



## Netflix Verileri Üzerinde TF-IDF Algoritması ve Kosinüs Benzerliği ile Bir İçerik Öneri Sistemi Uygulaması

**Özlem GELEMET**, İstanbul Gelişim Üniversitesi, Bilgisayar Mühendisliği Bölümü, Lisans Öğrencisi, ozlem\_gelemet@hotmail.com, 0000-0002-2443-2591

**Hakan AYDIN**, İstanbul Avcansaray Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Dr. Öğretim Üyesi, hakanaydin@ayvansaray.edu.tr, 0000-0002-0122-8512

**Ali ÇETİNKAYA**, İstanbul Gelişim Üniversitesi, Teknoloji Transfer Ofisi Uygulama ve Araştırma Merkezi, Araştırmacı, alcetinkaya@gelisim.edu.tr, 0000-0003-4535-3953

### ÖZ

Dijital platform kullanıcıları, bu platformların sunduğu özelleştirilmiş hizmetlerden yararlanmak ve bunları zaman ve mekan bağımsız olarak tüketmek istemektedirler. İnternet üzerinden yayım yapan bu platformlar arasında dünya çapında en yaygın olanlardan biri de Netflix'tir. Bu çalışmanın amacı TF-IDF (term frequency-inverse document frequency) algoritması ve Kosinüs Benzerliği (Cosine Similarity) ile Doğal Dil İşleme (NLP) ile Netflix kullanıcı verileri üzerinde bir içerik öneri sistemi uygulaması geliştirmektir. Bu bağlamda çalışmamızda yapılan analizler ile benzerlik yöntemleri ve uygun eşleşme verilerinin bulunması, böylelikle kullanıcılara kişisel bazda öneri yapılması hedeflenmiştir. Çalışma kapsamında hem Türkçe hem de diğer dillerdeki filmler ve diziler üzerinde farklı deneyler yapılmıştır. Yapılan deneyler neticesinde kosinüs benzerliği kullanılarak en yüksek benzerlik başarıları %91, en düşük benzerlik başarıları ise %43 olarak elde edilmiştir. Deneyler aynı veriler üzerinde kosinüs benzerliği ile birlikte TF-IDF algoritması ile yapıldığında ise başarı oranı %99 ile %80 arasında elde edilmiştir. Çalışma sonuçları, TF-IDF algoritması ve kosinüs benzerliği birlikte uygulanarak yapılan deneylerde, kosinüs benzerliği kullanılarak yapılan deneylere nazaran daha yüksek başarı oranının elde edildiğini ortaya koymaktadır. Çalışmamızın benzerlik yöntemleri ve uygun eşleşme verileri kullanılarak kişisel bazda öneri yapmayı hedefleyen içerik tabanlı öneri sistemi uygulamalarının geliştirilmesi bağlamında literatüre katkı sağlayacağı değerlendirilmektedir.

**Anahtar Kelimeler** : Yapay Zeka, Makine Öğrenmesi, Doğal Dil İşleme, Dil İşleyicileri, Veri Yönetimi



## A Content Recommendation System Application with TF-IDF Algorithm and Cosine Similarity on Netflix Data

### ABSTRACT

Digital platform users want to benefit from the customized services offered by these platforms and consume them regardless of time and place. Among these platforms that broadcast over the internet, one of the most common worldwide is Netflix. The purpose of this study is to develop a content recommendation system application on Netflix user data with TF-IDF (term frequency-inverse document frequency) algorithm and Natural Language Processing (NLP) with Cosine similarity. In this context, it is aimed to find similarity methods and suitable matching data with the analyzes made in our study, thus making suggestions to the users on a personal basis. Within the scope of the study, different experiments were carried out on films and TV series in both Turkish and other languages. As a result of the experiments, using cosine similarity, the highest similarity success was 91% and the lowest similarity success was 43%. When the experiments were performed with the TF-IDF algorithm on the same data, the success rate was between 99% and 80%. The results of the study reveal that a higher success rate is obtained in the experiments performed with the TF-IDF algorithm compared to the experiments using cosine similarity. It is considered that our study will contribute to the literature in the context of developing content-based recommendation system applications that aim to make suggestions on a personal basis using similarity methods and appropriate matching data.

**Keywords** : Artificial Intelligence, Machine Learning, Natural Language Processing, Language Handlers, Data Management

### EXTENDED ABSTRACT

Users of digital platforms want to benefit from the customized services offered by these platforms and consume them regardless of time and place. Among these platforms that broadcast over the Internet, Netflix is one of the most widely used worldwide. The objective of this study is to develop an application for a content recommendation system based on Netflix user data using the algorithm TF-IDF (Term Frequency-Inverse Document Frequency) and Natural Language Processing (NLP) with Cosine Similarity. In this context, we aim to find similarity methods and suitable matching data with the analyzes performed in our study to provide suggestions to users on a personal basis. In the study, various experiments were conducted with movies and TV series in both Turkish and other languages. In the cosine similarity experiments, the highest similarity success was 91% and the lowest was 43%. When the experiments with the TF-IDF algorithm were performed with the same data, the success rate ranged from 99% to 80%. The results of the study show that a higher success rate is obtained in the experiments with the TF-IDF algorithm than in the experiments with the cosine similarity. It is expected that our study will contribute to the literature related to the development of content-based recommender systems that aim to make suggestions on a personal basis using similarity methods and appropriate matching data.

## 1. GİRİŞ

Günümüzde dijitalleşme, yeni medya ortamları ve internet teknolojilerinin yoğun olarak kullanımı neticesinde dijital platformlarda kullanılan çevrimiçi içerik sitelerinin sayısının arttığı ve bu platformların yoğun olarak kullanıldığı görülmektedir. Özellikle içinde bulunduğumuz küresel COVID-19 pandemi sürecinde bu sitelerin kullanımı İnternet, Web 2.0 teknolojileri ve yeni medya teknolojileri temelinde giderek yaygınlaşmıştır. Ancak bu durum beraberinde bu sitelerde bir içerik yoğunluğuna neden olmuş ve kullanıcıları istenen bilgilere ulaşamama, zaman kaybı veya buna benzer sorunlarla karşı karşıya bırakmıştır. Bilginin talep edildiği anda kullanıcıların talep ettiği şekilde (metin, ses, görüntü, grafik vb.) ve talep edilen miktarda sağlanabilmesi önem arz etmektedir. Bu bağlamda karşımıza içerik öneri sistemleri çıkmaktadır. İçerik Yönetim Sistemleri kurumsal bilginin organize edilmesi, düzenlenmesi, denetlenmesi, arşivlenmesi, paylaşılmasını sağlayan sistemler olarak karşımıza çıkmaktadır. Günümüzde öneri sistemlerinin birçok farklı platformda sıkça kullanılmaya başlandığı görülmektedir. Öneri sistemleri kişilerin tercihleri ve zevklerine göre onlar için ilgilenebilecekleri farklı içerikler hazırlayarak kullanıcı değerlendirmesini en iyi hale getirmeyi hedefler. Öneri sistemleri, internetteki büyük verileri filtrelemek ve kullanıcıların en çok ilgileneceği bilgileri önlerine çıkarmak amacıyla kullanılır (Yücebaş, 2019). Büyük veriler üzerinde bilgi içeriklerine göre anlamsal sınıflandırma ile çalışmalar yapılabilmektedir (Yumusak vd., 2018). İçerik öneri sistemleri, özellikle çok fazla kullanıcıya sahip platformlarda kişisel olarak çıkarım yapmak için çok büyük faydalar sağlamaktadır. Öneri sistemlerinde farklı teknik ve yöntemler ile veriler filtrelenebilmekte ve sıralanabilmektedirler (Ahmad, 2017). Öneri sistemlerinde üç farklı yaklaşımdan söz edilebilir: Kullanıcı tabanlı yaklaşım, öge tabanlı yaklaşımı, hibrit öneri yaklaşım. Kullanıcı tabanlı yaklaşımda benzer müşteriler eşleştirilerek öneri sistemi buna göre çalıştırılır. Öge tabanlı yaklaşımda ise ürün benzerliğine dayalı işlem yapmaktadır. Hibrit yöntem ise diğer iki yaklaşımın karışımı olarak çalışan bir yaklaşımdır. Hibrit yaklaşım çok amaçlı optimizasyona dayalı karma bir öneri modeli sunar (Cai vd., 2020). Öneri sistemleri, arama motorlarının aksine kullanıcıların aradığını bulmaya çalışmaktan ziyade, onların isteyebileceği öğeleri bulmayı amaçlayan sistemlerdir (Wang vd., 2021). Öneri sistemleri, Web üzerindeki platformlar üzerinden izleyicilerin yalnızca kimlik bilgilerini değil, duygu ve deneyimlerini de veriye dönüştürmek suretiyle kullanıcıların beğenecekleri film önerileri sunarak müşteri memnuniyetini en üst düzeye taşımayı amaçlar (Monti vd., 2021 & Gasparetti vd., 2021, Zhang vd., 2021 & Thomas ve John, 2021 & Felfernig vd. 2021). Öneri sistemlerinde kullanıcılar belli bir ücret karşılığında içeriklere erişim hakkı kazanırlar ve kullanıcılar bu sistemlerden ilgi alanlarına göre filtreleme yapabilirler (Bennett ve Lanning, 2007). Günümüzde "Puhutv", "BluTV" gibi Yerel Dijital Platformlar yanında, ayrıca "Hulu", "DC Universe", "Amazon Prime Video", "YouTube Premium", "CBS All Access", "Netflix" gibi daha pek çok Küresel Dijital Platformun kullanıldığı görülmektedir. Gelecekte dijital yayıncılığın yaygınlaşmasıyla birlikte bunlar ve benzeri dijital yayıncılık platformları da çeşitlenebilecektir. İnternet üzerinden yayın yapan bu platformlar arasında dünya çapında

