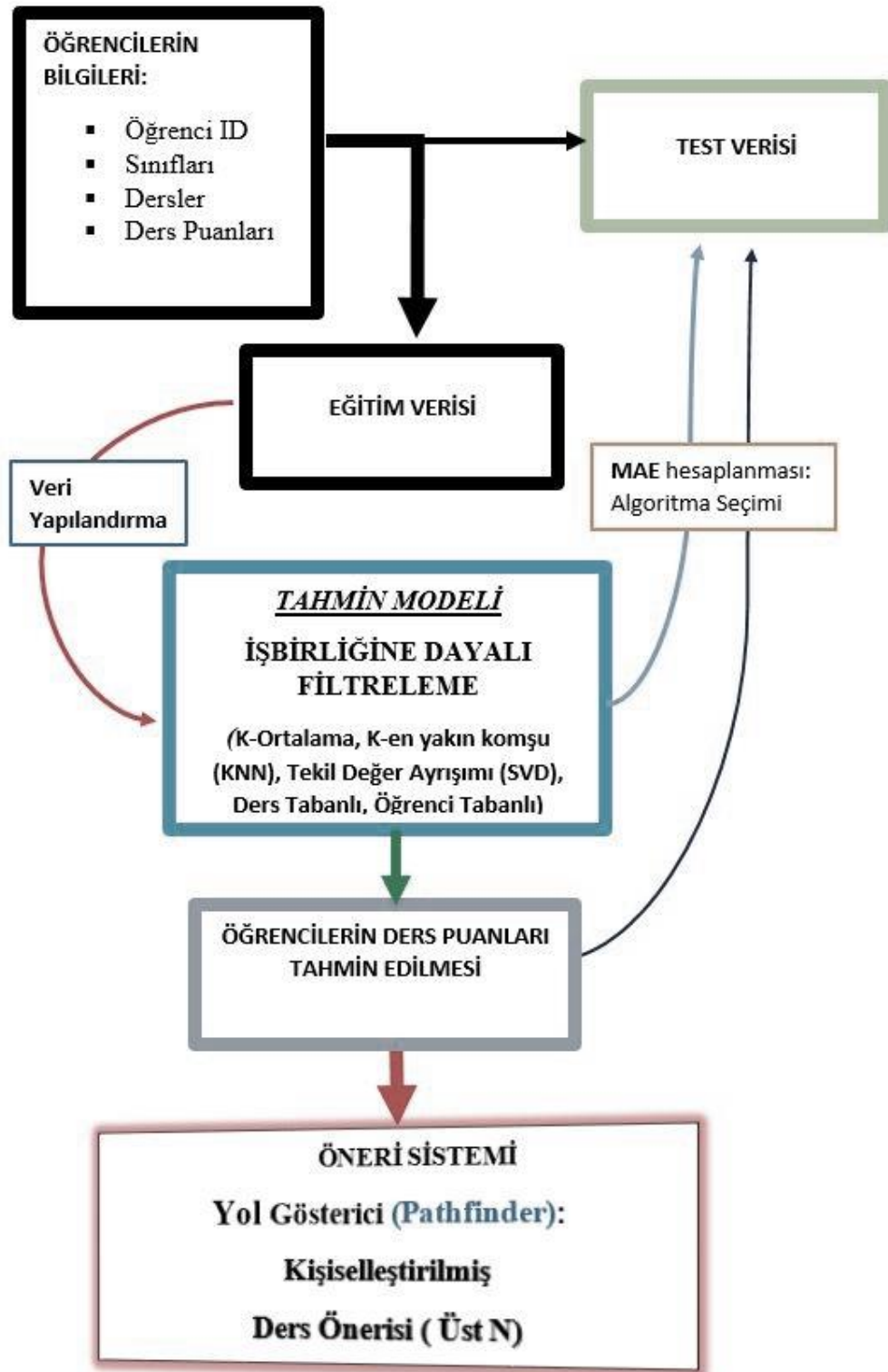
Şekil 3.10. Yükseköğretimde Öneri Sistemi İle Öğrenci Ders seçim Süreci

Üçüncü bölüm aynı zamanda öğrenci ve ara yüze bilgi sağlayan, önerilerin (tahminlerin) sunulduğu insan-bilgisayar ara yüzünün oluştuğu teknolojidir. Üçüncü bölümde öğrenci sistemle etkileşim sürecinin şematik gösterimi Şekil 3.10.' da verilmiştir.

3.18. İşbirliğine dayalı filtreleme ile ders başarı puanlarını tahmin süreci

Tezde önerilen ders seçme modelinde, işbirliğine dayalı filtreleme tekniğinden yararlanılmıştır. Bu modelde, veri tabanında öğrenci ID, sınıfı ve dönemi, dersleri ve ders başarı puanları yer almaktadır. Bu teknikte öğrencilerin geçmiş dönem ders başarı puanları rassal değişkeni olarak ele alınacaktır. Hafıza tabanlı bu iki öğrenci – öğrenci veya ders-ders benzerlik, geçmiş dönem başarı puanları yardımıyla pearson korelasyon değerleri hesaplanacaktır. Benzerliklere göre, öğrencilerin geçmiş dönem başarı puanları da tekrar tahmin edilmektedir. Böylece, bir test verisi olarak bütün öğrencilerden geçmiş dönem 20 ders başarı puanları sistemde belirlenerek, bu derslerin gerçek değerleri ile tahmin değerleri arasındaki MAE değerleri hesaplanacaktır.

Ayrıca MAE değeri, K-ortalama, K en yakın komşu ve tekil değer ayrışımı algoritmaları içinde benzer şekilde hesaplanmaktadır. Hesaplanan MAE değerine göre en iyi tahmin yapan algoritma ve ilgili algoritmadaki k, komşuluk sayısı belirlenir. Tekil değer ayrışımında ise, en küçük MAE değerinin elde edildiği ve stabil olduğu k (truncated) değeri baz alınır (Huang, vd., 2004). Tekil değer ayrışımında en küçük k değerinde elde edilen tahmin değerleri, ilgili öğrencinin karşılık gelen derslerine ait başarı puan değerleridir. Uygulanan algoritmalar sonucunda tahmin edilen başarı puanları ders isimleri ya da ders kodları (ders_ID) ile eşleştirilir. Daha sonra dersi seçecek öğrenciye genellikle büyükten küçüğe doğru sıralanmış şekilde en başarılı olacağından, en düşük notu alacak ders sıralı listesi oluşturulur. Oluşturulan liste çok uzun olması kullanışlı olmadığından ÜstN sıralı liste Üst N=10 ve 20 kadar desin sıralı biçimde öğrenciye öneri olarak sunulur (Parameswaran, vd., 2011).



Şekil 3.11. İşbirliğine dayalı filtreleme ile ders başarı puanlarını tahmin süreci

İşbirliğine dayalı filtreleme tekniği ile modellenen ders öneri sistemi Şekil 3.11.'de şematik olarak verilmiştir. Öneri sisteminde önerdiğimiz öneri modeli, kişiselleştirilmiş ders önerisi sunan yol göstericidir.

4. BULGULAR VE TARTIŞMA

Yükseköğretimde öğrenciler seçmeli dersleri seçerken, öğrenciye ve danışmana yol göstermesi açısından İşbirliğine dayalı filtreleme algoritmaları yardımıyla İktisat bölümü ve İşletme bölümü öğrencileri için aynı algoritmalar ayrı ayrı bölüm bazında uygulanarak henüz seçilmemiş derslerin başarı puanları tahmin edilmeye çalışılmıştır. Bu tezde öneri algoritmaları işlemlerinde Matlab Simulink 2017 Anadolu Üniversitesi lisanslı yazılımından yararlanılmıştır. Burada İşbirliğine dayalı filtreleme tekniği model algoritmalarından, k en yakın komşu algoritması (KNN), K-ortalama (k-means) algoritması ve gizil yapının keşfinde ve boyut indirgemedede en iyi performans veren tekil değer ayrışımı (SVD) algoritmasından yararlanılmıştır. Ayrıca İşbirliğine dayalı filtreleme tekniği hafıza tabanlı olan ders benzerliğine dayalı algoritmadan yararlanılmıştır. Model ve hafıza tabanlı algoritma sonuçlar, değerlendirme ölçütü olarak MAE hesaplamıştır. Uygulama sonucunda alan yazında en çok kullanılan ve öğrenci geçmiş başarı puanlarının 0-4 aralığında sürekli veri olmasından ötürü en iyi performans veren tahmin ölçüm metriği MAE değeri kısıt olarak alınmıştır. Ayrıca MAE, en sık kullanılan ve doğrudan yorumlamak için en kolay olanıdır. MAE için istatistiksel anlamlılık testi yapmak ve güven aralıklarını hesaplamak konusunda geniş bir araştırma literatürü vardır (Sarwar, vd., 2002). MAE ölçütünü seçme nedeni olarak, tahmin değerlerimizin gerçek değerlere en yakın olanını istemizden kaynaklıdır.

Önerilen ders öneri yaklaşımı için öneri kalitesini ölçmede, en belirgin ve yaygın olarak kullanılan ortalama mutlak hatası (MAE) metriği kullanılmıştır (Good, vd.1999). Ortalama mutlak hata (MAE), tahmin edilen bir derecelendirme ile kullanıcının gerçek derecelendirmesi arasındaki ortalama mutlak sapmayı ölçer (Nikos, vd.2012). Ortalama mutlak hata (MAE), ilk olarak N tane derecelendirme-tahmin çiftlerinin karşılıklı mutlak hatalarının toplanması ve daha sonra ortalamasının hesaplanmasıyla hesaplanır. Düşük MAE değerleri daha yüksek doğruluk gösterir. MAE en çok kullanılan metriktir çünkü hesaplamasının mekaniği basit ve anlaşılması kolaydır.

K en yakın komşuluk algoritması için komşu sayısı boyutu 10, 25, 50, 100, 250 olarak ve tüm kullanıcılar için sırasıyla kullanıcı tabanlı MAE değeri hesaplanmış ve sonuçlar karşılaştırılmıştır. Ayrıca Öğrenci – öğrenci benzerliği temelindeki yaklaşımda K

ortalamalar algoritması için komşu sayısı boyutu 2, 4, 8,12, 16 olarak ve tüm kullanıcılar için sırasıyla kullanıcı tabanlı MAE değeri hesaplanmış ve sonuçlar karşılaştırılmıştır.

Diğer taraftan üçüncü olarak Tekil değer ayrışımı algoritması için k değeri 2, 4, 8, 12, 16 olarak ve tüm kullanıcılar için sırasıyla kullanıcı tabanlı MAE değeri hesaplanmış ve sonuçlar karşılaştırılmıştır. Ders benzerliğine dayalı item-based tabanlı algoritma uygulamasında, henüz alınamamış ve notu olmayan dersler için sütunlara ortalama değerleri atanarak benzerlik hesaplanmıştır. Bu benzerlik hesaplanması sonrası k-ortalamalar kümeleme algoritması ile ders başarı tahmini yapılmıştır. Ders benzerliğine dayalı algoritmada komşu sayısı boyutu 5, 25, 50, 10, 250 olarak alınarak tüm dersler için MAE değeri hesaplanmış karşılaştırılarak yorumlanmıştır. Anadolu Üniversitesi İİBF'sine bağlı bölümler arasından en çok öğrencisi olan İktisat ve İşletme bölümü öğrencileri için seçmeli ders öneri sistemi uygulanmıştır. Sırasıyla İktisat bölümü öğrencileri için seçmeli ders öneri sisteminde kullanılan algoritma sonuçları ve İşletme bölümü öğrencileri için kullanılan algoritma sonuçları aşağıda başlıklar altında verilmiştir.

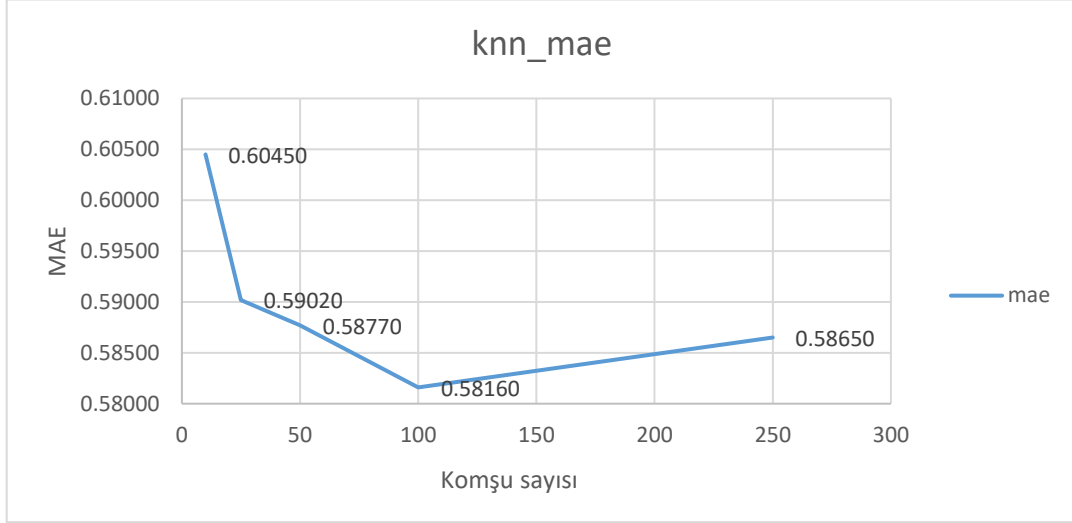
4.1. İktisat Bölümü Öğrencileri İçin Seçmeli Ders Öneri Algoritma Sonuçları

Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesine bağlı İktisat bölümü 2009-2014 eğitim dönemlerindeki öğrenci ders başarı puanlarından yararlanılarak, K-NN (K En Yakın Komşu) Algoritması, K Ortalama algoritması, Tekil değer ayrışımı (SVD) ile ders benzerliğine bağlı Ürün bazlı (Item base) algoritmaları sonuçları izleyen kısımlarda başlıklar halinde verilmiştir.

4.1.1. K-NN (K En Yakın Komşu) Algoritması Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde, İktisat bölümü öğrencileri için Öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda KNN algoritması sonucunda MAE ölçütü 0,58160 ile 0,6045 arasında hesaplanmıştır. En düşük MAE değeri 0,5865 ile komşu sayısının 100 olarak belirlenmesinde elde edilmiştir. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük

0,5865 olması yüksek bir hata ile tahmin yapıldığını göstermektedir. Komşu sayısına göre hesaplanan MAE değerleri Şekil 4.1’de verilmiştir.



Şekil 4.1. İktisat bölümü, K-En Yakın Komşuluk Algoritması sonucu MAE değerleri

Şekil 4.1 İncelendiğinde K En yakın komşuluk algoritması için en düşük MAE komşu sayısı yüz olduğunda bulunmuş ve daha sonra tekrar yükselme eğilimine girmiştir. Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmini komşu sayısı $k=100$ olduğunda elde edilmiştir.