en yaygın olanlardan biri de Netflix'tir. Üyelik sistemine dayanan çalışma prensibi ile üyeliklerin niteliği doğrultusunda kullanıcılar bu platform bünyesinde yer alan dizi ve filmleri izleyebilmektedirler. Bu sitenin içerikleri bulunulan ülkeye göre değişiklik gösterebilmektedir. Bu sitenin ayrıca oldukça büyük bir izleyici kitlesine sahip olduğu görülmektedir. Çocuklar, gençler ve yetişkinler gibi pek çok farklı kullanıcı kitlesine yönelik olarak içerik üreten Netflix sitesinin, günümüzde geniş bir kullanıcı kitlesine ulaşabilme gücüne ve yeteneğine sahip bir site olduğu görülmektedir.

Günümüzde kullanıcılar dijital platformlarda kullanılan çevrimiçi içerik sitelerinde bulunan içeriklerde arama yapmak ve bunları zaman ve mekan bağımsız olarak tüketmek istemektedirler. Bu bağlamda içinde bulunduğumuz Bilgi Çağında başta Yapay Zeka (YZ), Makine Öğrenmesi (ML) ve Doğal Dil İşleme (NLP) olmak üzere pek çok yöntem ve tekniklerden yararlanmak mümkündür. YZ uygulamalarının bir alt kategorisi olan NLP disiplini bilgisayar ile insan arasındaki etkileşimi dil üzerinden çözmeye odaklanır. NLP'de, ML uygulamalarının yardımıyla, metinler parçalara ayrılabilen, sınıflandırılabilen, duygu analizi yapılabilen, varlık ismi tanımlanabilmektedir. ML, bulunduğu ortamdan öğrenerek insan zekasını taklit etmek amacıyla tasarlanmış, gelişmekte olan bir hesaplama algoritmaları dalıdır (El Naqa ve Murphy, 2015 & Akay, 2020). ML algoritmaları özellikle büyük veriler için oldukça iyi sonuçlar verebilmektedir (Başer vd., 2021). Doküman benzerliği makinelerin dokümanlar arası bağlamları ayırt edip kelimeleri o bağlamlar içinde kullanması açısından önemli bir husustur. Bu çalışmanın konularından bir tanesi TF-IDF (term frequency–inverse document frequency) algoritması ve diğeri de kosinüs benzerliğidir. TF-IDF, terim ağırlıklandırma ölçümü kelimelerin metinler içinde geçme sıklığı bilgisine dayalıdır. Bir doküman TF-IDF ölçümü ile vektörler ile ifade edilebilir. Böylelikle bir doküman bilgisayarlar tarafından anlaşılabilir şekilde sayılar ile ifade edilebilir. Bu değer bir sözcüğün bulunduğu dokümanı ne kadar temsil ettiğini gösteren bir istatistiksel değer olarak ifade edilebilir. Kosinüs Benzerliği'nde eğer metinler birer vektör (yöney, vector) olarak düşünülürse iki vektörün birbirine göre olan ilişkisi bir açı ile ifade edilmektedir. Kosinüs Benzerliği, metinler arasındaki benzerliği vektörel olarak ölçmektedir. Metinlerde geçen kelimelerin metinde kaç kez geçtiği hesaplanır. Daha sonra her metin içerdiği kelimelerle 1 ve 0 şeklinde vektörel olarak ifade edilir. Her metin üç boyutlu uzayda vektörel olarak yerleştirildiğinde aralarındaki kosinüs açısı ne kadar küçük ise metinler birbirlerine o kadar yakındır. Tamamen birbiri ile ilişkisiz olan vektörler için ise kosinüs değeri 0 olurken tamamen birbirini zıddı olan dokümanlar için kosinüs değeri -1 olacaktır.

Bu çalışmanın amacı Netflix kullanıcı verileri üzerinde TF-IDF (term frequency–inverse document frequency) algoritması ve kosinüs benzerliği ile kullanıcıların yaptığı kişisel tercihlere göre aratılan, izlenen ya da beğenilen içeriklere göre yüksek doğruluk oranı ile yeni içerikler önerebilen bir içerik öneri sistemi uygulaması geliştirmektir. Çalışmada Kaggle

üzerinde yayınlanan ve 8807 adet veri içeren “Netflix Movie and Titles” isimli veri seti kullanılmıştır. Çalışma kapsamında hem Türkçe hem de diğer dillerdeki filmler üzerinde deneyler yapılmıştır. Çalışmada yapılan analizler ile benzerlik yöntemleri ve uygun eşleşme verilerinin bulunması, bu sayede kişisel bazda öneri yapılması hedeflenmiştir. Python programlama dilinde “seaborn”, “matplotlib” ve “wordcloud” gibi görselleştirme kütüphaneleri kullanılarak analizler yapılmıştır. Çalışma kapsamında gerçekleştirilen deneylerden elde edilen sonuçlar veri görselleştirme kullanılarak açıklanmıştır.

Bu çalışmanın diğer bölümleri şu şekilde düzenlenmiştir: Çalışmanın ikinci bölümünde NLP ve TF-IDF hakkında literatürde yer alan çalışmalar anlatılmıştır. Üçüncü bölümde öncelikle NLP, TF-IDF, ML, Kosinüs Benzerliği hakkında bilgiler verilmiştir. Daha sonra bu çalışma kapsamında izlenen yol ve gerçekleştirilen işlemler bir akış diyagramı kapsamında anlatılmıştır. Çalışmanın dördüncü bölümünde çalışma kapsamında gerçekleştirilen deneysel çalışmalara yer verilmiştir. Çalışmanın son bölümünde çalışma kapsamında elde edilen sonuçlar paylaşılmış ve gelecek çalışmalara değinilmiştir.