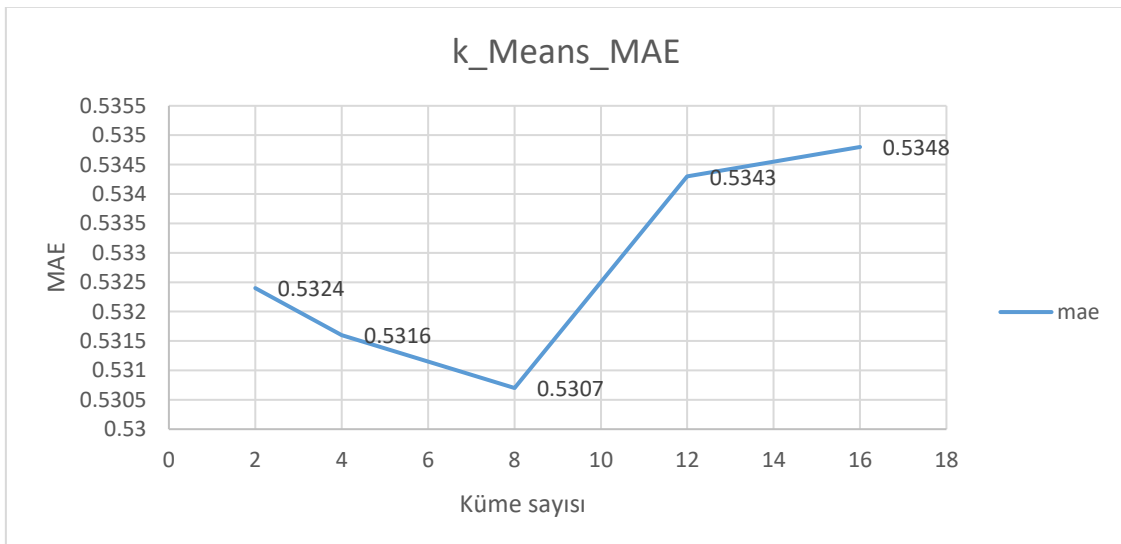
Öğr_ID	ARY111	HUK151	HUK152	İKT128	İKT216	Öğr_ID	ARY111	HUK151	HUK152	İKT128	İKT216
1	2.3	0	3.3	0	2.3	1.0	2.21	2.22	2.16	2.39	2.08
2	3.3	3.3	2.7	1	1	2.0	2.10	2.19	1.94	2.21	1.72
3	0	1.7	1.7	2.3	2.7	3.0	1.51	1.74	1.61	1.80	1.48
4	0	0	0	0	0	4.0	2.96	2.83	2.90	3.13	2.95
5	4	4	3	3.3	0	5.0	3.67	3.73	3.64	4.00	3.50
6	3	3	2.7	3.3	1.7	6.0	2.64	2.68	2.55	2.76	2.20
7	0	2.3	1.7	2	2.3	7.0	2.07	2.13	2.08	2.36	1.96
8	2.7	0	0	0	3	8.0	2.64	2.44	2.30	2.56	2.51
9	0	2	1.7	3.3	2	9.0	1.96	2.07	1.98	2.05	1.64
10	4	2.3	2	3.7	2	10.0	2.69	2.76	2.69	2.85	2.15
11	2.7	0	0	2.7	2.7	11.0	2.57	2.59	2.51	2.63	2.16
12	0	3	2.7	3	2.3	12.0	2.88	2.97	2.97	2.93	2.75
13	0	0	1.3	0	0	13.0	0.80	1.04	1.02	1.00	0.74
14	0	0	1.3	0	0	14.0	1.55	1.33	1.47	1.78	1.80
15	1.3	2.3	1.7	1.7	1	15.0	1.33	1.43	1.39	1.58	1.02
16	1.7	2.3	1.7	1.7	2	16.0	1.77	2.05	1.97	2.19	1.77
17	3.3	3.3	2.7	2.7	3	17.0	2.62	2.56	2.50	2.78	2.10
18	0	3.3	3.7	0	1.3	18.0	2.37	2.36	2.28	2.48	1.94
19	3.3	3	3	2.7	3	19.0	2.84	2.72	2.65	2.96	2.19
20	1.7	2	1.7	1.7	0	20.0	1.70	1.70	1.61	1.79	1.26

Şekil 4.2. İktisat bölümü k en yakın komşu algoritması sonucunda tahmin edilen ders puanları

.K en yakın komşuluk algoritması uygulaması sonucunda İktisat bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları dersler Şekil 4.2’de verilmiştir. Bu yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları göstermektedir. Şekil 4.2. incelendiğinde öğrencinin ders notu olmayan hücreler sıfır değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler sarı renkle belirtilirken şeklin sağ tarafında ise tahmin değerleri turuncu ile belirtilmiştir

4.1.2. K-Means (K Ortalamalar) Algoritması Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde veri setinde, İktisat bölümü öğrencileri için Öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda k-ortalama algoritması sonucunda MAE ölçütü 0,5307 ile 0,5348 arasında hesaplanmıştır. En düşük MAE değeri 0,5307 ile komşu sayısının 8 olarak belirlenmesinde elde edilmiştir. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük 0,5307 olması yüksek bir hata ile tahmin yapıldığını göstermektedir. Komşu sayısına göre hesaplanan MAE değerleri Şekil 4.3’de verilmiştir.



Şekil 4.3. İktisat bölümü, k-ortalama algoritması sonucu mae değerleri

Şekil 4.3 İncelendiğinde K Ortalama algoritması için en düşük MAE komşu sayısı $k=8$ olduğunda bulunmuş ve daha sonra tekrar yükselme eğilimine girmiştir. Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmini komşu sayısı $k=8$ olduğunda elde edilmiştir.

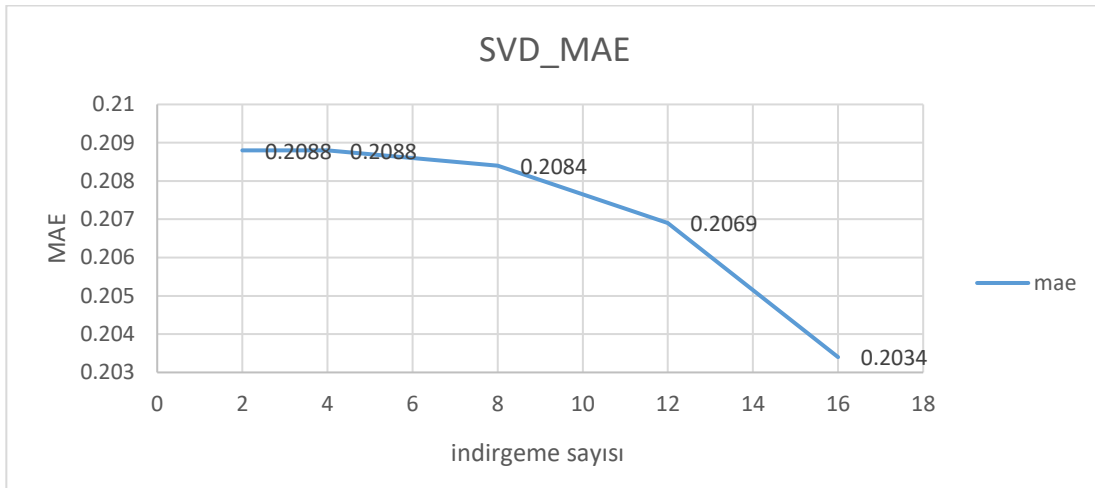
Öğr_ID	ARY111	HUK151	HUK152	İKT128	İKT216	Öğrenci_ID	ARY111	HUK151	HUK152	İKT128	İKT216
1	2.3	0	3.3	0	2.3	1	2.26	2.15	2.46	2.49	2.29
2	3.3	3.3	2.7	1	1	2	2.60	2.42	2.02	2.22	1.69
3	0	1.7	1.7	2.3	2.7	3	2.45	1.92	1.93	1.86	1.90
4	0	0	0	0	0	4	2.97	2.71	3.52	3.71	2.82
5	4	4	3	3.3	0	5	3.66	3.80	3.76	3.81	3.66
6	3	3	2.7	3.3	1.7	6	2.88	2.72	2.92	3.18	2.79
7	0	2.3	1.7	2	2.3	7	2.97	2.48	2.38	2.37	2.34
8	2.7	0	0	0	3	8	2.95	2.51	2.28	2.46	3.08
9	0	2	1.7	3.3	2	9	1.35	1.91	1.88	1.68	1.38
10	4	2.3	2	3.7	2	10	2.84	2.58	2.37	2.87	2.25
11	2.7	0	0	2.7	2.7	11	2.94	2.69	2.49	2.65	2.95
12	0	3	2.7	3	2.3	12	3.56	3.18	3.12	3.12	3.05
13	0	0	1.3	0	0	13	1.00	1.00	1.00	1.00	1.00
14	0	0	1.3	0	0	14	1.51	1.51	1.28	1.51	2.13
15	1.3	2.3	1.7	1.7	1	15	1.53	1.49	1.43	1.53	1.15
16	1.7	2.3	1.7	1.7	2	16	1.82	2.09	1.95	2.38	1.86
17	3.3	3.3	2.7	2.7	3	17	2.93	2.80	2.91	3.23	2.86
18	0	3.3	3.7	0	1.3	18	2.12	2.53	2.58	2.44	2.03
19	3.3	3	3	2.7	3	19	2.84	2.86	2.68	2.95	2.53
20	1.7	2	1.7	1.7	0	20	1.64	1.54	1.56	1.72	1.36

Şekil 4.4. İktisat Bölümü K-ortalama algoritma tahmin değeri

K -ortalama algoritması uygulaması sonucunda İktisat bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları dersler Şekil 4.4’de verilmiştir. Bu yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları göstermektedir. Şekil 4.4. incelendiğinde öğrencinin ders notu olmayan hücreler sıfır değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler sarı renkle belirtilirken şeklin sağ tarafında ise tahmin değerleri turuncu ile belirtilmiştir.

4.1.3. Tekil Değer Ayrışımı (SVD) Algoritması Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde, İktisat bölümü öğrencileri için Öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda Tekil değer algoritması sonucunda MAE ölçütü 0,288 ile 0,2034 arasında hesaplanmıştır. En düşük MAE değeri 0,2034 ile komşu sayısının 8 olarak belirlenmesinde elde edilmiştir. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük 0,2034 olması düşük bir hata ile tahmin yapıldığını göstermektedir. Bu duruma öğrenci başarı notu tahmininde en iyi performans tekil değer ayrımı (SVD) ile sağlanmıştır. K sayısına göre hesaplanan MAE değerleri Şekil 4.5.'de verilmiştir.



Şekil 4.5. İktisat bölümü, tekil değer ayrışımı algoritması sonucu mae değerleri

Şekil 4.5 incelendiğinde Tekil değer ayrışımı algoritması için en düşük MAE k değeri 16 olduğunda bulunmuştur. Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmin k sayısı 16 olduğunda elde edilmiştir.

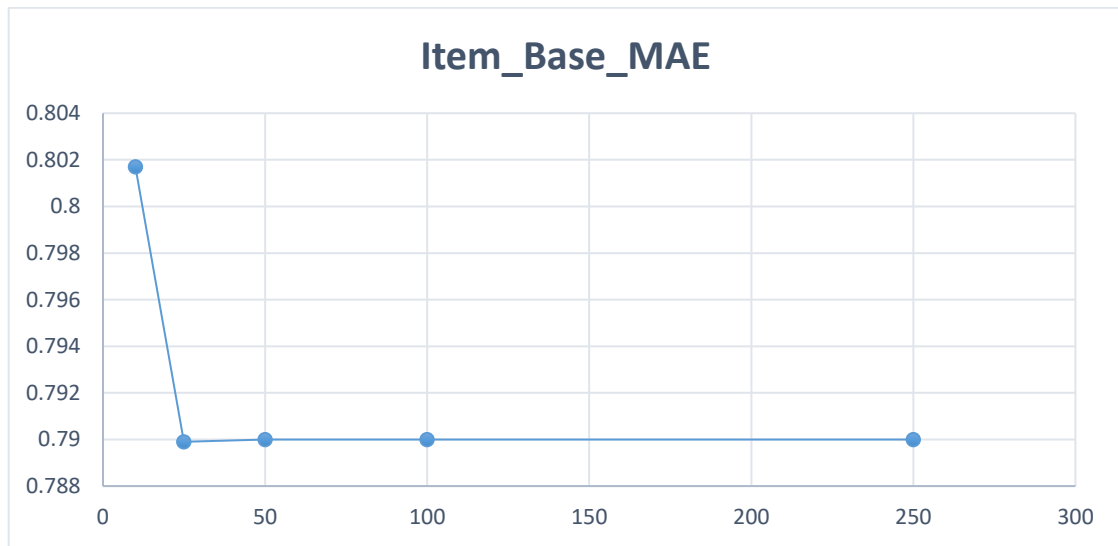
Öğr_ID	ARY111	HUK151	HUK152	İKT128	İKT216	Öğrenci_ID	ARY111	HUK151	HUK152	İKT128	İKT216
1	2.3	0	3.3	0	2.3	1	2.474	2.472	2.471	2.471	2.471
2	3.3	3.3	2.7	1	1	2	2.347	2.344	2.343	2.348	2.344
3	0	1.7	1.7	2.3	2.7	3	2.131	2.131	2.133	2.131	2.131
4	0	0	0	0	0	4	2.596	2.595	2.595	2.597	2.595
5	4	4	3	3.3	0	5	3.075	3.073	3.072	3.076	3.072
6	3	3	2.7	3.3	1.7	6	2.814	2.812	2.810	2.816	2.811
7	0	2.3	1.7	2	2.3	7	2.553	2.552	2.551	2.554	2.551
8	2.7	0	0	0	3	8	2.705	2.703	2.701	2.706	2.702
9	0	2	1.7	3.3	2	9	2.538	2.536	2.535	2.539	2.536
10	4	2.3	2	3.7	2	10	2.868	2.865	2.863	2.869	2.864
11	2.7	0	0	2.7	2.7	11	2.849	2.847	2.846	2.850	2.847
12	0	3	2.7	3	2.3	12	3.057	3.055	3.054	3.058	3.055
13	0	0	1.3	0	0	13	2.662	2.660	2.659	2.663	2.660
14	0	0	1.3	0	0	14	2.698	2.696	2.695	2.699	2.696
15	1.3	2.3	1.7	1.7	1	15	1.914	1.912	1.911	1.915	1.912
16	1.7	2.3	1.7	1.7	2	16	2.324	2.322	2.320	2.325	2.321
17	3.3	3.3	2.7	2.7	3	17	2.847	2.844	2.843	2.848	2.844
18	0	3.3	3.7	0	1.3	18	2.804	2.802	2.802	2.804	2.802
19	3.3	3	3	2.7	3	19	3.081	3.079	3.077	3.083	3.078
20	1.7	2	1.7	1.7	0	20	2.086	2.084	2.083	2.087	2.084

Şekil 4.6. İktisat bölümü ders notu tekil değer ayrışımı algoritması ($k=16$) tahmin değeri

Başarı notu tahmininde en başarılı sonuç tekil değer ayrışımı sonucunda elde edilmiştir. İktisat bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları dersler Şekil 4.6'de verilmiştir. Bu yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları göstermektedir. Şekil 4.6. incelendiğinde öğrencinin ders notu olmayan hücreler sıfır değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler sarı renkle ve şeklin sağ tarafında ise tahmin değerleri sarı renkle belirtilmiştir.

4.1.4. Ders Benzerliğine Dayalı Algoritma Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde veri setinde, İktisat bölümü öğrencileri için öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda ders benzerliğine dayalı ürün benzerlik (item-base) algoritması sonucunda MAE ölçütü 0,79 ile 0,8017 arasında hesaplanmıştır. En düşük MAE değeri 0,79 ile komşu sayısının 50 ve üzerinde MAE değeri değişmemiş sabit kalmıştır. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük 0,79 olması büyük bir hata ile tahmin yapıldığını göstermektedir. Bu duruma öğrenci başarı notu tahmininde kötü performans ders benzerliğine dayalı item-based algoritması sonucunda elde edilmiştir. Komşu sayısına göre hesaplanan MAE değerleri Şekil 4.7’de verilmiştir.



Şekil 4.7. İktisat bölümü, ders benzerliği algoritması sonucu mae değerleri

Şekil 4.7 İncelendiğinde Ders benzerliğine dayalı algoritması için en düşük MAE komşu sayısı 25 olduğunda bulunmuş ve bundan sonra 0,79 olarak değişmemiştir. Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmin komşu sayısı 25 olduğunda elde edilmiştir. Başarı notu tahmininde en başarısız MAE sonucu ders benzerliği sonucunda elde edilmiştir.

Ders benzerliği algoritması uygulaması sonucunda İktisat bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları dersler Şekil 4.8’de verilmiştir. Bu

yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları göstermektedir. Şekil 4.8. incelendiğinde öğrencinin ders notu olmayan hücreler sıfır değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler yeşil renkle ve şeklin sağ tarafında ise tahmin değerleri yeşille belirtilmiştir.

Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424	Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424
1	0	0	0	0	0	1.0	2.47	2.47	2.47	2.47	2.51
2	0	0	0	0	0	2.0	2.34	2.34	2.34	2.34	2.39
3	0	0	0	0	0	3.0	2.13	2.13	2.13	2.13	2.15
4	0	0	0	1.7	0	4.0	2.59	2.59	2.59	2.59	2.61
5	0	0	0	0	0	5.0	3.07	3.07	3.07	3.07	3.10
6	2	0	0	0	0	6.0	2.81	2.81	2.81	2.81	2.85
7	0	2.7	0	1	1.3	7.0	2.55	2.55	2.55	2.55	2.57
8	0	3	0	1.7	0	8.0	2.70	2.70	2.70	2.70	2.72
9	2.3	2.3	1	1.3	0	9.0	2.53	2.53	2.53	2.53	2.57
10	0	3	3	3	0	10.0	2.86	2.86	2.86	2.86	2.91
11	0	0	0	1	0	11.0	2.84	2.84	2.84	2.84	2.88
12	0	0	0	0	0	12.0	3.05	3.05	3.05	3.05	3.07
13	0	4	0	2	3	13.0	2.66	2.66	2.66	2.66	2.70
14	0	0	0	0	0	14.0	2.69	2.69	2.69	2.69	2.72
15	2.7	0	0	3.3	2.7	15.0	1.91	1.91	1.91	1.91	1.95
16	2	2	0	2.3	2	16.0	2.32	2.32	2.32	2.32	2.37
17	0	2	3	3	0	17.0	2.84	2.84	2.84	2.84	2.89
18	0	3	0	3	0	18.0	2.80	2.80	2.80	2.80	2.82
19	0	0	0	0	0	19.0	3.07	3.07	3.07	3.07	3.12
20	0	0	0	0	0	20.0	2.08	2.08	2.08	2.08	2.12

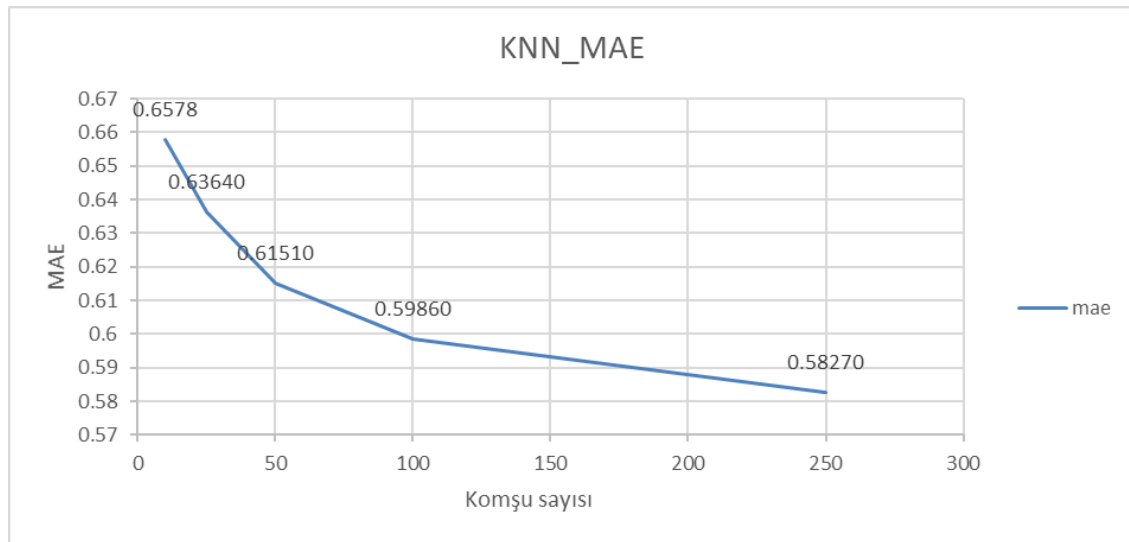
Şekil 4.8. İktisat bölümü ders benzerliği ders notu tahmin değeri

4.2. İşletme Bölümü Öğrencileri İçin Seçmeli Ders Öneri Algoritma Sonuçları

Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesine bağlı İşletme bölümü 2009-2014 eğitim dönemlerindeki öğrenci ders başarı puanlarından yararlanılarak, K-NN (K En Yakın Komşu) Algoritması, K Ortalama algoritması, Tekil değer ayrışımı (SVD) ile ders benzerliğine bağlı Ürün bazlı (İtem base) algoritmaları sonuçları aşağıda verilmiştir.

4.2.1. K-NN (K En Yakın Komşu) Algoritması Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde, İşletme bölümü öğrencileri için Öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda KNN algoritması sonucunda MAE ölçütü 0,5827 ile 0,6578 arasında hesaplanmıştır. En düşük MAE değeri 0,5827 ile komşu sayısının 250 olarak belirlenmesinde elde edilmiştir. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük 0,5827 olması yüksek bir hata ile tahmin yapıldığını göstermektedir. Komşu sayısına göre hesaplanan MAE değerleri Şekil 4.9'de verilmiştir.



Şekil 4.9. İşletme bölümü, k-en yakın komşuluk algoritması sonucu mae değerleri

Şekil 4.9. İncelendiğinde K En yakın komşuluk algoritması için en düşük MAE komşu sayısı yüz olduğunda bulunmuş ve daha sonra tekrar yükselme eğilimine girmiştir.

Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmini komşu sayısı $k=250$ olduğunda elde edilmiştir.

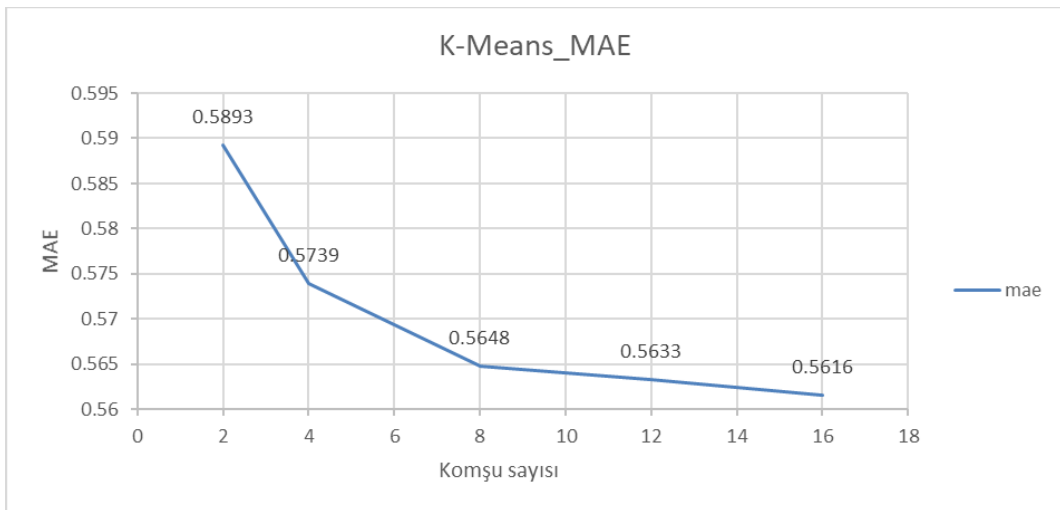
K en yakın komşu algoritması uygulaması sonucunda İşletme bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları dersler Şekil 4.10 'de verilmiştir. Bu yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları göstermektedir. Şekil 4.10.. incelendiğinde öğrencinin ders notu olmayan hücreler sıfır değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler yeşil renkle ve şeklin sağ tarafında ise tahmin değerleri yeşille belirtilmiştir.

Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424	Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424
1	0	0	0	0	0	1	1.51	1.51	1.51	1.51	1.51
2	0	0	0	0	0	2	2.27	2.04	1.91	1.95	1.91
3	0	0	0	0	0	3	1.51	1.43	1.51	1.28	1.28
4	0	0	0	1.7	0	4	3.30	3.28	3.13	2.93	3.00
5	0	0	0	0	0	5	1.30	1.30	1.30	1.30	1.30
6	2	0	0	0	0	6	2.03	2.03	2.03	2.03	2.03
7	0	2.7	0	1	1.3	7	2.22	2.05	2.24	2.19	2.07
8	0	3	0	1.7	0	8	3.25	3.36	3.02	3.40	3.19
9	2.3	2.3	1	1.3	0	9	3.29	3.09	2.93	2.85	3.09
10	0	3	3	3	0	10	3.26	3.34	3.19	3.28	3.13
11	0	0	0	1	0	11	2.32	2.33	2.24	2.34	2.39
12	0	0	0	0	0	12	1.28	2.07	2.79	2.67	1.69
13	0	4	0	2	3	13	3.72	3.41	3.33	3.58	2.98
14	0	0	0	0	0	14	2.54	2.54	2.45	2.55	2.45
15	2.7	0	0	3.3	2.7	15	3.01	2.87	2.92	3.57	2.67
16	2	2	0	2.3	2	16	3.28	3.43	3.79	3.31	3.36
17	0	2	3	3	0	17	2.69	2.69	2.68	2.69	2.68
18	0	3	0	3	0	18	3.76	3.65	3.59	3.87	3.65
19	0	0	0	0	0	19	2.33	2.33	2.30	2.34	2.34
20	0	0	0	0	0	20	2.13	2.13	2.13	2.13	2.10

Şekil 4.10. İşletme bölümü, k-en yakın komşuluk algoritması ders notu tahmin değerleri

4.2.2. K Ortalamalar Algoritması Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde veri setinde, İşletme bölümü öğrencileri için Öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda k-ortalama algoritması sonucunda MAE ölçütü 0,5616 ile 0,5893 arasında hesaplanmıştır. En düşük MAE değeri 0,5616 ile komşu sayısının 16 olarak belirlenmesinde elde edilmiştir. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük 0,5616 olması yüksek bir hata ile tahmin yapıldığını göstermektedir. Komşu sayısına göre hesaplanan MAE değerleri Şekil 4.11’de verilmiştir.



Şekil 4.11. İşletme bölümü, k-ortalama algoritması sonucu mae değerleri

Şekil 4.11. İncelendiğinde K Ortalama algoritması için en düşük MAE komşu sayısı k=16 olduğunda bulunmuş ve daha sonra tekrar yükselme eğilimine girmiştir. Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmini komşu sayısı k=16 olduğunda elde edilmiştir.

K -ortalama algoritması uygulaması sonucunda İşletme bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları dersler Şekil 4.12’de verilmiştir. Bu yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları göstermektedir. Şekil 4.12. incelendiğinde öğrencinin ders notu olmayan hücreler sıfır

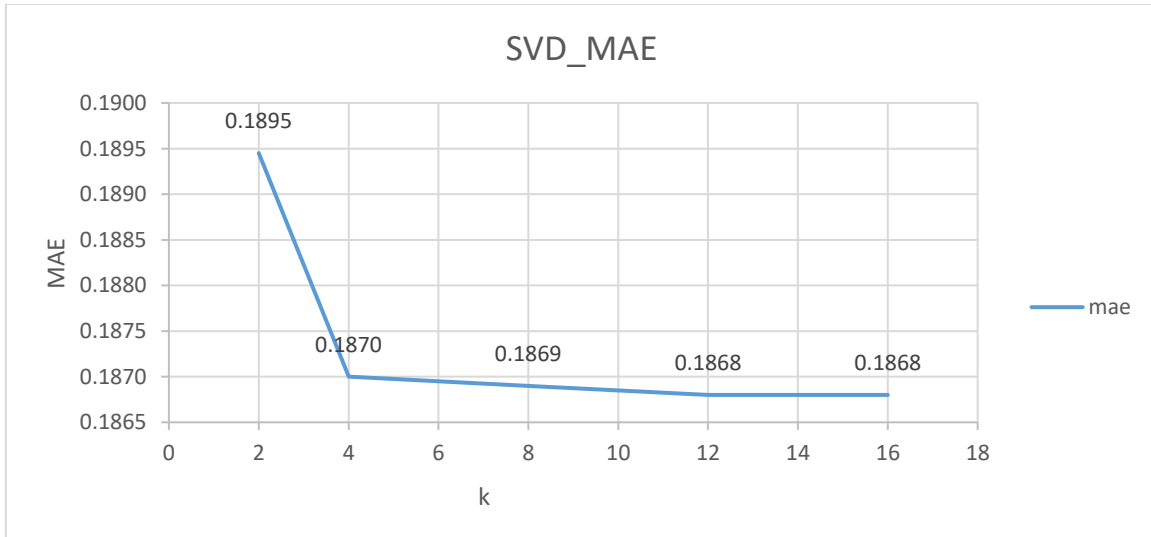
değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler yeşil renkle ve şeklin sağ tarafında ise tahmin değerleri yeşil renkle belirtilmiştir.

Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424	Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424
1	0	0	0	0	0	1	3.84	3.67	3.27	3.49	3.22
2	0	0	0	0	0	2	3.29	3.26	3.09	3.42	3.61
3	0	0	0	0	0	3	2.23	2.22	2.47	2.53	2.06
4	0	0	0	1.7	0	4	3.41	3.41	3.40	3.39	3.41
5	0	0	0	0	0	5	4.00	4.00	4.00	3.97	3.89
6	2	0	0	0	0	6	2.96	3.59	3.48	3.74	2.92
7	0	2.7	0	1	1.3	7	2.73	2.48	2.77	2.94	2.51
8	0	3	0	1.7	0	8	4.00	4.00	3.31	4.00	3.81
9	2.3	2.3	1	1.3	0	9	2.54	2.63	2.89	2.89	2.52
10	0	3	3	3	0	10	3.79	4.00	4.00	4.00	3.61
11	0	0	0	1	0	11	3.10	3.27	3.28	3.28	3.36
12	0	0	0	0	0	12	3.36	3.58	3.69	3.73	3.27
13	0	4	0	2	3	13	2.04	2.13	2.22	1.99	2.12
14	0	0	0	0	0	14	1.51	1.51	1.54	1.51	1.51
15	2.7	0	0	3.3	2.7	15	2.27	2.41	2.23	2.70	2.32
16	2	2	0	2.3	2	16	2.81	2.69	2.62	3.10	2.39
17	0	2	3	3	0	17	3.48	3.71	4.00	4.00	3.64
18	0	3	0	3	0	18	2.86	2.57	2.53	3.43	2.84
19	0	0	0	0	0	19	3.98	4.00	4.00	3.93	4.00
20	0	0	0	0	0	20	2.00	2.41	2.53	2.72	2.25

Şekil 4.12. İşletme bölümü K-ortalama algoritma sonucu tahmin ders notu

4.2.3. Tekil Değer Ayrışımı (SVD) Algoritması Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde, İşletme bölümü öğrencileri için Öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda Tekil değer algoritması sonucunda MAE ölçütü 0,1868 ile 0,1895 arasında hesaplanmıştır. En düşük MAE değeri 0,1895 ile $k=12$ ve $k=16$ olarak belirlenmesinde elde edilmiştir. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük 0,1868 olması düşük bir hata ile tahmin yapıldığını göstermektedir. Bu duruma öğrenci başarı notu tahmininde en iyi performans tekil değer ayrışımı (SVD) ile sağlanmıştır. K sayısına göre hesaplanan MAE değerleri Şekil 4,13’de verilmiştir.