## 2. LİTERATÜR İNCELEMESİ

Literatürde NLP kapsamında TF-IDF ile yapılmış çeşitli çalışmalar mevcuttur. (Ahmed, 2017 & Korkmaz, 2021) tarafından yapılan çalışmalarda, haberlerin gerçekliğinin tespitinde TF-IDF yöntemi kullanıldığı, bu bağlamda TF-IDF ile haber içeriklerindeki kelimelerin ağırlıklandırıldığı görülmektedir. Literatürde özellikle duygu analizi maksadıyla TF-IDF ile yapılmış farklı çalışmalar olduğu da görülmektedir (Martineau ve Finin, 2009 & Paltoglou ve Thelwall, 2010 & Mee vd., 2021). Bu çalışmalarda duyguları temsil eden kelimelere dikkat edilerek gerekli ağırlıklandırmaların yapıldığı görülmektedir. (Sjarif vd., 2019) tarafından yapılan çalışmada TF-IDF ve Random Forest (RF) algoritması kullanılarak istenmeyen mesajların filtrelenmesinin yapıldığı görülmektedir. Literatürde Yapay Sinir Ağları (YSA) kullanılarak TF-IDF ile yapılmış farklı çalışmalar olduğu da görülmektedir (Marcinčuk vd., 2021 & Chaipornkaew ve Banditwattanawong, 2021 & Thakkar ve Chaudhari, 2020). (Öztürk vd., 2020) tarafından yapılan çalışmada NLP ve alınan metinsel veriler için TF-IDF kullanılmıştır. Literatürde TF-IDF ile temel kelimelere ağırlıklar verilerek sınıflandırmaya dayalı farklı çalışmalar da mevcuttur (Mohammed ve Omar, 2020 & Jiang vd., 2021 & Cahyani ve Patasik, 2021). (Albayrak, 2020) tarafından yapılan çalışmada lisansüstü seviyede açılan düşünülen disiplinler arası bir dersin içeriğinin hazırlanması için veri madenciliği tekniklerinden doğal dil işleme yöntemleri kullanılmıştır. Literatür taramamızda bilgi güvenliği konusunda da TF-IDF ile yapılmış çalışmaların olduğu tespit edilmiştir (Al-Rimy vd., 2020 & Jalilifard vd., 2021 & Rani ve Bidhan, 2021).

Literatürde kosinüs benzerliği kapsamında yapılmış çeşitli çalışmalar da mevcuttur. (Beyaz ve Yaylı, 2021) tarafından yapılmış çalışmada röntgen verisinden çıkarım yaparak hastalara teşhis konulması araştırılmıştır. (Önden ve Çallı, 2019) tarafından yapılan çalışmada sosyal medya üzerinde yapılan inceleme ile yenilikçi ürünlerin tüketicilerin tutumlarına

etkileri araştırılmıştır. Literatür taramamızda yüz tespiti konusunda kosinüs benzerliği ile yapılmış çalışmaların olduğu tespit edilmiştir (Nguyen ve Bai, 2010 & Zheng vd., 2015 & Chen, 2021). Bu çalışmalarda yüz verileri arasındaki vektör uzaklığının hesaplandığı, bu hesaplama sonuçlarının yüz doğrulama sistemleri bağlamında kosinüs benzerliği ile değerlendirildiği görülmektedir. (Meltem ve Çamurcu, 2011) tarafından yapılan çalışmada İnternette bulunan web sayfalarında bulunan dokümanlara erişmek veya ekrana getirmek konusunun araştırıldığı, belgeleri kümeleme tekniğinin kullanıldığı ve belgeler kümelenirken kosinüs uzaklıklarının kullanıldığı görülmektedir. (Polat ve Körpe, 2018) tarafından yapılan çalışmada kosinüs benzerliği ile vektör uzayında belgelerin yakınlıklarının tespit edilmesinin araştırıldığı görülmektedir.

### 3. MATERYAL ve YÖNTEMLER

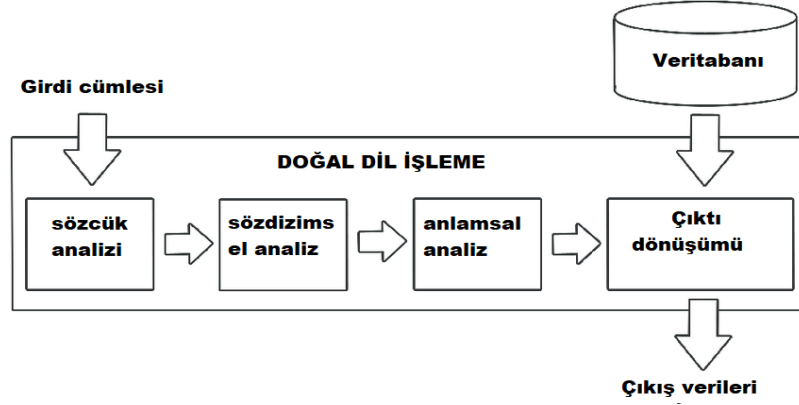
#### 3.1. Doğal Dil İşleme (NLP)

Doğal dil işleme, insan dilinin YZ ve bilgisayar biliminin kesişimi olarak ifade edilebilir (Şekil 1). Doğal dil işleme, insan dilini yapay zeka teknikleri ile bilgisayar bilimlerini kullanarak çalışan bir yöntemdir.



Şekil 1: Doğal Dil İşleme Yapısı

Doğal dil işleme ise günlük konuşma dilinin bilgisayar tarafından algılanarak üzerinde çeşitli işlemler yapılması anlamına gelmektedir. Bu işlemlere algılanan dilden çıkarımlar yapılması, yeni cümleler üretilebilmesi ve farklı varlıklarla bağlantı kurulabilmesi örnek verilebilir. Bağlantı kurulan varlıklar bir kişi, kurum vb. olarak düşünülebilir. Genel olarak doğal dil işleme yapısının çalışma mantığı Şekil 2’de verilmiştir (Bulut, 2020).



Şekil 2: Doğal Dil İşlemenin Çalışma Mantığı

### 3.2. TF-IDF Algoritması

TF-IDF (Terim frekansı-ters belge frekansı) algoritması bir kelimenin doküman içerisindeki önemini gösteren ağırlıklandırmaları yapmak için kullanılan bir yöntemdir. Bu yöntemde TF-IDF değeri TF ve IDF değerlerinin ayrı ayrı hesaplanarak birbirleri ile çarpılmasıyla hesaplanır. TF değeri kelimenin ilgili dokümandaki tekrar sayısının dokümandaki toplam sayısına oranı ile hesaplanır. IDF değeri ise veri setindeki toplam doküman sayısının, seçilen kelimenin geçtiği toplam doküman sayısına oranının logaritma değeri olarak hesaplanır. Sonuç olarak hesaplanan TF ve IDF değeri birbiri ile çarpılarak seçilen kelimenin ağırlık değeri hesaplanır. Denklem (1)'de kullanılan  $i$  değeri cümle içerisindeki bir kelimeyi,  $d$  ise belgeyi temsil etmektedir. Bu değerler kullanılarak öncelikle TF değeri elde edilmektedir.

$$w_{i,d} = t_{fi,d} \times \log(n/d_{fi}) \quad (1)$$

Belgede geçen terim sayısının, en fazla geçen değer sayısına oranı Denklem (2)'de belirtilmiştir.

$$t_{fi,d} = \frac{f_{ri,d}}{d_{fi}} \quad (2)$$

Denklem (3)'te belge içerisinde değerin kaç kez tekrarlandığı belirtilmektedir.

$$f_{t,d} \quad (3)$$

Bu değer kullanılarak Denklem (4)'te ki normalizasyon işlemi yapılmaktadır.

$$\log(1 + f_{t,d}) \quad (4)$$

Denklem (5)'te ağırlıklandırma uygulanan dokümanlara 0,5 ile 1 arasında değerler çıktı olarak verilir. Ayrıca hesaplanmış olan ham frekanslar için standardizasyon işlemi yapılmaktadır.

$$0.5 + 0.5 \frac{f_{t,d}}{\max f_{t,d}} \quad (5)$$

Denklem (6)'da ise TF değeri için en büyük ölçeklendirme hesaplanmaktadır.

$$K + (1 - K) ft, d \max ft, d \quad (6)$$

### 3.3. Makine Öğrenmesi (ML)

ML, bir yandan insan zekasına yaklaşılmaya çalışırken, bir diğer yandan insanların vereceği komutlara ihtiyaç duymayan algoritmaları içerir. ML, temel olarak eğitilecek yeterli sayıda veriyi kullanarak, bu veriden anlamlı sonuçlar çıkartmayı hedefler.