Şekil 4.13. İşletme bölümü, tekil değer ayrışımı algoritması sonucu mae değerleri

Şekil 4.13 incelendiğinde Tekil değer ayrışımı algoritması için en düşük MAE k değeri 16 olduğunda bulunmuştur. Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmin k sayısı 12 ve 16 olduğunda elde edilmiştir. $K=12$ olarak alınabilir. Daha sonra MAE değeri stabil bir değere dönüşmüştür. Başarı notu tahmininde en başarılı sonuç tekil değer ayrışımı sonucunda elde edilmiştir. İşletme bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları derslere ilişkin şekil 4.14’de verilmiştir. Bu yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları ise Şekil 4.5’de verilmiştir. Şekil 4.14 incelendiğinde öğrencinin ders notu olmayan hücreler

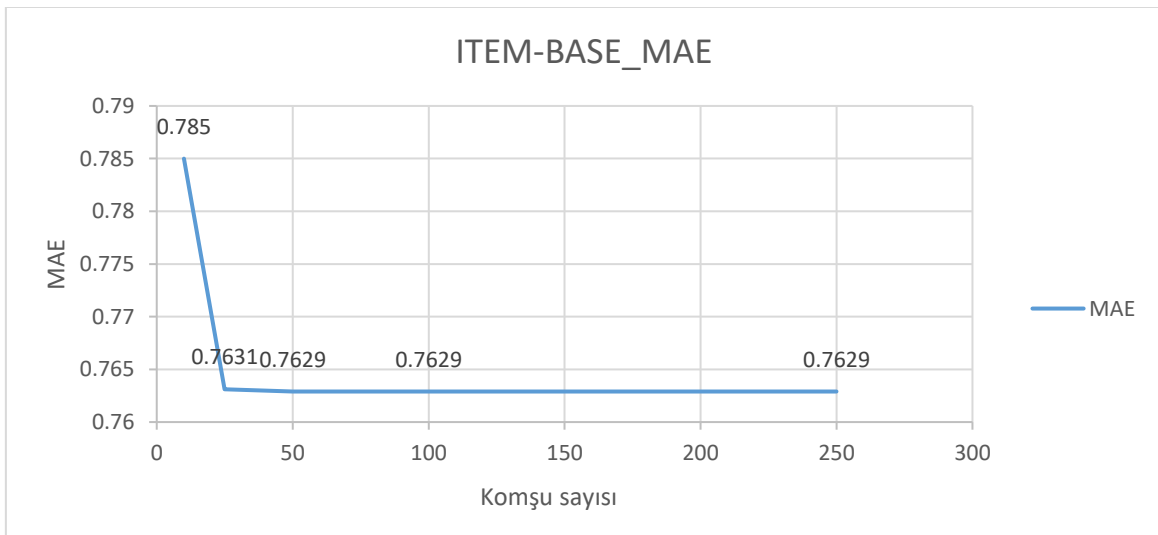
sıfır değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler yeşil renkle ve şeklin sağ tarafında ise tahmin değerleri yeşil renkle belirtilmiştir.

Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424	Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424
1	0	0	0	0	0	1.0	2.47	2.47	2.47	2.47	2.49
2	0	0	0	0	0	2.0	2.34	2.34	2.34	2.34	2.39
3	0	0	0	0	0	3.0	2.13	2.13	2.13	2.13	2.15
4	0	0	0	1.7	0	4.0	2.59	2.59	2.59	2.59	2.61
5	0	0	0	0	0	5.0	3.07	3.07	3.07	3.07	3.10
6	2	0	0	0	0	6.0	2.81	2.81	2.81	2.81	2.85
7	0	2.7	0	1	1.3	7.0	2.55	2.55	2.55	2.55	2.57
8	0	3	0	1.7	0	8.0	2.70	2.70	2.70	2.70	2.74
9	2.3	2.3	1	1.3	0	9.0	2.53	2.53	2.53	2.53	2.56
10	0	3	3	3	0	10.0	2.86	2.86	2.86	2.86	2.91
11	0	0	0	1	0	11.0	2.84	2.84	2.84	2.84	2.88
12	0	0	0	0	0	12.0	3.05	3.05	3.05	3.05	3.07
13	0	4	0	2	3	13.0	2.66	2.66	2.66	2.66	2.68
14	0	0	0	0	0	14.0	2.69	2.69	2.69	2.69	2.72
15	2.7	0	0	3.3	2.7	15.0	1.91	1.91	1.91	1.91	1.95
16	2	2	0	2.3	2	16.0	2.32	2.32	2.32	2.32	2.35
17	0	2	3	3	0	17.0	2.84	2.84	2.84	2.84	2.87
18	0	3	0	3	0	18.0	2.80	2.80	2.80	2.80	2.82
19	0	0	0	0	0	19.0	3.07	3.07	3.07	3.07	3.12
20	0	0	0	0	0	20.0	2.08	2.08	2.08	2.08	2.12

Şekil 4.14. İşletme bölümü tekil değer ayrışımı algoritması (k=12) tahmin ders puanları

4.2.4. Ders Benzerliğine Dayalı Algoritma Sonuçları

Üniversite öğrencilerinin seçmeli ders seçiminde veri setinde, İşletme bölümü öğrencileri için öneri sistemi mimarisi olan İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda ders benzerliğine dayalı ürün benzerlik (item-base) algoritması sonucunda MAE ölçütü 0,785 ile 0,7629 arasında hesaplanmıştır. En düşük MAE değeri 0,7629 ile $k=50$ ve üzerinde MAE değeri değişmemiş sabit kalmıştır. Bu değer öğrencinin herhangi bir dersten gerçek değeri ile tahmin değeri arasındaki farkların ortalaması olarak bulunur. MAE değerinin en düşük 0,7629 olması büyük bir hata ile tahmin yapıldığını göstermektedir. Bu duruma öğrenci başarı notu tahmininde kötü performans ders benzerliğine dayalı item-based algoritması sonucunda elde edilmiştir. Komşu sayısına göre hesaplanan MAE değerleri Şekil 4.15’de verilmiştir.



Şekil 4.15. İşletme bölümü ders benzerliği algoritması sonucu MAE değerleri

Şekil 4.15 incelendiğinde Ders benzerliğine dayalı algoritması için en düşük MAE $k=50$ olduğunda bulunmuş ve bundan sonra 0,7629 olarak değişmemiştir. Bu nedenle burada denenen algoritma sonucunda en uygun başarı notu tahmin $k=50$ olduğunda elde edilmiştir. Başarı notu tahmininde en başarısız MAE sonucu ders benzerliği sonucunda elde edilmiştir.

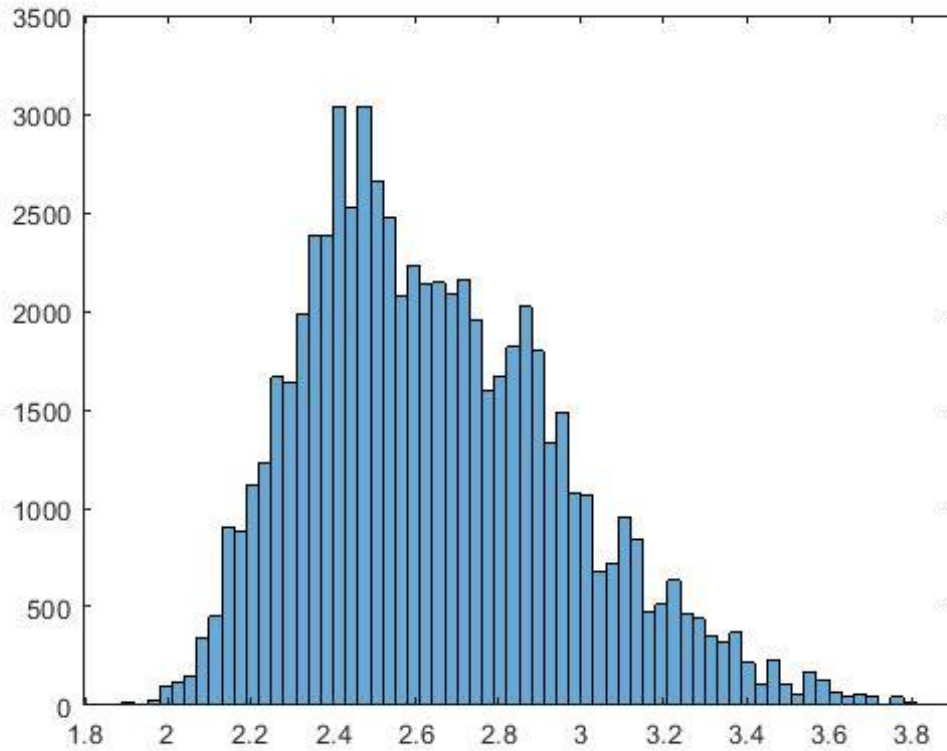
K -ortalama algoritması uygulaması sonucunda İşletme bölümünden örnek olarak ilk 20 öğrencinin geçmiş notları ve henüz almadıkları dersler Şekil 4.16’de verilmiştir. Bu

yirmi öğrencinin tekil değer ayrışımı algoritması sonucunda tahmin edilen notları göstermektedir. Şekil 4.16. incelendiğinde öğrencinin ders notu olmayan hücreler sıfır değeri ile doldurularak normalleştirilmiştir. Ayrıca, tahmin edilecek başarı notu, sıfır olan hücreler yeşil renkle ve şeklin sağ tarafında ise tahmin değerleri yeşil renkle belirtilmiştir.

Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424	Öğr_ID	BEÖ155	İŞL402	İŞL404	İŞL405	İŞL424
1	0	0	0	0	0	1	2.66	2.76	2.67	2.50	3.08
2	0	0	0	0	0	2	2.40	1.80	2.27	2.19	2.62
3	0	0	0	0	0	3	2.62	2.23	2.46	2.25	2.66
4	0	0	0	1.7	0	4	1.87	2.01	1.83	1.94	1.65
5	0	0	0	0	0	5	2.54	2.78	2.89	2.79	2.64
6	2	0	0	0	0	6	2.69	2.67	3.03	2.30	2.51
7	0	2.7	0	1	1.3	7	2.58	3.06	2.98	2.81	2.97
8	0	3	0	1.7	0	8	0.80	0.80	0.80	0.97	1.04
9	2.3	2.3	1	1.3	0	9	1.77	1.23	1.76	1.85	1.90
10	0	3	3	3	0	10	1.30	1.03	1.56	1.42	1.89
11	0	0	0	1	0	11	1.64	2.12	1.66	1.74	2.17
12	0	0	0	0	0	12	2.33	2.72	2.31	2.74	2.81
13	0	4	0	2	3	13	2.84	2.38	2.26	2.67	2.66
14	0	0	0	0	0	14	3.09	2.89	2.75	2.83	3.14
15	2.7	0	0	3.3	2.7	15	1.50	1.56	1.51	1.62	1.76
16	2	2	0	2.3	2	16	1.85	1.78	1.88	1.83	1.95
17	0	2	3	3	0	17	2.46	2.28	2.36	2.48	2.31
18	0	3	0	3	0	18	3.03	2.94	2.85	3.14	3.72
19	0	0	0	0	0	19	1.73	2.42	2.18	1.96	2.68
20	0	0	0	0	0	20	2.46	2.29	2.23	2.49	2.61

Şekil 4.16. İşletme bölümü ders benzerliğine algoritması sonucunda tahmin ders puanları

Tezde; Yükseköğretimde öneriş sistemlerine dayalı des seçme modeli olarak, işbirliğine dayalı filtreleme teknikleri hafıza ve model bazlı algoritmalar denenmiştir. Bu algoritmaların bulguları aşamalı halinde açıklanmıştır. Hafıza bazlı algortima ders benzerliği sonucu tahmin değerlendirme ölçütü MAE oldukça yüksek çıkmıştır. Model bazlı algoritmalar, K-ortalama, K en yakın Komşuluk ve Tekil değer ayrışımı algoritması uygulanmıştır.



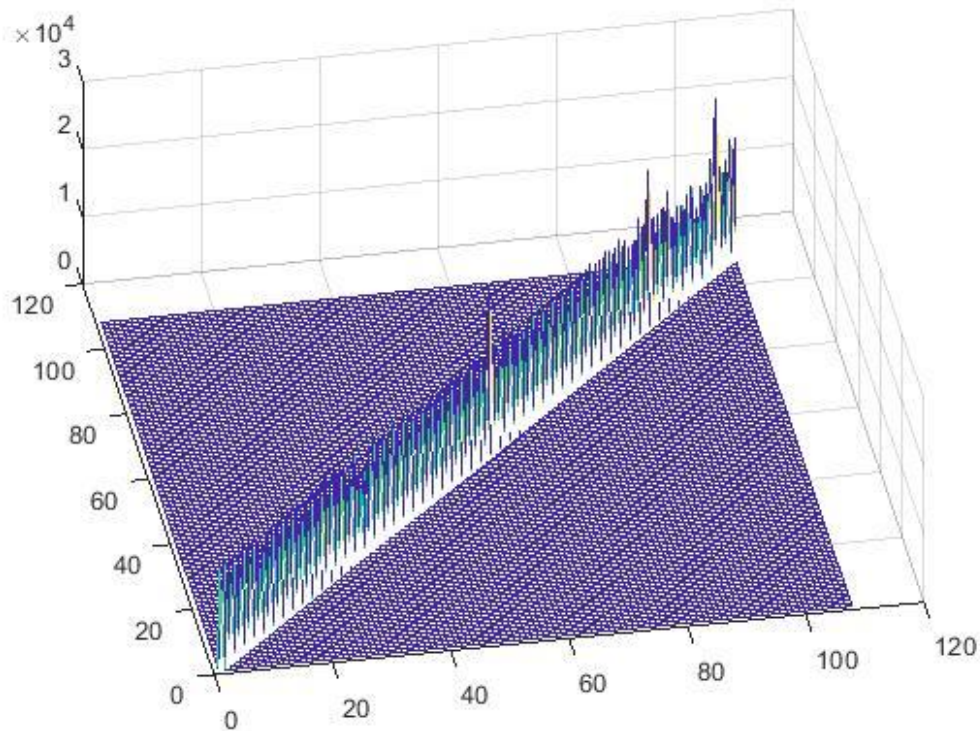
Şekil 4.17. İktisat bölümü tekil değer ayrışımı test verisi tahmin ders başarı puanları grafiği.

Model bazlı algoritmalar içinde en düşük MAE değeri Tekil değer ayrışımı algoritması sonucunda elde edilmiştir. İktisat bölümü veri matrisinden, test verisi $n = 108$ ders seçilmiş ve seçilen derslere ilişkin elde edilen tahmin ders başarı puanları histogram grafiği Şekil 4.17.' de verilmiştir.

	1	2	3	4	5	6	7	8	9	10
1	1.5178e+04	6.9840	6.7037	6.7042	6.7045	6.9069	7.0135	6.7530	6.4804	6.2910
2	6.9840	1.6222e+04	7.0192	7.0356	7.0523	7.2615	7.3754	7.1071	6.7741	6.5727
3	6.7037	7.0192	1.4983e+04	6.7196	6.7342	6.9380	7.0436	6.7839	6.5010	6.3145
4	6.7042	7.0356	6.7196	1.5048e+04	6.7684	6.9578	7.0621	6.8030	6.4969	6.3274
5	6.7045	7.0523	6.7342	6.7684	1.5119e+04	6.9745	7.0799	6.8197	6.5013	6.3135
6	6.9069	7.2615	6.9380	6.9578	6.9745	1.5786e+04	7.2862	7.0233	6.6990	6.5030
7	7.0135	7.3754	7.0436	7.0621	7.0799	7.2862	1.6404e+04	7.1332	6.8019	6.6015
8	6.7530	7.1071	6.7839	6.8030	6.8197	7.0233	7.1332	1.5302e+04	6.5442	6.3753
9	6.4804	6.7741	6.5010	6.4969	6.5013	6.6990	6.8019	6.5442	1.4380e+04	6.1424
10	6.2910	6.5727	6.3145	6.3274	6.3135	6.5030	6.6015	6.3753	6.1424	1.3597e+04

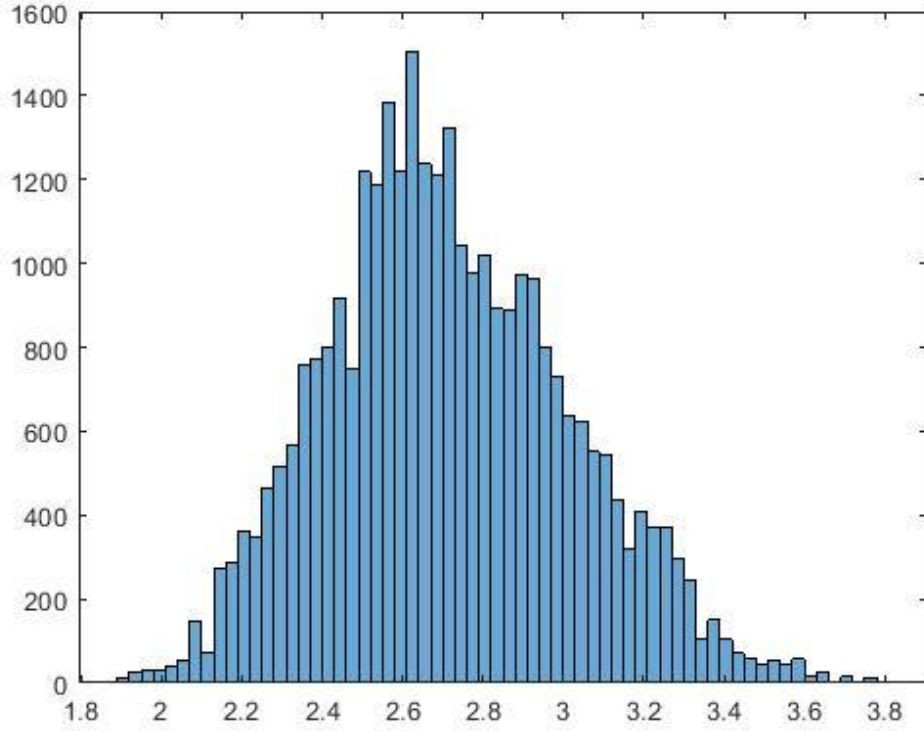
Şekil 4.18. İktisat bölümü verissetinden hesaplanan özdeğer – özvektör matrisi

Ayrıca Tekil değer ayrıştı algoritması sonucunda Σ değerlerini hesaplamak amacıyla İktisat bölümü ve $m=2196$ öğrenci, $n=108$ ders için oluşturulan verimatrixinden hesaplanan özdeğer – özvektör matrisi ekran görüntüsü Şekil 4.18.'de verilmiştir. Σ özdeğer özvektör matrisi grafiği Şekil 4.18.'da verilmiştir. Şekil 4.19. incelendiğinde taban mor renkli alan,sıfır değeri ile doldurulup, köşegen elemanların karekökleri alınarak asıl köşegene yerleştirilirse öz vektörler hesaplanmaktadır. Özdeğerleri ve Köşegenden yükselen sütunlar ise özvektörleri göstermektedir.



Şekil 4.19. İktisat bölümü verissetinden hesaplanan özdeğer – özvektör matrisi grafiği

İşletme bölümü veri matrisinden, test verisi $n = 108$ ders seçilmiş ve seçilen derslere ilişkin elde edilen tahmin ders başarı puanları histogram grafiği Şekil 4.20.' de verilmiştir.

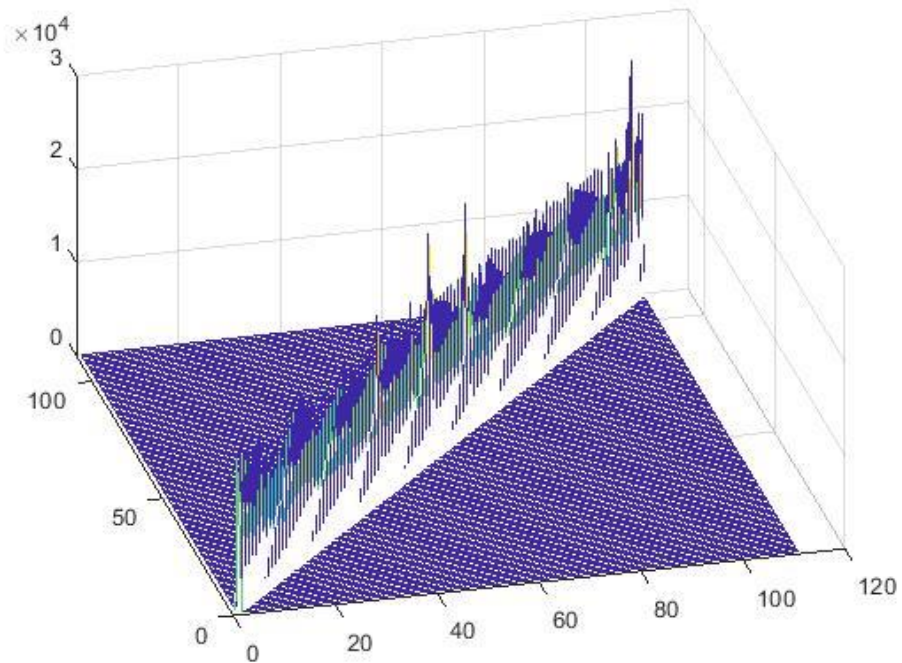


Şekil 4.20. İşletme bölümü tekil değer ayrışımı test verisi tahmin ders başarı puanları grafiği.

	1	2	3	4	5	6	7	8	9	10
1	1.6453e+04	7.5562	7.2103	7.3019	6.7990	7.3245	7.5706	6.7087	7.0072	6.6908
2	7.5562	1.7750e+04	7.5039	7.6090	7.0940	7.5971	7.8793	6.9858	7.2939	6.9497
3	7.2103	7.5039	1.6249e+04	7.2519	6.7706	7.2458	7.5197	6.6799	6.9732	6.6458
4	7.3019	7.6090	7.2519	1.6711e+04	6.8808	7.3403	7.6219	6.7743	7.0736	6.7264
5	6.7990	7.0940	6.7706	6.8808	1.5018e+04	6.8333	7.0826	6.3693	6.6181	6.3087
6	7.3245	7.5971	7.2458	7.3403	6.8333	1.6595e+04	7.6077	6.7470	7.0446	6.7206
7	7.5706	7.8793	7.5197	7.6219	7.0826	7.6077	1.7805e+04	6.9948	7.3140	6.9451
8	6.7087	6.9858	6.6799	6.7743	6.3693	6.7470	6.9948	1.4719e+04	6.5606	6.2267
9	7.0072	7.2939	6.9732	7.0736	6.6181	7.0446	7.3140	6.5606	1.5809e+04	6.4570
10	6.6908	6.9497	6.6458	6.7264	6.3087	6.7206	6.9451	6.2267	6.4570	1.4467e+04

Şekil 4.21. İşletme bölümü verisetinden hesaplanan özdeğer – özvektor matrisi

Ayrıca Tekil değer ayrışması algoritması sonucunda \sum değerlerini hesaplamak amacıyla İşletme bölümü ve $m=2239$ öğrenci, $n=111$ ders için oluşturulan verimatrixinden hesaplanan özdeğer – özvektor matrisi ekran görüntüsü Şekil 4.21.'de verilmiştir. Şekil 4.22. incelendiğinde taban mor renkli alan, sıfır değeri ile doldurulup, köşegen elemanların karekökleri alınarak asıl köşegene yerleştirilirse öz vektörler hesaplanmaktadır özdeğerleri ve köşegenden yükselen sütunlar ise özvektörleri göstermektedir.



Şekil 4.22. İşletme bölümü verisetinden hesaplanan özdeğer – özvektor matrisi grafiği

4.3. İktisat Bölümü İçin Ders Önerisi

Yükseköğretimde işbirliğine dayalı filtreleme seçmeli ders öneri modeli oluşturmada, en iyi sonucu veren SVD tekniği ile edilmiştir. Bu nedenle ders öneri sunumu İktisat bölümündeki 2196 öğrenci arasından rassal 5 öğrenci için ve karşılaştırma yapılabilmesi amacıyla seçili 30 ders için SVD ders tahminleri kişiselleştirilmiş bir öneri olarak örnek model Şekil 4.23.'de verilmiştir. İşbirliğine dayalı ders seçmede öneri modeli; “yol gösterici (pathfinder)” olarak isimlendirdiğimizden öneride yol gösterici oklar kullanılmıştır. Yeşil renkli oklar yüksek başarı, sarı renkli oklar ortalama bir başarı ve kırmızı renkli oklar ise başarısız veya düşük puan tahmin edildiğini göstermektedir.

DERSLER	Öğr_ID_101	DERSLER	Öğr_ID_877	DERSLER	Öğr_ID_914	DERSLER	Öğr_ID_1145	DERSLER	Öğr_ID_1945
	ÖNERİ		ÖNERİ		ÖNERİ		ÖNERİ		ÖNERİ
İKT304	↑ 2.5	TÜR152	↑ 2.7	İKT301	↑ 2.3	İKT338	↑ 2.7	İŞL101	↑ 3.1
İKT301	↑ 2.5	TÜR151	↑ 2.6	MAT198	↑ 2.3	İKT337	↑ 2.7	MAT198	↑ 3.1
MAT198	↑ 2.3	İŞL101	↑ 2.6	İKT304	↑ 2.2	İKT301	↑ 2.6	TÜR151	↑ 3.1
İŞL101	↑ 2.2	İKT301	↑ 2.6	TÜR152	↑ 2.2	İKT304	↑ 2.6	İKT301	↑ 3.1
TÜR152	↑ 2.1	İKT304	↑ 2.5	TÜR151	↑ 2.2	İŞL101	↑ 2.6	İKT304	↑ 3.1
TÜR151	↑ 2.1	MAT198	↑ 2.5	İKT218	↑ 2.2	TÜR152	↑ 2.6	TÜR152	↑ 3.1
İKT310	↑ 2.0	İKT310	↑ 2.4	İŞL101	↑ 2.1	MAT198	↑ 2.6	İKT310	↑ 3.1
İKT303	⇒ 2.0	İKT303	↑ 2.4	İKT310	↑ 2.1	TÜR151	↑ 2.6	İKT303	↑ 3.1
İKT220	⇒ 1.9	İKT220	↑ 2.3	İKT303	↑ 2.0	İKT310	↑ 2.6	İKT220	↑ 3.1
MAT197	⇒ 1.9	MAT197	↑ 2.3	İKT220	↑ 2.0	İKT303	↑ 2.6	MAT197	↑ 3.1
İKT218	⇒ 1.8	İKT218	⇒ 2.2	MAT197	↑ 1.9	İKT220	↑ 2.6	İKT218	↑ 3.1
HUK126	⇒ 1.7	HUK126	⇒ 2.2	İKT302	↑ 1.9	MAT197	↑ 2.6	HUK126	↑ 3.0
İKT128	⇒ 1.7	İKT128	⇒ 2.1	HUK126	⇒ 1.8	İKT302	↑ 2.5	İKT128	↑ 2.9
İKT302	⇒ 1.6	İKT302	⇒ 2.1	İKT128	⇒ 1.8	İKT218	⇒ 2.5	İKT302	↑ 2.9
İKT219	⇒ 1.6	İKT219	⇒ 2.1	İKT219	⇒ 1.7	HUK126	⇒ 2.5	İKT219	⇒ 2.8
İKT308	↓ 1.5	İKT308	⇒ 2.0	İKT309	⇒ 1.7	İKT128	⇒ 2.5	İKT308	⇒ 2.8
ARY111	↓ 1.4	ARY111	⇒ 2.0	İKT308	⇒ 1.6	İKT219	⇒ 2.4	ARY111	⇒ 2.7
İKT217	↓ 1.4	İKT217	⇒ 1.9	ARY111	⇒ 1.6	İKT308	⇒ 2.4	İKT217	⇒ 2.7
İST206	↓ 1.3	İST206	⇒ 1.9	İKT217	⇒ 1.5	ARY111	⇒ 2.4	İST206	⇒ 2.6
MUH115	↓ 1.3	MUH115	⇒ 1.8	İST206	⇒ 1.5	İKT217	⇒ 2.3	MUH115	⇒ 2.6
İKT309	↓ 1.2	İKT309	↓ 1.8	MUH115	⇒ 1.4	İST206	⇒ 2.3	İKT309	⇒ 2.5
HUK151	↓ 1.1	HUK151	↓ 1.7	HUK151	↓ 1.3	MUH115	⇒ 2.3	HUK151	⇒ 2.5
İKT216	↓ 1.1	İKT216	↓ 1.7	İKT216	↓ 1.3	İKT309	⇒ 2.3	İKT216	⇒ 2.4
MUH116	↓ 1.0	MUH116	↓ 1.6	MUH116	↓ 1.2	HUK151	⇒ 2.2	MUH116	⇒ 2.4
İST205	↓ 1.0	İST205	↓ 1.6	İST205	↓ 1.2	İKT216	⇒ 2.2	İST205	↓ 2.3
HUK152	↓ 1.0	HUK152	↓ 1.6	HUK152	↓ 1.1	MUH116	↓ 2.2	HUK152	↓ 2.3
MLY205	↓ 1.0	MLY205	↓ 1.5	MLY205	↓ 1.1	İST205	↓ 2.2	MLY205	↓ 2.2
İKT338	↓ 1.0	İKT338	↓ 1.5	İKT338	↓ 1.0	HUK152	↓ 2.1	İKT338	↓ 2.2
İKT337	↓ 1.0	İKT337	↓ 1.4	İKT337	↓ 1.0	MLY205	↓ 2.1	İKT337	↓ 2.1
SOS124	↓ 1.0	SOS124	↓ 1.4	SOS124	↓ 1.0	SOS124	↓ 2.0	SOS124	↓ 2.1

Şekil 4.23. İktisat bölümü tekil değer ayrışımı algoritması sonucu elde edilen tahmin başarı paunları ile ders seçme öneri örneği.