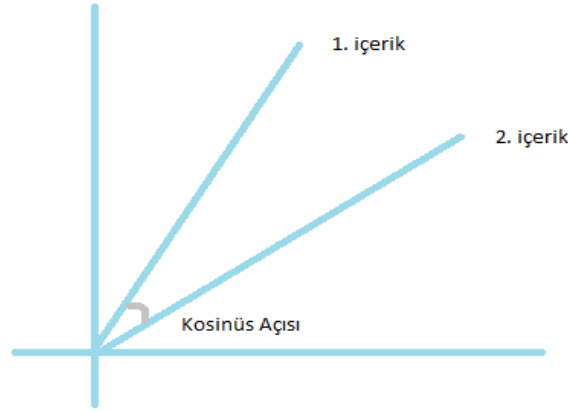
Şekil 3: Derin Öğrenme ve Makine Öğrenmesi İlişkisi

Şekil 3'teki şemada ML'nin DL ile olan ilişkisi gösterilmiştir (Kılıç, 2021). ML'de verilerin çıktı değerlerinin, hangi özelliklerin olması veya olmaması durumunda gerçekleştiğini tespit ederek bir tahminleme yapılması durumu söz konusudur. Veri seti ne kadar büyük ve içerisindeki örneklerin ayırt edici özelliği ne kadar fazla ise ML algoritmalarının başarısı o derece artmaktadır. ML teknik ve yöntemleri derin öğrenmeyi de içine almaktadır. Derin öğrenme (Deep Learning, DL) henüz üzerinde işlem yapılmamış verileri ifade eden bir yapıdır (Gürsakar, 2017).

### 3.4. Kosinüs Benzerliği

Kosinüs Benzerliği metinler arasındaki vektörel uzaklığı ölçerek benzerliğin bulunmasını temel almaktadır. İçerikler birer vektör olarak düşünüldüğünde Kosinüs Benzerliği iki vektörün çarpımının iki vektörün boylarının çarpımına oranı olarak hesaplanır (Amin ve Garg, 2019 & Carrasco ve Rosillo, 2021 & Xia vd., 2015 & Tata ve Patel, 2007 & Liao vd., 2021). Şekil 4 üzerinde A ve B içerikleri arasında oluşan açının görüntüsü yer almaktadır. Söz konusu şekilde belirtilen kosinüs açısının boyutuna göre benzerlik değeri elde edilir.

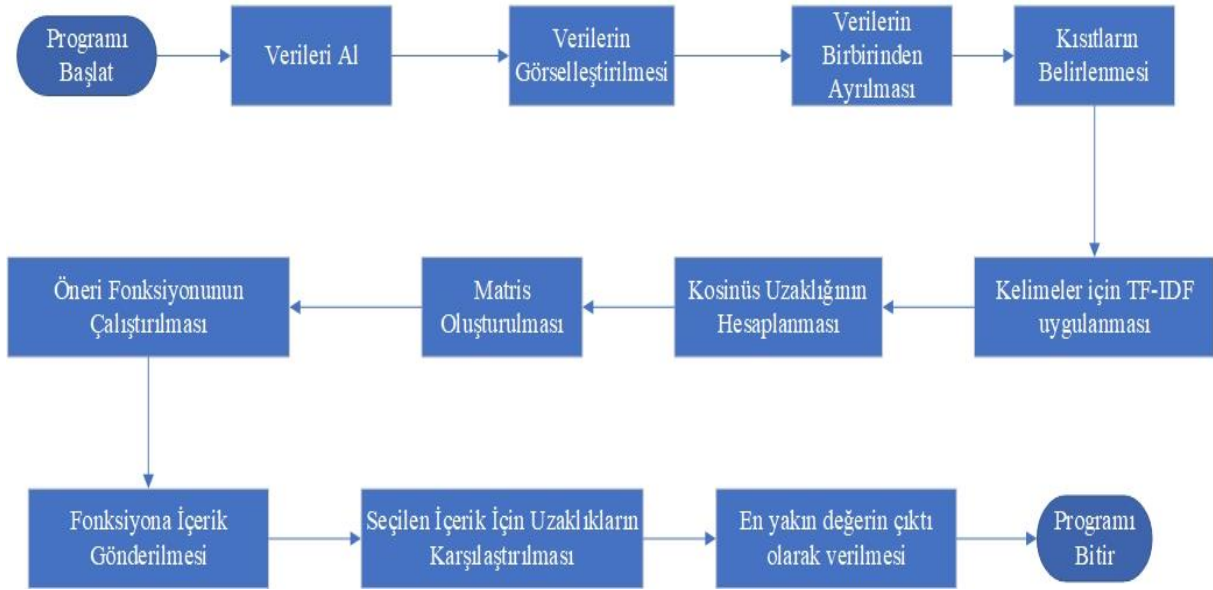




Şekil 4: Kosinüs Benzerliği Grafiği

### 3.5. Sistemin Akış Şeması

Çalışmamız Şekil 5 üzerinde gösterilen akış diyagramını dikkate alınarak gerçekleştirilmiştir. Söz konusu akış diyagramında gösterildiği gibi çalışmamızın ilk aşamasını veri setinin çalışmaya dahil edilmesi oluşturmaktadır. Veri setinin daha iyi analiz edilebilmesi amacıyla veri görselleştirme işlemlerinden de faydalanılmıştır. Çalışmamızda bir içerik öneri sistemi gerçekleştirilmesi amaçlandığından veri setinde yer alan verilerden özellik seçimi yapılmıştır. Çalışmada hem TF-IDF algoritmasından hem de kosinüs benzerliğinden yararlanılmıştır. Çalışmada Python programlama dili kullanılmıştır. Bu programlama dilinin seçilme nedeni kütüphanelerinin açık kaynak kodlu olmasıdır (Kumaş, 2021). Python dili dünya üzerinde çok yaygın kullanılmaktadır. Çalışmada "nltk" kütüphanesinden de yararlanılmıştır. NLTK (Natural Language Toolkit - Doğal Dil Araç Seti), İngilizce dili için geliştirilen bir doğal dil işleme kütüphanesidir. Bu kütüphane Steven Bird ve Edward Loper tarafından Pennsylvania Üniversitesi'nde geliştirilmiştir. NLTK metinleri kolayca işleyebilir, çıkarımlar yapabilir, etiketlendirebilir. Kelimelerden anlam çıkarma, kelimelerin kökünü bulma gibi işlemler NLTK kütüphanesi ile kolayca gerçekleştirilebilir (NLTK, 2021). Bu kütüphane kelimelerin birbirinden ayrılması amacıyla kullanılmıştır. Çalışma kapsamında veri görselleştirme ve matematiksel işlemler için "numpy" kütüphanesi, grafikler için "seaborn" kütüphanesi ve veri seti üzerinde çalışabilmek için de "pandas" kütüphaneleri kullanılmıştır.