4.4. İşletme Bölümü İçin Ders Önerisi

Ders öneri sunumu örneği için İşletme bölümündeki 2239 öğrenci arasından rassal 5 öğrenci için ve karşılaştırma yapılabilmesi amacıyla seçili 30 ders için SVD ders tahminleri kişiselleştirilmiş bir öneri olarak örnek model Şekil 4.24.'de verilmiştir. İşbirliğine dayalı ders seçmede öneri modeli; “yol gösterici (pathfinder)” olarak isimlendirdiğimizden öneride yol gösterici oklar kullanılmıştır. Yeşil renkli oklar yüksek başarı, sarı renkli oklar ortalama bir başarı ve kırmızı renkli oklar ise başarısız veya düşük puan tahmin edildiğini göstermektedir.

DERSLER	Öğr_ID_25	DERSLER	Öğr_ID_652	DERSLER	Öğr_ID_733	DERSLER	Öğr_ID_1278	DERSLER	Öğr_ID_2098
DERSLER	ÖNERİ	DERSLER	ÖNERİ	DERSLER	ÖNERİ	DERSLER	ÖNERİ	DERSLER	ÖNERİ
İŞL111	↑ 3.6	İŞL111	↑ 3.4	İŞL111	↑ 3.4	İŞL111	↑ 3.6	İŞL111	↑ 3.7
İKT452	↑ 3.4	İNG150	↑ 3.3	İST206	↑ 3.3	BİL150	↑ 3.4	İST206	↑ 3.5
İST206	↑ 3.3	İST206	↑ 3.3	İŞL113	↑ 3.2	İST206	↑ 3.3	BİL285	↑ 3.5
İKT101	↑ 3.3	İŞL113	↑ 3.2	İNG175	↑ 3.1	İŞL113	↑ 3.2	İŞL113	↑ 3.4
İŞL113	↑ 3.2	İNG175	↑ 3.1	HUK151	↑ 3.1	BİL285	↑ 3.2	İKT209	↑ 3.4
İNG175	↑ 3.1	BİL285	↑ 3.1	BİL285	↑ 3.1	İNG175	↔ 3.1	HUK251	↑ 3.4
BİL285	↑ 3.1	HUK251	↑ 3.1	İŞL112	↑ 3.0	HUK251	↔ 3.1	İŞL112	↔ 3.0
HUK151	⇒ 3.1	HUK151	↑ 3.0	FİN301	↑ 2.9	İŞL112	↔ 3.0	İST205	↔ 2.9
İŞL112	⇒ 3.0	İŞL112	↑ 3.0	İŞL202	↑ 2.9	BİL151	↔ 3.0	İŞL202	↔ 2.9
İST205	⇒ 2.9	İST205	↑ 2.9	FİN302	⇒ 2.9	İST205	↔ 2.9	HUK151	↔ 2.9
İŞL202	⇒ 2.9	İŞL202	↑ 2.9	BİL811	⇒ 2.8	İŞL202	↔ 2.9	İSN315	↔ 2.9
HUK305	⇒ 2.9	HUK152	⇒ 2.8	İNG150	⇒ 2.8	BİL811	↔ 2.9	HUK152	↔ 2.8
BİL811	⇒ 2.8	İKT452	⇒ 2.8	İNG176	⇒ 2.8	HUK151	↔ 2.9	BİL811	↔ 2.8
İNG150	⇒ 2.8	BİL811	⇒ 2.8	İŞL103	⇒ 2.8	HUK152	↔ 2.8	İKT452	↔ 2.8
FİN302	⇒ 2.8	İŞL103	⇒ 2.8	HUK251	⇒ 2.7	İKT452	↔ 2.8	İŞL103	↔ 2.8
İŞL103	⇒ 2.8	FİN301	⇒ 2.7	İKT101	⇒ 2.7	İNG150	↔ 2.8	FİN301	↔ 2.7
İKT209	⇒ 2.8	İKT101	⇒ 2.7	HUK152	⇒ 2.7	İŞL103	↔ 2.8	İNG175	↔ 2.7
FİN301	⇒ 2.7	HUK305	⇒ 2.7	BİL150	⇒ 2.6	BEÖ155	↔ 2.8	İKT101	↔ 2.7
İST435	⇒ 2.6	İST435	⇒ 2.6	İST435	⇒ 2.6	FİN301	↔ 2.7	HUK305	↔ 2.7
İKT102	⇒ 2.6	BİL151	⇒ 2.6	BİL151	⇒ 2.6	İKT101	↔ 2.7	İST435	↔ 2.6
BİL151	⇒ 2.6	BSİ304	⇒ 2.5	BSİ304	⇒ 2.6	HUK305	↔ 2.7	BİL151	↔ 2.6
İŞL214	⇒ 2.6	İŞL214	⇒ 2.5	BEÖ155	⇒ 2.5	İST435	↔ 2.6	İKT102	↔ 2.6
BSİ304	↓ 2.6	FİN302	⇒ 2.5	İŞL214	⇒ 2.5	İŞL214	↔ 2.5	BSİ304	↔ 2.5
ARY209	↓ 2.5	ARY209	⇒ 2.5	ARY209	⇒ 2.5	BSİ304	↔ 2.5	İNG150	↔ 2.5
BEÖ155	↓ 2.4	BEÖ155	↓ 2.3	İKT209	↓ 2.3	FİN302	↔ 2.5	İŞL214	↔ 2.5
İSN315	↓ 2.3	İKT209	↓ 2.3	İST205	↓ 2.3	ARY209	↔ 2.5	FİN302	↔ 2.5
HUK251	↓ 2.3	İSN315	↓ 2.3	İSN315	↓ 2.3	İKT209	↓ 2.3	İNG176	↔ 2.5
BİL150	↓ 2.1	BİL150	↓ 2.1	İKT452	↓ 2.2	İSN315	↓ 2.3	ARY209	↓ 2.5
İNG176	↓ 2.1	İNG176	↓ 2.1	HUK305	↓ 2.0	İNG176	↓ 2.1	BEÖ155	↓ 2.3
HUK152	↓ 2.1	İKT102	↓ 1.9	İKT102	↓ 1.9	İKT102	↓ 1.9	BİL150	↓ 2.1

Şekil 4.24. İktisat bölümü tekil değer ayrışımı algoritması sonucu elde edilen tahmin başarı paunları ile ders seçme öneri örneği.

5. SONUÇ VE ÖNERİLER

Bu tezde, İşbirliğine dayalı filtrelemeye dayalı ders seçme öneri modeli önerilmektedir. İşbirliğine dayalı filtreleme tekniği, bu alandaki çalışmalar incelendiğinde, öğrencilere ders seçimi önerisi için öğrenci puanlarını tahmin edilmesinde ve öğretim kalitesi değerlendirmesinde yararlanılan en başarılı kişiselleştirilmiş öneri teknolojisidir. Ders önerisi sistemi, öğrenciler tarafından seçilen derslerin en iyi kombinasyonunu tahmin etmeyi amaçlamaktadır. Bu amaçla ders öneri sisteminde İşbirliğine dayalı filtreleme tekniği ile öğrencinin sahip olduğu başarı notundan yararlanarak gelecekte seçeceği düşünülen derslere ait akademik başarı tahmini yapan öneri algoritmaları karşılaştırılmıştır. Öncelikle, öğrencilerin birbirine olan akademik başarı notu benzerlikleri için kümeleme algoritmalarından yararlanılmıştır. Benzerlik derecesi öğrencinin akademik derece olarak en benzer kümeleri oluşturmada ve benzerlik yükleri yardımıyla öğrenci için ders tavsiyeleri verilmesinde kullanılmıştır.

İşbirliğine dayalı filtreleme tekniği ile ders önerme yönteminde veri matrisinin kısıtlılık problemi olan eksik başarı puanlarının tahmin edilmesini hedefleyen algoritmalar geliştirilmiştir. Kümeleme algoritmaları arasından literatürde en yaygın ve etkili tahmin sonuçları veren K-Ortalamlar ve KNN algoritmaları uygulanmıştır. Bu algortmadan birisi kümeleme tekniklerine dayalı olarak öncelikle veri matrisi üzerinden öğrenci ve dersler için Pearson korelasyon katsayısı yardımıyla benzerlikler hesaplanmış ve k-ortalamlar ve KNN kümeleme algoritmaları uygulanmıştır. Kümeleme, nesnelerin gruplarını o şekilde bulur ki, bir grup içindeki nesnelere birbirlerine benzerken, başka gruptaki nesnelere farklıdır. Kümeleme en önemli denetimsiz(unsupervised) öğrenme tekniğidir (Dol Aher ve Lobo, 2012). Daha sonra öğrencinin başarı notunu tahmin için İşbirliğine dayalı filtreleme tekniklerinden birisi olan ders - ders benzerliğine dayalı olarak algoritma uygulanmıştır. Ayrıca, öğrencinin almış olduğu ders başarı puanlarından oluşan matristeki seyreklik problemi için, boyut indirgeme yöntemi olan SVD tekniği ile gizil yapının ortaya çıkarılması ile ders öneri algoritmaları denenmiştir. Veri matrisindeki kısıt seyreklik problemi kümeleme algoritmaları için benzerlikler hesaplanmasında ve uygun komşulukların belirlenmesinde bir sorun olarak karşımıza çıkmıştır. Bu sorunun giderilmesinde veri matrisi için boyut indirgeme tekniği olan tekil değer ayrıştırma (SVD) uygulanmıştır.

İşbirliğine dayalı filtreleme algoritması sonucunda ders seçme önerisi, öğrencilere seçmeli derslere ilişkin öngörülen bir bilgi oluşturarak, öğrencinin ders seçiminde belirsizliği etkin bir şekilde azaltmaya yarayacaktır. Ayrıca Üniversitelerdeki, fakülte yönetimine ders planlaması yapmada hangi derslerin önerilmesinin belirlenmesinde en etkin ve en az hatayla tahmin edilmesinde kullanılacak algoritma ya da algoritmalarından yararlanılacağını gerçek veri seti üzerinde test ederek yardımcı olmaya çalışılmıştır. Bu uygulanan öğrenci benzerliğine dayalı K-means, KNN, SVD ile ders benzerliğine dayalı (Item base), algoritmaları sonucunda, öğrencileri seçmeli derslerin seçimindeki belirsizliği etkili bir şekilde azaltarak, öğrencilere seçecekleri derslerin başarı hedefine yönelik olarak öneride bulunabileceğini sonucuna ulaşılmıştır. Ancak gerçek veri seti üzerinde uygulanan algoritmalar sonucunda SVD en iyi performansı göstermiştir. Algoritmalar, Anadolu Üniversitesi İİBF İktisat ve İşletme bölümleri için ayrı ayrı uygulanmış ve aşağıdaki bulgular elde edilmiştir.

İktisat bölümü için, sırasıyla kullanıcı tabanlı yaklaşımda ilk olarak, KNN algoritması sonucunda MAE ölçütü 0,5865 ile 0,6045 arasında hesaplanmıştır. En düşük MAE değeri 0,5865 ile komşu sayısının 250 olarak belirlenmesinde elde edilmiştir.

İkinci olarak İktisat bölümü için, K-ortalamlar tekniği sonucunda MAE ölçütü 0,5307 ile 0,5348 arasında hesaplanmıştır. En düşük MAE değeri 0,5307 ile komşu sayısının 8 olarak belirlenmesinde elde edilmiştir.

Üçüncü olarak, İktisat bölümü için Tekil değer algoritması sonucunda MAE ölçütü 0,288 ile 0,2034 arasında hesaplanmıştır. SVD ile $k=8$ değerinde en düşük MAE değeri 0,2034 olarak bulunmuştur. Son olarak (item-base) Ders benzerliğine dayalı algoritması için en düşük MAE komşu sayısı 25 olduğunda bulunmuş ve bundan sonra 0,79 olarak değişmemiştir.

İşletme bölümü için, İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda KNN algoritması sonucunda MAE ölçütü 0,5827 ile 0,6578 arasında hesaplanmıştır. En düşük MAE değeri 0,5827 ile komşu sayısının 250 olarak belirlenmesinde elde edilmiştir.

İkinci olarak İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda k-ortalama algoritması sonucunda MAE ölçütü 0,5616 ile 0,5893 arasında hesaplanmıştır. En düşük MAE değeri 0,5616 ile komşu sayısının 16 olarak belirlenmesinde elde edilmiştir.

Üçüncü olarak, İşbirliğine dayalı filtreleme tekniklerinden birisi olan kullanıcı tabanlı yaklaşımda Tekil değer algoritması sonucunda MAE ölçütü 0,1868 ile 0,1895 arasında hesaplanmıştır. En düşük MAE değeri 0,1895 ile $k=12$ ve $k=16$ olarak belirlenmesinde elde edilmiştir. SVD ile $k=12$ değerinde en düşük MAE değeri 0,1895 olarak bulunmuştur.

Son olarak İşletme bölümü için, (item-base) Ders benzerliğine dayalı algoritması için MAE ölçütü 0,785 ile 0,7629 arasında hesaplanmıştır. En düşük MAE değeri 0,7629 ile $k=50$ ve üzerinde MAE değeri değişmemiş sabit kalmıştır.

Seçmeli ders seçimi için İşbirliğine dayalı filtreleme tekniklerinden en çok kullanılan ve en iyi tahmin elde edilen algoritmalar denenmiş ve bunun sonucunda bir ders öneri sistemi mimarisi için başarı puanları tahmini elde edilmiştir. Bu algoritmalarından en düşük MAE sonucu $k=8$ için 0,02034 olarak Tekil değer ayrışımı sonucunda ulaşılmıştır. Bu sonuca göre ders öneri sisteminde, öğrenci ders başarılarının tahmininde gizil ilişkilerin ortaya konulması ve boyut azaltarak hesaplama hızında da büyük artış ve en düşük hata ile (SVD) tekil değer ayrışımı algoritması en iyi sonucu vermiştir.

Öğrenci için ders öneri sistemlerinin psikolojik bağlamda da ele alınması, öğrencinin güveninden daha çok eğitsel madencilikte «son mil» problemi olarak tanımlanan mezuniyetini etkileyecek ve engeller oluşmasına neden olabilir. Üniversite öğrencileri için ders seçiminde öneri sistemlerini ticari öneri sistemlerinden ayırmak gerekmektedir. Çünkü ticari öneri sistemlerinde doğru olmayan bir öneri ile belki bir tüketici kaybı yaşanmasına veya güven kaybı oluşmasına neden olunabilir.

Birçok araştırmada öneri sisteminin başarısının kullanıcı tatminine dayandığı söylenmektedir. Ticari öneri sistemleri uygulamalarında, tatmin kullanıcının satın alınan ve iade edilmeyen ürünler temelinde yapılmaktadır. Ticari olmayan öneri sistemlerinde ise

kullanıcının ne ölçüde tatmin olduğu sorgulanmaktadır. Seçmeli ders için tasarladığımız öneri sisteminde, öğrenciler (kullanıcılar) için tatmin öğrenciler üzerinde yapılan anketler ile ölçülebileceği gibi öneri sisteminin öncesinde ve sonrasında da seçmeli derslerin not ortalamaları arasında anlamlı farklılıkların bulunması bize önerilerin performansına ilişkin ipuçları verebilir. Ayrıca öneri sistemi ile her bir öğrenci için önerisi sistemi sonrası seçtiği derslerdeki tahmin edilen başarı notu ile seçilen derse ait gerçekleşen başarı notu karşılaştırması ve genel başarı ortalamaları farklarının karşılaştırılması bize sistemin performansı hakkında bilgi sağlayacaktır.