**Şekil 5:** Sistemin Akış Diyagramı

## 4. DENEYSEL ÇALIŞMALAR

### 4.1. Veri Seti İşlemleri

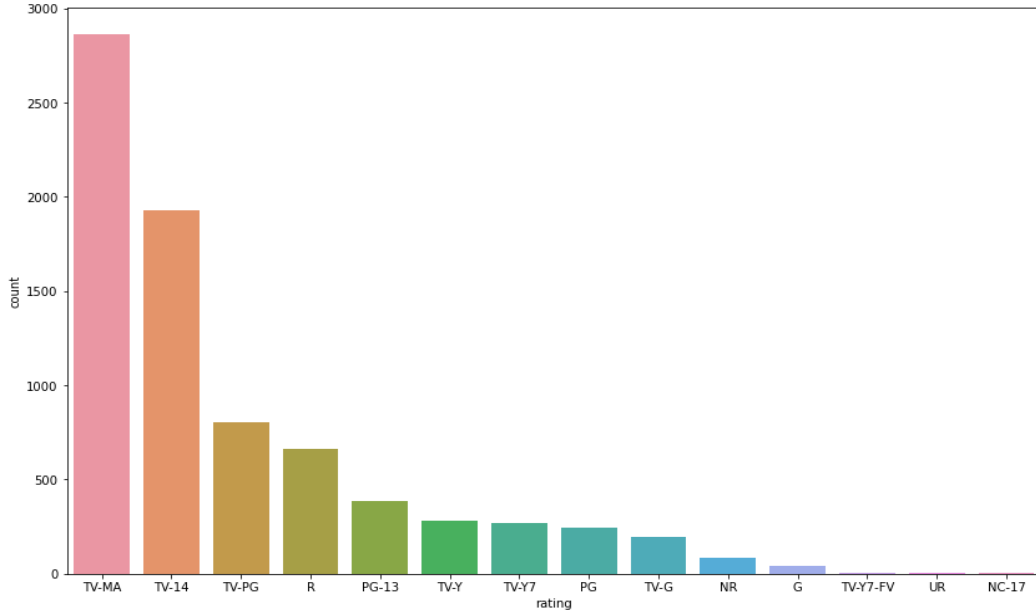
Çalışma kapsamında 2019 yılında Kaggle platformunda yayınlanan veri seti kullanılmıştır (Bansal, 2021). Bu veri seti içerisinde yapım yılları 1925 ve 2020 yılları arasında olan 8807 adet benzersiz içerik bulunmaktadır. Veri seti içerdiği geniş tarih aralığı ve yüksek veri sayısı bakımından diğer veri setlerinden ayrılmaktadır. Bu veri seti içerisinde 11 Adet giriş verisi mevcuttur. Bu çalışma içerisinde veri seti üzerinde çalışmanın başarısını arttırabilmek için 3 adet giriş verisi seçilmiştir. Bu veri seti Netflix arama motoru olan Flixable'dan alınmış olan bir veri setidir. Veri seti içerisinde her içerik için eşsiz bir id değeri bulunur. Bu veri setinde film veya dizi olma durumu, bunların isimleri, yapım ülkeleri, reyting türleri, içerik oyuncularları, içeriğin türü, yayınlanma tarihi, içeriğin türüne göre yayınlanma süreleri ve özet bilgileri yer almaktadır. Tablo 1 üzerinde bu çalışma kapsamında kullanılan veri setinde yer alan özellikler sunulmuştur.

**Tablo 1:** Veri Seti Özellikleri

	show_id	type	title	cast	country	date_added	release_year	rating	duration	listed_in	description
0	81145628	Movie	Norm of the North: King Sized Adventure	Alan Marriott, Andrew Toth,	United States, India, South	September 9, 2019	2019	TV-PG	90 min	Children & Family Movies,	Bedore Planning ....

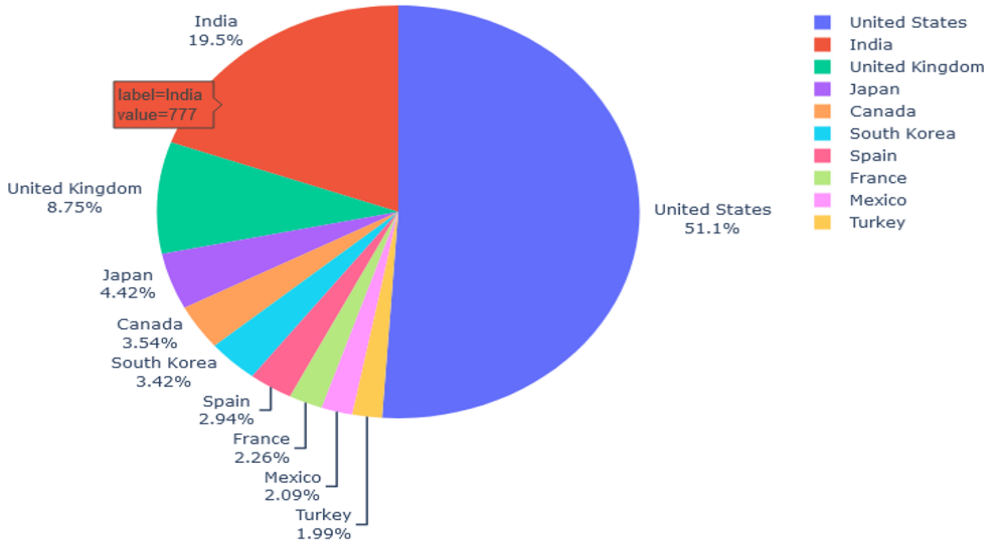
				Brian Dobson,...	Korea, China					Comedies	
1	80117401	Movie	Jandino: Whatever it Takes	Jandino Asporaat	United Kingdom	September 9, 2016	2016	TV-MA	94 min	Stand-Up Comedy	Jandino Asporaat ...
2	70234439	TV Show	Transformers Prime	Peter Cullen, Sumalee Montano, Frank Welker, ...	United States	September 8, 2018	2013	TV-Y7-FV	1 Season	Kids' TV	With the help ....
3	80058654	TV Show	Transformers : Robots in Disguise	Will Friedle, Darren Criss, Constance Zimmer, ...	United States	September 8, 2018	2016	TV-Y7	1 Season	Kids' TV	When a prison ...

Şekil 6'da çalışmamız kapsamında hazırladığımız ve izlenme oranlarını gösteren grafik yer almaktadır.



Şekil 6: İzlenme Oranları Grafiği

Şekil 7'de çalışmamız kapsamında hazırladığımız ve ülkelere göre içerik yoğunluğunu gösteren grafik gösterilmektedir.



Şekil 7: Ülkelere Göre İçerik Yoğunluğu

Çalışmamızda kelime görselleştirmesi için “wordcloud” kütüphanesinden yararlanılmıştır. Bu kütüphanenin çalışma içerisinde kullanımı Şekil 8 üzerinde gösterilmiştir.



Şekil 8: Worldcloud Kütüphanesinin Akış Şeması

```
from wordcloud import WordCloud
plt.subplots(figsize=(25,15))
wordcloud = WordCloud(
    background_color='black',
    width=1920,
    height=1080
).generate(" ".join(ds1.title))
plt.imshow(wordcloud)
plt.axis('off')
plt.savefig('cast.png')
plt.show()
```

Algoritma 1: Worldcloud Sözde (Psoduiodo) Kod Yapısı

Wordcloud kütüphanesinin kullanım amacı veri setinde bulunan içeriklerin isimleri içerisinde geçen kelimelerin yoğunluğunu bulmaktır. Wordcloud eklenerek öncelikle oluşacak görsel fiziksel özellikler verilmiştir. Bu özellikler boyut bilgisi, arkaplan rengi gibi özelliklerdir. Sonrasında wordcloud kütüphanesinin bir metodu olan generate ile oluşturulacak şekilde görselin boyutları belirtilmiştir. Wordcloud nesnesine arkaplan rengi verilerek, yükseklik ve genişliği belirtilmiştir. "Join" ile veri setinin hangi kolonunda işlem yapılacağı belirtilerek o kolon kelimeleri üzerinde gösterimi yapılmıştır. Şekil 9'da çalışmamız kapsamında hazırladığımız ve içeriklerde bulunan kelimelerin yoğunluk haritasını gösteren görsel yer almaktadır.