Ticari öneri sistemi algoritmalarında aynı veri seti üzerinden en iyi tahmin ve en iyi öneri oluşturmaya çalışılırken, ders öneri sistemleri kapsam ve kural gereği herkese açık veri olmadığından, Ancak uygulama sonrası öğrencilerin tahmin ve gerçekleşen başarı kıyaslamasıyla doğruluğu tespit edilebilir. Ticari uygulamalarda örneğin film öneri sisteminde geri bildirim çok kısa sürede alınırken, Öğrenci ders öneri sisteminde bu en az bir dönem ve daha uzundur. Ticari uygulamalarda adaptif öneri sistemleri, zaman ve kişisel alışkanlıkların değişimine bağlı olarak çeşitli uygulamalar ve algoritmalar ile test edilebilir. Geri bildirim süresinin kısa olması bu imkânı sağlamaktadır. Ders öneri sistemlerinde geri bildirim süresinin daha uzun olması, ayrıca önerilen dersin öğrencinin başarısını artırmaya yönelik olması, öğrenci öneri sistemlerinin her seferinde yeniden tasarlanması ve yeniden hesaplanması gerekmektedir. Ticari bir öneri de önerilen romantik bir film türünü çok beğenmesi, onun daha sonraki önerilecek Romantik film türünde daha çok beğenmesi üzerinde etkisi yoktur. Bu benzer beğenisini artırmaz, diğer bir ifade ile duygularına etkisi yoktur.

Ders öneri sisteminde «başarı başarıyı getirir» sözü ile başarılı bir öğrencinin motivasyonun artması, derse katılımı, özgüveninin yükselmesi belki de ders öneri sistemlerinin gizil katkısıdır. Böylelikle daha sonra seçeceği derslere ilişkin, tıpkı film önerisindeki gibi derslere karşı duygularında bir etkisi olmada, ön yargıları yıkarak motivasyonu artırabilir. Öğrencinin derslerine ilişkin başarı odaklı motivasyonun artması başarılı olma olasılığını da artıracaktır. Korelasyon ile nedensellik için sonbaharda yaprakların sararıp yere düşmesi örneğindeki gibi, yaprakların sararıp düşmesine neden olan soğuk değil, güneş ışınlarının yatay gelmesiydi. Bu bağlamda öğrencilerin motivasyonlarının artması ve başarılı olacağı inancı, yapraklara ilkbahar ve yaz aylarındaki

gelen dik güneş ışını gibidir. Böylece öğrenci sistemden kopmadan, eğitimine ara vermeden, başarılı bir mezuniyetle iş hayatına dâhil olacaktır.

Bu tez çalışması sonucunda, üniversiteler de öğrenciler için ders seçiminde yararlı olacak, bir ders öneri sisteminin alt yapısını oluşturacak model öneri sunulmuş ve modelde yararlanılacak bazı algoritmalar denenmiş ve sonuçları tartışılmıştır. Yükseköğretim kurumlarında öğrenci ders seçiminde karşılaşılan problemlerin çözümünde, bu tez de sunulan öneri sistemi modelinin yararlı olacağını düşünüyoruz.

Yükseköğretimde büyük bir öğrenci kitlesine sahip kitlesel eğitim veren Açık ve Uzaktan eğitim kurumları için e-danışman sistem oldukça önemlidir. Açık ve uzaktan eğitim gören öğrencilerin ders atamasında bir danışman desteği yoktur. Bu nedenle öğrencilere ders seçiminde yol gösterici, seçeceği derslerden başarı puanlarını tahmin eden ve genel not ortalaması GNO hesaplayan otomatik bir öneri sistemine ihtiyaç vardır. Önerilen “yol gösterici” veya “pathfinder” e-danışman sistemi uzaktan eğitim kurumlarının bu sorunu çözeceğine inanmaktayız. Ayrıca “yol gösterici” veya “pathfinder” e-danışman ders öneri sistemini kurmayı planlayan üniversitelere ve bu konuda araştırma yapacaklara yol göstermesini umut ediyoruz.

KAYNAKLAR DİZİNİ

- Adomavicius, G. ve Tuzhilin, A., 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6).
- Almazro, D. ve diğerleri, 2010. A Survey Paper on Recommender Systems.
- Arnold, K. E. ve Pistilli, M. D., 2012. Course signals at Purdue: Using learning analytics to increase student success. %1 içinde New York, NY. ACM: Proceedings of the 2nd International Conference on Learning Analytics ve Knowledge.
- Baker, K., 2005. Singular Value Decomposition Tutorial”.
- Barneveld, A. v., Arnold, K. E. ve Campbell, J. P., 2012. Analytics in Higher Education:
- Bendakir, N. A. E., 2006. Using Association Rules for Course Recommendation. Boston,
- Bennet, J. L., 2007. The Netflix Prize. San Jose, California,
- Brooks, C. ve Thompson, C., 2017. Predictive Modelling in Teaching and Learning. %1 içinde The Handbook of Learning Analytics. Alberta, Canada: Society for Learning Analytics Research (SoLAR), pp. 61-68.
- Chatti, M., Dyckhoff, A., Schroeder, U. ve Thüs, H., 2012. A Reference Model for Learning Analytics. In *International Journal of Technology Enhanced Learning*, pp. 1-22.
- Chen, S., Ma, B. ve Zhang, K., 2009. On the similarity metric and the distance metric. May, Issue 410, pp. 2365-2376.
- Chih, M. C., Lee, H. ve Chen, Y., 2005. Personalized E-Learning System Using Item Response Theory. *Computers ve Education*.

KAYNAKLAR DİZİNİ (devam)

- Crookston, B. B., 1972. A developmental view of academic advising as teaching.. Journal of College Student Personne, pp. 12-17.
- Davenport, T. H., Harris, J. G. ve Morison, R., 2010. Analytics at work: Smarter decisions, better results.. Boston,: Harvard Business Press..
- Denley, T., 2013. Degree Compass: A Course Recommendation System. <https://er.educause.edu/articles/2013/9/degree-compass-a-course-recommendation-system> [Erişildi: 6 3 2018].
- Diaz, V. ve Brown, M., 2012. Learning Analytics: A Report on the ELI Focus Session, basım yeri bilinmiyor: EDUCAUSE Learning Initiative.
- Dol Aher, S. ve Lobo, L. M. R. J., 2012. Prediction of Course Selection by Student using Combination of Data Mining Algorithms in E-Learning.. nternational Journal of Computer Applications., Issue 40, pp. 1-7.
- Everitt, B. S., Sabine, L., Morven, L. ve Daniel, S., 2011. Cluster Analysis. United Kingdom: John Wiley ve Sons Ltd..
- Farzan, R. ve Brusilovsky, P., 2006. Social Navigation Support in a Course Recommendation System.. %1 içindedaptive Hypermedia and Adaptive Web-Based Systems.. Berlin, Heidelberg:
- Field, A., 2009. Discovering Statistics Using SPSS. 3. dü. London: SAGE Publications Ltd.
- Fisher, R. J., 1993. Social desirability bias and the validity of indirect questioning. Journal of Consumer Research.
- G. Strang, 2009. Introduction to Linear Algebra. Wellesley: Cambridge Press.

KAYNAKLAR DİZİNİ (devam)

- Gadanho, S. ve Lhuillier, N., 2007. Addressing uncertainty in implicit preferences: Proceedings of the 2007 ACM conference on Recommender systems.
- Gan, G., Ma, C. ve Wu, J., 2007. Data Clustering: Theory, Algorithms, and Applications. Siam.
- Goldberg, D., Nichols, D., Oki, B. M. ve Terry, D., 1992. Using collaborative filtering to weave an information tapestry. ACM, 35(12), pp. 61-70.
- Golub, G. H. ve Loan, C. F. V., 1996. Matrix Computations. Baltimore and London: JHU Press.
- Golumbic, M., C. Markovich, M., Tsur, S. ve Schild, U., 1986. A knowledge-based expert system for student advising. IEEE, pp. 120-123.
- Good, N. S. ve diğerleri, 1999. Combining collaborative filtering with personal agents for better recommendations.
- Greene, K., 2006. "The \$1 million Netflix challenge". Technology Review.
- Gürsakal, N., 2017. Büyük Veri. 3. dü. Bursa: DORA.
- Harmelen, M. V. ve Workman, D., 2012. Analytics for learning and teaching. CETIS Analytics Series,.
- Hartigan, J. A., 1975. Clustering Algorithms. 99. New York, NY, USA. John Wiley ve Sons, Inc..
- Heisserer, D. L. ve Parette, P., 2002. Advising at-risk students in college and university settings.. College student journal, Mart.

KAYNAKLAR DİZİNİ (devam)

- Herlocker, J., 2004. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems, Issue 22, pp. 5-53.
- Hernández-del-Olmo, F. ve Gaudioso, E., 2008. Evaluation of recommender systems: A new approach. Expert Systems and Application, Issue 35, pp. 790-830.
- Huang, Z., Chen, H. ve Zeng, D., 2004. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. New York, NY, USA, ACM, pp. 116-142.
- Isinkaye, F., Folajimi, Y. ve Ojokoh, B., 2015. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal, Issue 16.
- Jahrer, M., Töschler, A. ve Legenstein, R., 2010. Combining predictions for accurate recommender systems. , pp. 693-702.
- Jannach, D., Zanker, M., Ge, M. ve Gröning, M., 2012. Recommender Systems in Computer Science and Information Systems. International Conference on Electronic Commerce and Web Technologies, pp. 76-87.
- Jason Burby, A. B. ve . W. S. C., 2007. <https://www.digitalanalyticsassociation.org/>.
[Çevrimiçi]
https://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitionsVol1.pdf [Erişildi: 24 Nisan 2018].
- Johnson, L. ve diğerleri, 2013. NMC Horizon Report: 2013 Higher Education Edition.. Austin, TEKSAS:
- Johnson, L., S Adams, B., V, E. ve A, F., 2014. NMC Horizon Report: 2014 Higher Education Edition.. Austin, Teksas, The New Media Consortium, pp. 201-204.

KAYNAKLAR DİZİNİ (devam)

Johnson, L. ve diğerleri, 2011. The 2011 Horizon Report.. Austin, Texas.:

Johnson, R. ve Wichern, D., 2007. Applied Multivariate Statistical Analysis. Prentice Hall, Upper Saddle River, NJ: Pearson.

Jolliffe, I., 1986. Principal Component Analysis and Factor Analysis. basım yeri bilinmiyor:Springer series in statistics.

Joyce, T. N. R., 1958. The Thesaurus Approach To Information Retrieval. American Documentation 9.

Kaufman, L. ve Rousseeuw, P., 1990. Finding Groups in Data: An Introduction To Cluster Analysis. New York: John Wiley.

Koren, Y., Bell, R. ve Volinsky, C., 2009. Matrix Factorization Techniques for Recommender Systems. in Computer, pp. 30-37.

Lavoué, É. D. H. V. K. B. J. ve . P.-S. M., 2017. Data Driven Approaches in Digital Education. Tallinn, Estonia,

Lemire, D., Boley, H., McGrath, S. ve Ball, M., 2005. Collaborative filtering and inference rules for context-aware learning object recommendation. Interactive Technology and Smart Education, 2(3), pp. 179-188.

Mahmood, T. ve Ricci, F., 2009. Improving recommender systems with adaptive conversational strategies., ACM, pp. 73-82.

Manouselis, N., Drachsler, H., Verbert, K. ve Duval, E., 2012. Recommender Systems for Learning. Springer.

KAYNAKLAR DİZİNİ (devam)

- Manouselis, N., Drachler, H., Verbert, K. ve Duval, E., 2013. Recommender Systems for Learning. Springer.
- Marques., J. M. ve Dario, P., 1994. The 'Black Sheep Effect': Social Categorization, Rejection of Ingroup Deviates, and Perception of Group Variability. 1(5).
- Middleton, S. E., Alani, H. ve De Roure, D. C., 2002. Exploiting Synergy Between Ontologies and Recommender Systems. Semantic web conference, WWW2002, pp. 1-10.
- Murray, H. G. ve Renaud, R., 1995. Disciplinary differences in teaching and learning: Implications for practice. New Directions for Teaching and Learning, 1995(64).
- Parameswaran, A., Venetis, P. ve Garcia-Molina, H., 2011. Recommendation systems with complex constraints: A course recommendation perspective.. ACM Transactions on Information Systems.
- Picciano, A. G., 2012. The Evolution of Big Data and Learning Analytics in American Higher Education.. Journal of Asynchronous Learning Networks, 16(3), pp. 9-20.
- Resnick, P. ve Varian, H. R., 1997. Recommender systems. , pp. 56-58.
- Ricci, F., Rokach, L., Shapira, B. ve Kantor, P. B., 2010. Recommender Systems Handbook. Berlin, Heidelberg: Springer-Verlag.
- Sarwar, B., Karypis, G., Konstan, J. ve Riedl, J., 2000. Application of dimensionality reduction in recommender system: A case study. Boston, MA, USA, , pp. 682-693.
- Sarwar, B. M., Karypis, G., Konstan, J. ve Riedl, J., 2002. Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering.

KAYNAKLAR DİZİNİ (devam)

Sharples, M. ve diğerleri, 2013. Innovating Pedagogy 2013. , pp. 1-42.

Siemens, G., 2010. What are Learning Analytics?. [Çevrimiçi]

Erişim: <http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/> [Erişildi: 1 Nisan 2018].

Singh, K. M., Zenawi, Z. ve Neerja, S., 2008. Impact of Learning Analytics on Curriculum Design and Student Performance. Hershey PA, USA: IGI Global.

Steele, G., Leonard, M., Haberle, C. ve Lipschultz, W., 2013. Technology and Academic Advising: Perceptions and Practices in Higher Education. NACADA Technology in Advising Commission Sponsored Survey.

Stewens, J., 1996. Applied multivariate statistics for the social science. New Jersey: Lawrence Erlbaum Associates.

Su, X. ve Khoshgoftaar, T. M., 2009. A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence.

Tabachnick, B. G. ve Fidell, L., 2001. Using multivariate statistics. 4. dü. Boston: Allyn and Bacon..

Tatlıdil, H., 2002. Uygulamalı Çok Değişkenli İstatistiksel Analiz. Ankara: Ziraat Matbaacılık. TDK, 2018. Türk Dil Kurumu. [Çevrimiçi] http://www.tdk.gov.tr/index.php?option=com_gtsve_elime=LAB%C4%B0RENT [Erişildi: 14 Kasım 2018].

Theodoridis, S. ve Koutroumbas, K., 2008. Pattern Recognition. 4. dü. Orlando, FL, ABD.: Academic Press, Inc..