Şekil 9: İçeriklerde Bulunan Kelimelerin Yoğunluk Haritası

#### 4.2. Kosinüs Benzerliği ile Gerçekleştirilen Deneysel Çalışmalar

Kosinüs benzerliğini kullanarak yaptığımız bu deneylerde 12 farklı örnek kullanılmıştır. Bu deneylere ilişkin seçilen filmler ve deneylerin sonuçlarına ilişkin hususlar aşağıdadır:

- Çalışmanın 1'inci deneyinde "Solo Berserk: The Golden Age Arc II - The Battle for Doldrey" filmi seçilmiş ve kosinüs benzerlik skoru hesaplanmıştır. Kosinüs benzerlik metodu kullanıldıktan sonra en benzer film "Berserk: The Golden Age Arc I - The Egg of the King" olarak hesaplanmıştır. "Berserk: The Golden Age Arc I - The Egg of the King" filminin seçilen film ile benzerlik skoru 0.917663'tür. Bu değer filmlerin birbirlerine çok yakın olduklarını, oyuncu kadrolarının, türlerinin neredeyse aynı olduğunu göstermektedir.
- Çalışmanın 2'nci deneyinde seçilen "Düğün Dernek" filmi için öneri sistemi çalıştırılmıştır. Bu deneyde seçilen filmin devam filmi olan "Düğün Dernek 2: Sünnet" filmi önerilmiştir. Benzerlik skoru 0.916667 olarak hesaplanmıştır. Oyuncu kadrosu, yapım ülkesi ve türü neredeyse aynı olan bu iki film için hesaplanan benzerlik skoru yüksek ölçülmüştür.
- Çalışmanın 3'üncü deneyinde "Eyyvah Eyyvah" filmi kullanılmıştır. Bu film için yine serinin devam filmi olan "Eyyvah Eyyvah 2" filmi önerilmiştir. Benzerlik skoru 0.886405 olarak hesaplanmıştır. Oyuncu kadrosu, yapım ülkesi ve türü neredeyse aynı olan bu iki film için hesaplanan benzerlik skoru başarılıdır.
- Çalışmanın 4'üncü deneyinde "Ancient Aliens" filmi seçilmiştir. Burada oyuncu kadrosundan daha çok yapım ülkesi ve tür bilgileri benzerlik skorunu artırmakta etkili

olmuştur. Önerilen film “The Curso Oak Island”, benzerlik skoru ise 0.866025 olarak hesaplanmıştır. Yine benzerlik skoruna göre oldukça benzer bir filmidir.

- Çalışmanın 5’inci deneyinde “Gilmore Girls” filmi kullanılmıştır. Önerilen film “Gilmore Girls: A Year in the Life” filmidir. Yine bir devam filmi önerilmiştir, ancak buradaki fark oyuncu kadrosunda olan değişikliklerden kaynaklı olarak Deney-1 ve Deney-2’ ye göre benzerlik skoru düşüş göstermiştir.
- Çalışmanın 6’ncı deneyinde “Stranger Things” içeriği seçilmiştir. ABD yapımı olan ve bilim kurgu türündeki bu içerik için “Beyond Stranger Thing” adlı içerik önerilmiştir. Benzerlik skoru 0.741941 olarak hesaplanmıştır. Hesaplanan değer benzerlik kısıtlarının iki içerik baz alındığında oldukça yakın olduğunu göstermektedir. İçeriklerin türü farklı olsa da oyuncu kadrosu ve yapım ülkesi benzerliği yüksektir. Bu da skorun başarısını olumlu yönde etkilemiştir.
- Çalışmanın 7’nci deneyinde seçilen “Monster High: Boo York, Boo York” içeriği için “Monster High: Frights, Camera, Action!” içeriği önerilmiştir. Aralarındaki benzerlik skoru ise 0.585369 olarak hesaplanmıştır. Bu skor iki içerik kısıtlarına bakıldığında tür ve yapım ülkelerinin çok benzer olduğunu ancak oyuncu kadrosunda değişiklikler olduğunu göstermektedir.
- Çalışmanın 8’inci deneyinde “Goldie & Bear” içeriği için “Hatchimals | Adventures in Hatchtopia” içeriği önerilmiştir. Benzerlik skoru 0.547723’ tür. İki içerik de animasyon türünde ve ABD yapımıdır. Bu kısıtlar bakımından oldukça benzerdir ancak oyuncu kadrosu farklarından dolayı benzerlik skorunda düşüş görülmüştür.
- Çalışmanın 9’uncu deneyinde “MINDHUNTER” dizisi seçilmiştir. Bu dizi için veri setinden “The Sinner” dizisi önerilmiştir. İki dizi de gerilim türünde ve ABD yapımıdır. Bu açılarından birbirine benzer olan bu iki içerik oyuncu kadroları karşılaştırıldığında farklı oyunculara sahip içerikler olduğu görülmüştür. “MINDHUNTER” dizisi ve önerilen dizi olan “The Sinner’ın” benzerlik oranı 0.500000’dır. tür ve yapım ülkesi açısından aynı olan bu iki içerik oyuncu kadrosu açısından birbirlerinden ayrılmaktadır. Bu da benzerlik skorunun düşmesine sebep olmuştur.
- Çalışmanın 10’uncu deneyinde seçilen “When the Camellia Blooms” dizisi için 0.490290 benzerlik skoru ile “Something in the Rain” dizisi önerilmiştir. Bu iki dizi yapım ülkesi bakımından aynıdır ancak tür bakımından ve oyuncu kadrosu olarak farklıdır. Burada iki kısıtta yaşanan farklılıklar nedeniyle benzerlik skorunda düşüş yaşanmıştır.
- Çalışmanın 11’inci deneyinde “Friends” dizisi seçilmiştir. Öneri olarak “Episodes” dizisi verilmiştir. İki dizi arasındaki benzerlik skoru 0.476731 olarak hesaplanmıştır. Yapım ülkesi ve tür bakımından çok benzer olan bu iki dizi oyuncu kadroları

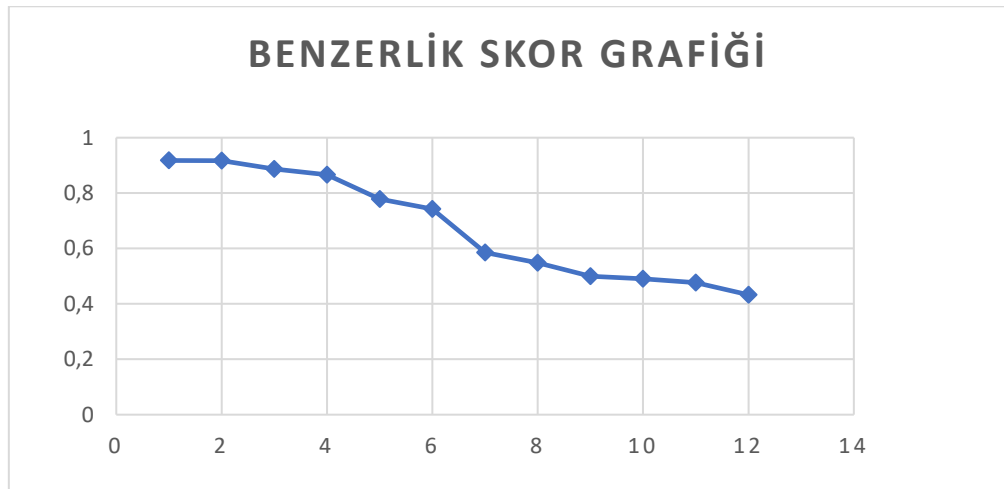
bakımından yüksek benzerliğe sahip değildir. İki dizi için sadece bir oyuncunun benzer olduğu görülmüştür. Bu durum başarıyı etkilemiştir.

- Çalışmanın 12'inci deneyinde "13 Reasons Why" dizisi için "13 Reasons Why: Beyond the Reasons" adlı içerik önerilmiştir. Bu iki içerik için oyuncu kadroları neredeyse aynıdır, ancak tür bakımından farklılık göstermektedir. Önerilen içerik "13 Reasons Why" içeriğinin belgeselidir. İki içerik için benzerlik skoru 0.432789'dur.

Tablo 2'de kosinüs benzerliği yöntemi kullanılarak gerçekleştirilen ve yukarıda anlatılan 12 adet farklı deneyin sonuçları yer almaktadır.

**Tablo 2:** Kosinüs Benzerliği ile Yapılan Deneyler

No	Seçilen İçerik Adı	Önerilen İçerik Adı	Benzerlik Skoru
1	Berserk: The Golden Age Arc II - The Battle for Doldrey	Berserk: The Golden Age Arc I - The Egg of the King	0.917663
2	Düğün Dernek	Düğün Dernek 2: Sünnet	0.916667
3	Eyyvah Eyyvah	Eyyvah Eyyvah 2	0.886405
4	Ancient Aliens	The Curso Oak Island	0.866025
5	Gilmore Girls	Gilmore Girls: A Year in the Life	0.777778
6	Stranger Things	Beyond Stranger Things	0.741941
7	Monster High: Boo York, Boo York	Monster High: Frights, Camera, Action!	0.585369
8	Goldie & Bear	Hatchimals   Adventures in Hatchtopia	0.547723
9	MINDHUNTER	The Sinner	0.500000
10	When the Camellia Blooms	Something in the Rain	0.490290
11	Friends	Episodes	0.476731
12	13 Reasons Why	13 Reasons Why: Beyond the Reasons	0.432789



**Şekil 10:** Kosinüs Benzerliği Deneyleri Benzerlik Skor Grafiği

Şekil 10'da kosinüs benzerliği yöntemi ile gerçekleştirilen deneylerin benzerlik skor grafiği yer almaktadır. Yapılan deneyler neticesinde kosinüs benzerliği kullanılarak en yüksek benzerlik başarısı %91, en düşük benzerlik başarısı ise %43 olarak elde edilmiştir.

#### **4.3. TF-IDF Yöntemi ve Kosinüs Benzerliği ile Gerçekleştirilen Deneysel Çalışmalar**

TF-IDF yöntemini kosinüs benzerliği ile birlikte kullanarak yaptığımız deneylerde 12 farklı örnek kullanılmıştır. Bu deneylere ilişkin seçilen filmler ve deneylerin sonuçlarına ilişkin hususlar aşağıdadır:

- Seçilmiş olan 1 numaralı film “Solo Berserk: The Golden Age Arc II - The Battle for Doldrey” için tf-idf kullanılarak ve kosinüs benzerlik skoru hesaplanarak başarılı bir öneri yapılmaya çalışılmıştır. Tf-idf uygulanmış olan verilere daha sonra kosinüs benzerlik metodu eklenerek en benzer film “Berserk: The Golden Age Arc I - The Egg of the King” olarak hesaplanmıştır. “Berserk: The Golden Age Arc I - The Egg of the King” filminin seçilen film ile benzerlik skoru 0.986486’dır. Bu değer filmlerin birbirlerine çok yakın olduklarını, oyuncu kadrolarının, türlerinin neredeyse aynı olduğunu göstermektedir.
- 2 numaralı deneyde ise seçilen “Düğün Dernek” filmi için öneri sistemi çalıştırılmıştır. Bu deneyde seçilen filmin devam filmi olan “Düğün Dernek 2: Sünnet” filmi önerilmiştir. Benzerlik skoru 0.994937 olarak hesaplanmıştır. Oyuncu kadrosu ve türü neredeyse aynı olan bu iki film için hesaplanan benzerlik skoru yüksek ölçülmüştür.
- 3 numaralı deney ise “Eyyvah Eyyvah” filmidir. Bu film için yine benzer oyuncu kadrosuna sahip bir film olan “Locked on You” filmi önerilmiştir. Benzerlik skoru 0.997701 olarak hesaplanmıştır. Oyuncu kadrosu ve türü neredeyse aynı olan bu iki film için hesaplanan benzerlik skoru yüksek başarıya sahiptir.
- 4 numaralı deneyde “Ancient Aliens” filmi seçilmiştir. Burada oyuncu kadrosundan ve yapım ülkesi bilgileri benzerlik skorunu artırmakta etkili olmuştur. Önerilen film “The Curso Oak Island”, benzerlik skoru ise 0.998625 olarak hesaplanmıştır. Yine benzerlik skoruna göre oldukça benzer bir filmidir.
- 5. deneyde “Gilmore Girls” filmi kullanılmıştır. Önerilen film “Gilmore Girls: A Year” in the Life filmidir. Yine bir devam filmi önerilmiştir ancak buradaki fark oyuncu kadrosunda olan değişikliklerden kaynaklı olarak Deney-1 ve Deney-2’ ye göre benzerlik skoru düşüş göstermiştir. 0.971061 olarak ölçülen değer yine yüksek bir başarıya sahiptir.
- 6. deneyde “Stranger Things” içeriği seçilmiştir. ABD yapımı olan ve bilim kurgu türündeki bu içerik için “Beyond Stranger Thing” adlı içerik önerilmiştir. Benzerlik skoru 0.940772 olarak hesaplanmıştır. Hesaplanan değer benzerlik kısıtlarının iki içerik



baz alındığında oldukça yakın olduğunu göstermektedir. İçeriklerin oyuncu kadrosu ve yapım ülkesi benzerliği yüksektir. Bu da skorun başarısını olumlu yönde etkilemiştir.

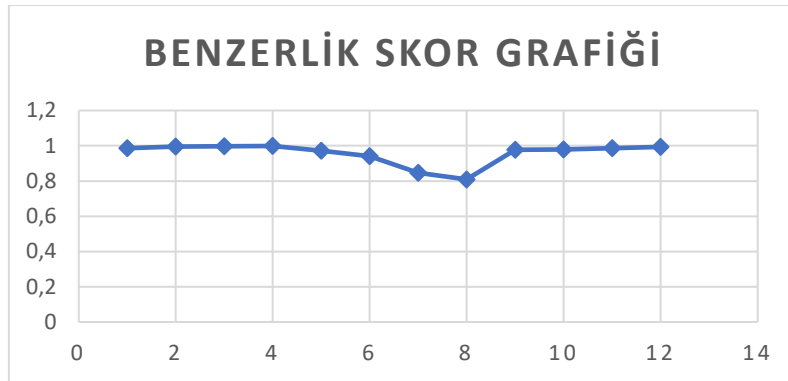
- 7. deneyde seçilen “Monster High: Boo York, Boo York” içeriği için “Monster High: Frights, Camera, Action!” içeriği önerilmiştir. Aralarındaki benzerlik skoru ise 0.846154 olarak hesaplanmıştır. Bu skor iki içerik kısıtlarına bakıldığında yapım ülkelerinin çok benzer olduğunu ancak oyuncu kadrosunda değişiklikler olduğunu göstermektedir. Bu değişiklik sebebiyle başarıda düşüş yaşanmıştır.
- 8 numaralı deney olan “Goldie & Bear” içeriği için “Scooby-Doo!: Mystery Incorporated” içeriği önerilmiştir. Benzerlik skoru 0.80904’tür. İki içerik de animasyon türünde ve ABD yapımıdır. Bu bakımdan oldukça benzerdir ancak oyuncu kadrosu farklarından dolayı benzerlik skorunda düşüş görülmüştür.
- 9. deneyde “MINDHUNTER” dizisi seçilmiştir. Bu dizi için veri setinden “Dave Chappelle: Sticks & Stones” dizisi önerilmiştir. İki dizi de gerilim türünde ve ABD yapımıdır. Bu açılarından birbirine benzer olan bu iki içerik oyuncu kadroları karşılaştırıldığında farklı oyunculara sahip içerikler olduğu görülmüştür. “MINDHUNTER” ve önerilen dizi olan “Dave Chappelle: Sticks & Stones’un” benzerlik oranı 0.977805’tir. Tür ve yapım ülkesi açısından aynı olan bu iki içerik oyuncu kadrosu açısından birbirlerinden ayrılmaktadır. Bu da benzerlik skorunun düşmesine sebep olmuştur.
- 10 numaralı deneyde seçilen “When the Camellia Blooms” dizisi için 0.978138 benzerlik skoru ile “Rugrats in Paris: The Movie” içeriği önerilmiştir. Bu iki içerik yapım ülkesi bakımından aynıdır ancak oyuncu kadrosu olarak farklıdır. Burada iki kısıtta yaşanan farklılıklar nedeniyle benzerlik skorunda düşüş yaşanmıştır.
- 11. deneyde “Friends” dizisi seçilmiştir. Öneri olarak “Man with a Plan” dizisi verilmiştir. İki dizi arasındaki benzerlik skoru 0.986053 olarak hesaplanmıştır. Yapım ülkesi bakımından çok benzer olan bu iki dizi oyuncu kadroları bakımından da yüksek benzerliğe sahiptir. İki dizi için sadece bir oyuncunun benzer olduğu görülmüştür. Bu durum başarıyı etkilemiştir.
- 12. deneyde “13 Reasons Why” dizisi için “13 Reasons Why: Beyond the Reasons” adlı içerik önerilmiştir. Bu iki içerik için oyuncu kadroları neredeyse aynıdır. Önerilen içerik “13 Reasons Why” içeriğinin belgeselidir. İki içerik için benzerlik skoru 0.99322’dir.