KAYNAKLAR DİZİNİ (devam)

KAYNAKLAR DİZİNİ (devam)

- Towle, B. ve Clark, Q., 2000. Knowledge Based Recommender Systems Using Explicit User Models.
- Unelsrød, H. F., 2011. Design and Evaluation of a Recommender System for Course Selection.. pp. 1-56.
- Venkatesan, M. ve Thangadurai, K., 2016. Collaborative Filtering Using Data Mining and Analysis. ve :IGI Global.
- Vialardi, C., Bravo, J. ve Ortigosa, A., 2009. Recommendation in Higher Education Using Data Mining Techniques. International Working Group on Educational Data Mining, Paper presented at the International Conference on Educational Dat, pp. 190-199.
- Vialardi, C., Bravo, J., Shafti, L. ve Ortigosa, A., 2009. Recommendation in Higher Education Using Data Mining Techniques,. International Working Group on Educational Data Mining, Paper presented at the International Conference on Educational Dat, pp. 190-199.
- Wall, M., Rechtsteiner, A. ve Rocha, L., 2003. Singular Value Decomposition and principal component analysis, in A Practical Approach to Microarray Data Analysis,. [Çevrimiçi] Erişim: <http://public.lanl.gov/mewall/kluwer2002.html>
- Young, M. ve Muller, j., 2015. Three educational scenarios for the future: Lessons from the sociology of knowledge. %1 içindeCurriculum and the Specialization of Knowledge. ve :Routledge, pp. 76-91.

EK AÇIKLAMALAR

Ek Açıklama - A: Knn matlab iktisat bölümü algoritma

Ek Açıklama - B: Knn matlab işletme bölümü algoritma

Ek Açıklama - C: Tekil Değer Ayrışımı (Svd) Algoritma

EK AÇIKLAMALAR

Ek Açıklama - A: Knn matlab algoritma

```

load iktisat;
load targetItemsList_push; %hedef iktisat listesini yükler
%load target2;
targetItemsList=targetItemsList_push;
MLP=iktisat;
q = targetItemsList';
% sigma = 0
neighbors = 25; % k= 25 komşu sayısı değişebilir.

ALLMean = zeros(size(MLP,1),1);
ALLStd = zeros(size(MLP,1),1);
for tuidx=1:size(MLP,1)
    ALLMean(tuidx,1) = ComputeMean(MLP(tuidx,:));
    ALLStd(tuidx,1) = ComputeStd(MLP(tuidx,:));
end
% start producing predictions
Pr = zeros(size(MLP,1), size(q,2));

for auidx=1:size(MLP,1)
    auidx;
    % create active user and train users dataset
    activeUser = MLP(auidx,:);
    trainUsers = MLP;
    trainUsers(auidx,:)=[];
    TUMean = ALLMean; TUMean(auidx)=[];

```

Ek Açıklama - A: Knn matlab algoritma (devam)

```

TUSStd = ALLStd; TUSStd(auidx)=[];

for itemid=1:size(q,2)
%
    targetItem = q(itemid);
    tmpActiveUser = activeUser;
    tmpActiveUser(targetItem) = 0;
    avg = ComputeMean(tmpActiveUser);
    stddev = ComputeStd(tmpActiveUser);

    possibleNeighbors = trainUsers(trainUsers(:,targetItem)~=0,:);
    possibleNeighborsMean = TUMean(trainUsers(:,targetItem)~=0);
    possibleNeighborsStd = TUSStd(trainUsers(:,targetItem)~=0);
        sim = ComputeSimilarities(possibleNeighbors, possibleNeighborsMean,
            possibleNeighborsStd, tmpActiveUser);% user ve komşuları arasındaki
            pearson benzerliğini hesaplar

% produce prediction normally
[Y,I]=sort(sim,'descend');
if size(possibleNeighbors,1)>neighbors, limit=neighbors;
else limit=size(possibleNeighbors,1); end
sub1=0; sub2=0;

for k=1:limit
    if possibleNeighbors(I(1,k),targetItem) ~= 0
        if ~isnan(Y(1,k))

```

Ek Açıklama - A: Knn matlab algoritma (devam)

```

        sub1      =      sub1      +      (possibleNeighbors(I(1,k),targetItem)-
possibleNeighborsMean(k))*Y(1,k);
        sub2 = sub2 + Y(1,k);
        end
        end
    end
    if sub2~=0
        Pr(auidx,itemid) = avg+sub1/sub2; %ders için prediction hesaplanır.
    else
        Pr(auidx,itemid) = avg;
    end
    if Pr(auidx,itemid)<0, Pr(auidx,itemid)=0;
    elseif Pr(auidx,itemid)>4, Pr(auidx,itemid)=4;
    end

    end
end

k=0;
MAE=0;
for au=1:size(MLP,1)
    for i=1:size(targetItemsList,1)
        if MLP(au,targetItemsList(i,1))~=0 veve ~isnan (Pr(au,i))
            MAE = MAE + abs(Pr(au,i)-MLP(au,targetItemsList(i,1)));
            k=k+1;
        end
    end
end

```

Ek Açıklama - A: Knn matlab algoritma (devam)

```

end
MAE = MAE / k%(20*size(targetItemsList,1));
knnMAE=MAE;
save knnMAE knnMAE
save k k
save Pr Pr
ElapsedTime = toc
MAE
function similarities = ComputeSimilarities(TrainData, TUMean, TUSTd,
ActiveUserVector)
    AUMean = ComputeMean(ActiveUserVector);
    AUStd = ComputeStd(ActiveUserVector);
    similarities=zeros(1,size(TrainData,1));

    for tuidx=1:size(TrainData,1)
        for itemidx=1:size(TrainData,2)
            if ActiveUserVector(1,itemidx)~=0 veve TrainData(tuidx,itemidx)~=0
                similarities(1,tuidx) = similarities(1,tuidx) + (ActiveUserVector(1,itemidx)
- AUMean)*(TrainData(tuidx,itemidx) - TUMean(tuidx,1));
            end
        end
        similarities(1,tuidx) = similarities(1,tuidx)/(AUStd*TUSTd(tuidx,1));
    end

function MeanVal = ComputeMean(Vector)% 1*n vector
    MeanVal = mean(Vector(Vector~=0));

function StdVal = ComputeStd(Vector)% 1*n vector

StdVal = std(Vector(Vector~=0));

```


Ek açıklama -B: Item Base Matlab Algoritma

```

load isletme;
load targetItemsList_push_isletme
MLP=isletme;
qp = targetItemsList_push_isletme';
%sigma = 2;
neighbors = 10;
A=MLP;%(1:5,:);
for auidx=1:size(A,1)
    auidx
        for itemid=1:size(qp,2)
            targetItem = qp(itemid);
            activeUser2=MLP(auidx,:);
            activeUser2(targetItem)=0;
            activeUserItem=find(activeUser2~=0);
            % Active userın rate verdiği itemlar ile targer item arasındaki benzerliği
hesaplar
            for i=1:size(activeUserItem,2)
                iindex=activeUserItem(1,i);
                sub1=0;sub2=0; sub2_1=0;sub2_2=0;
                corating=[];
                corating =find(A(:,iindex)ve A(:,targetItem)); %target item ile
activeuseritemin rate verdiği itemlara rate veren ortak userları bulur.
                for k=1:size(corating,1)
                    kindex=corating(k,1);
                    if (A(kindex,iindex)~=0 veve A(kindex,targetItem)~=0)
                        indexOfRatings = find(A(kindex,:)~=0);
                        ratings(1,1:size(indexOfRatings,2)) = A(kindex,indexOfRatings);
                        average=mean(ratings);
                        ratings=[];

```

Ek.-B. Item Base Matlab Algoritma (devam)

```

sub1=sub1 + (A(kindex,iindex)-average)*(A(kindex,targetItem)-average);
    sub2_1=sub2_1+ (A(kindex,iindex)-average)^2;
        sub2_2=sub2_2+ (A(kindex,targetItem)-average)^2;
    end
    sub2=sqrt(sub2_1*sub2_2);
    if sub2~=0
        itemsim(1,i)=sub1/sub2;
    else
        itemsim(1,i)=0;
    end
end
end
sim=itemsim;
%   targetItem = qp(itemid);
%   tmpitem=targetItem;
%
%   activeUser2=MLP(auidx,:);
%   activeUser2(auidx,targetItem)=0;
%   activeUserItem=find(activeUser2~=0);
%   save activeUserItem activeUserItem;
%   for i=1:size(activeUserItem,2)
%       sub1=0;sub2=0; sub2_1=0;sub2_2=0;
%       for k=1:size(A,1)
%           if (A(k,i)~=0 veve A(k,targetItem)~=0)
%               indexOfRatings = find(A(k,:)~=0);
%               ratings(1,1:size(indexOfRatings,2)) = A(k,indexOfRatings);
%               average=mean(ratings);
%               ratings=[];

```

Ek Açıklama – B: Item Base Matlab Algoritma (devam)

```

%      sub1=sub1 + (A(k,i)-average)*(A(k,targetItem)-average);
%      sub2_1=sub2_1+ (A(k,i)-average)^2;
%      sub2_2=sub2_2+ (A(k,targetItem)-average)^2;
%      end
%      sub2=sqrt(sub2_1*sub2_2);
%      if sub2~=0
%          itemsim(1,i)=sub1/sub2;
%      else
%          itemsim(1,i)=0;
%      end
%      end
%      end
%      sim=itemsim;
%      coRated=[];
%      %find avd ve standart sapma
tmpOrgActiveUser = MLP(auidx,:);
tmpOrgActiveUser(targetItem) = 0;
ratings = tmpOrgActiveUser(tmpOrgActiveUser~=0);
avg = mean(ratings);
stddev = std(ratings,1);
%      produce prediction kNN
[Y,I]=sort(sim,'descend');
%      if size(activeUserItem,2)>neighbors, limit=neighbors;
%      else limit=size(activeUserItem,2); end
%      sub1=0; sub2=0;
%      for k=1:limit
%          if Y(1,k)>0

```

Ek Açıklama – B: Item Base Matlab Algoritma (devam)

```

        sub1 = sub1 + MLP(auidx,I(k))*Y(k);
        sub2 = sub2 + Y(k);
    end
end
if sub2~=0
    Prp(auidx,itemid) = avg+stddev*(sub1/sub2);
else
    Prp(auidx,itemid) = avg;
end
if Prp(auidx,itemid)<1, Prp(auidx,itemid)=1;
elseif Prp(auidx,itemid)>4, Prp(auidx,itemid)=4;
end
save Prp Prp;
k=0;
MAE=0;
for au=1:size(MLP,1)
    for i=1:size(targetItemsList_push,1)
        if MLP(au,targetItemsList_push(i,1))~=0 veve ~isnan (Prp(au,i))
            MAE = MAE + abs(Prp(au,i)-MLP(au,targetItemsList_push(i,1)));
            k=k+1;
        end
    end
end
end
MAE = MAE / k%(20*size(targetItemsList,1));
save MAE MAE
save k k
save Prp Prp;

```

Ek Açıklama - C: Tekil Değer Ayırışımı (Svd) Algoritma

```

function SVD
clear all
clc
load iktisat;
MLP=iktisat;
%load R;
%fillerSize = [25];
%attackSize = 15;
%load MLP;
%segmentUsers = LocateSegmentUsers(dMLP);
for userid=1:size(MLP,1)
    indexOfRatings = find(MLP(userid, :)~=0);
    ratings(1,1:size(indexOfRatings,2)) = MLP(userid,indexOfRatings);
    average=mean(ratings);
    standarddeviation=std(ratings,1);
    rtd=find(MLP(userid,)==0);
    MLP(userid,rtd)=average;
end
load targetItemsList_push
save MLP MLP
%MLP=ders;
qp = targetItemsList_push';
a=MLP;
%for o=1:1
sdR=std(MLP);
meR=mean(sdR);
meR2=meR*meR; %meR2 is square of meR
an4=MLP;

```

Ek Açıklama - C: Tekil Değer Ayrışımı (Svd) Algoritma (devam)

```

g=size(an4,1); %total user number
an5=an4'*an4;
an6=(1/g)*an5;
d=rank(an6);
for i=1:size(a,2),
    for j=1:size(a,2),
        if i==j
            an6(i,j)=an5(i,j)-(meR2); %Exception takes place
        end
    end
end
save an6 an6
[Vd,Sd] = eig(an6);
S1=Sd;
for i=1:size(Sd,2)
    for j=1:size(Sd,2)
        if i==j
            S1(i,j)=Sd(size(Sd,2)+1-i,size(Sd,2)+1-j); %S1 has descending order
diagonal!
        end
    end
end
S2=sqrt(S1);
S3=S2(1:d,1:d); %S3 reduced to d version of S2
for i=1:size(S3,1)
    S4(i)=sqrt(g)*S3(i,i); %S3 is multiplied with square root of g!
end
V1=Vd;

```

Ek Açıklama - C: Tekil Değer Ayrışımı (Svd) Algoritma (devam)

```

    for i=1:size(Vd,2)
V1(:,i)=Vd(:,size(Vd,2)+1-i);% Column are taken symmetry like V[0 1 2] to V[2 1 0];
    end
    V2=V1(1:end,1:d); % V2 reduced to d version of V1
    Ud=[];
    for i=1:d
        Ud(:,i)=(1/S4(i))*an4*V2(:,i); %Ud is reduced to d initially
    end
    k=16;
    URd=Ud(1:end,1:k);
    SRd=S3(1:k,1:k);
    VRd=V2(1:end,1:k);
    SQd=sqrt(SRd);
    ad2d=URd*SQd;
    ad3d=SQd*VRd';
    PRd=ad2d*ad3d;
    %save P
    %Test items are used here!
    load targetItemsList_push;
    %load targetItemsList_nuke;
    titems=targetItemsList_push;
    %titemsN=targetItemsList_nuke;
    c1=repmat(titems,50,1);
    for i=1:size(MLP,1)
        standarddeviations(i)=ComputeStd(MLP(i,:));

        meanvalues(i)=ComputeMean(MLP(i,:));    %Calculate row mean and stand
dev.

        %standarddeviations(i)=std(k1(n));

```

Ek Açıklama - C: Tekil Değer Ayırışımı (Svd) Algoritma (devam)

```

end
for u=1:size(MLP,1)
    for itemid=1:size(qp,2)
        paq1d(u,itemid)=PRd(u,titems(itemid));
        paqd(u,itemid)=meanvalues(u)+standarddeviations(u)*paq1d(u,itemid);
        if paqd(u,itemid)>4, paqd(u,itemid)=4;
        elseif paqd(u,itemid)<1, paqd(u,itemid)=1;
        end
    end
end
end
save paqd paqd;
k=0;
MAE=0;
for au=1:size(MLP,1)
    for i=1:size(targetItemsList_push,1)
        if MLP(au,targetItemsList_push(i,1))~=0 veve ~isnan (paqd(au,i))
            MAE = MAE + abs(paqd(au,i)-MLP(au,targetItemsList_push(i,1)));
            k=k+1;
        end
    end
end
end
svdMAE = MAE / k%(20*size(targetItemsList,1));
save svdMAE svdMAE
save k k
%save Prp Prp;
return
end

```


ÖZGEÇMİŞ

Bülent Batmaz

İstatistik Bilgi Sistemleri

Eğitim

Yüksek Lisans	2002	Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Sayısal Yöntemler Anabilim Dalı
Lisans	1993	Anadolu Üniversitesi Fen Fakültesi İstatistik Bölümü
Lise	1986	Gaziantep Lisesi

İş

1995 --- Anadolu Üniversitesi Açıköğretim Fakültesi, Öğretim Görevlisi

Kişisel Bilgiler

Doğum Yeri ve Yılı: Gaziantep / 1968

Cinsiyet: Erkek

Yabancı Dil: İngilizce