Tablo 3 üzerinde TF-IDF ve kosinüs benzerliği birlikte kullanarak gerçekleştirilen ve yukarıda anlatılan 12 farklı deneyin sonuçları verilmiştir.

**Tablo 3:** TF-IDF ve Kosinüs Benzerliği ile Yapılan Deneyler

Deney No	Seçilen İçerik Adı	Önerilen İçerik Adı	Benzerlik Skoru
1	Berserk: The Golden Age Arc II - The Battle for Doldrey	Berserk: The Golden Age Arc III - The Advent	0.986486
2	Düğün Dernek	Düğün Dernek 2: Sünnet	0.994937
3	Eyyvah Eyyvah	Locked On You	0.997701
4	Ancient Aliens	The Curso Oak Island	0.998625
5	Gilmore Girls	Gilmore Girls: A Year in the Life	0.971061
6	Stranger Things	Yes, God, Yes	0.940772
7	Monster High: Boo York, Boo York	Monster High: Frights, Camera, Action!	0.846154
8	Goldie & Bear	Scooby-Doo!: Mystery Incorporated	0.80904
9	Mindhunter	Dave Chappelle: Sticks & Stones	0.977805
10	When the Camellia Blooms	Rugrats in Paris: The Movie	0.978138
11	Friends	Man with a Plan	0.986053
12	13 Reasons Why	13 Reasons Why: Beyond the Reasons	0.99322

Şekil 11'de TF-IDF yöntemi ile kosinüs benzerliğinin birlikte kullanılarak gerçekleştirildiği deneylerin benzerlik skor grafiği yer almaktadır.



**Şekil 11:** TF-IDF ile Kosinüs Benzerliği Deneyleri Benzerlik Skor Grafiği

Çalışma kapsamında TF-IDF algoritması ve kosinüs benzerliği yöntemlerinin birlikte kullanılarak yapıldığı deneyler neticesinde başarı oranı %99 ile %80 arasında elde edilmiştir.

## 5. SONUÇ

Günümüzde dijital yayıncılık platformlarının içerik yoğunluğunun artması durumu bu platformların kullanıcılarını talep edilen içeriği bulamama problemi ile karşı karşıya bırakmıştır. İnternet üzerinden yayın yapan bu platformlar arasında dünya çapında en yaygın olanlardan birisi de Netflix'tir. Çalışmamızda TF-IDF algoritması ve kosinüs benzerliği ile Netflix kullanıcı verileri üzerinde bir içerik öneri sistemi uygulaması geliştirilmiştir. Çalışmamız kapsamında hem Türkçe hem de diğer dillerdeki filmler ve diziler üzerinde farklı deneyler yapılmıştır. Yapılan deneyler neticesinde kosinüs benzerliği kullanılarak en yüksek

benzerlik başarısı %91, en düşük benzerlik başarısı ise %43 olarak elde edilmiştir. Deneyler aynı veriler üzerinde kosinüs benzerliği ile birlikte TF-IDF algoritmasının kullanılması ile başarı oranı %99 ile %80 arasında elde edilmiştir. Çalışma sonuçları, TF-IDF algoritması ile yapılan deneylerde, kosinüs benzerliği kullanılarak yapılan deneylere nazaran daha yüksek başarı oranının elde edildiğini ortaya koymaktadır. Benzerlik ölçütlerinde kosinüs benzerliği en iyi performansı sağlayan benzerliklerden birisidir. Ancak çalışmamız kapsamında gerçekleştirilen deneylerin sonuçlarına göre kosinüs benzerliği ile birlikte TF-IDF algoritması kullanılmasının, kosinüs benzerlik ölçütünün tek başına kullanılmasına nazaran daha yüksek başarı oranı elde edilmiştir. TF-IDF terim ağırlıklandırma ölçümü kelimelerin metinler içinde geçme sıklığı bilgisine dayalıdır. Terim ağırlıklandırma, metin sınıflandırmada sonuçlar üzerinde doğrudan etkili olan önemli bir adımdır. Elde edilen sonuçlar NLP çalışmalarında TF-IDF'in kullanım etkinliğini ortaya koymaktadır. Günümüzde TF-IDF'in, sorgu kelimesi ile en alakalı sonuçları sunmak için arama motorları tarafından kullanılması da bu durumu destekler niteliktedir. Çalışmamızın benzerlik yöntemleri ve uygun eşleşme verileri kullanılarak kişisel bazda öneri yapmayı hedefleyen içerik tabanlı öneri sistemi uygulamalarının geliştirilmesi bağlamında literatüre katkı sağlayacağı değerlendirilmektedir.

Gelecek çalışmalarımızda, uygulamamızın mevcut özelliklerine ilave olarak kullanıcı yorumlarını ve oylamalarını da çalışmamıza eklemeyi, kullandığımız veri setinde bulunan "tanım (description)" kolonunun daha da zenginleştirerek kosinüs benzerliği ve TF-IDF yöntemlerinin ayrıca bu kolon üzerinde de uygulamayı, veri setimizdeki örnek sayılarını güncel verilerle daha da zenginleştirmeyi ve böylelikle uygulamamızı daha da geliştirmeyi planlıyoruz.

## KAYNAKÇA

- Ahmad, S. (2017). Survey on recommendation system using data mining and clustering techniques. *International Journal for Research in Engineering Application and Management (IJREAM)*, 3(9), 68-72.
- Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments* (pp. 127-138). Springer, Cham.
- Akay, E. C., Soydan, N. T. Y., & Gacar, B. K. (2020). Makine öğrenmesi ve ekonomi: bibliyometrik analiz. *PressAcademia Procedia*, 12(1), 104-105.
- Albayrak, A. (2020). Doğal Dil İşleme Teknikleri Kullanılarak Disiplinler Arası Lisansüstü Ders İçeriği Hazırlanması. *Bilişim Teknolojileri Dergisi*, 13(4), 373-383.
- Al-Rimy, B. A. S., Maarof, M. A., Alazab, M., Alsolami, F., Shaid, S. Z. M., Ghaleb, F. A., ... & Ali, A. M. (2020). A pseudo feedback-based annotated TF-IDF technique for dynamic crypto-ransomware pre-encryption boundary delineation and features extraction. *IEEE Access*, 8, 140586-140598.
- Amin, D. M., & Garg, A. (2019). Performance Analysis of Data Mining Algorithms. *Journal of Computational and Theoretical Nanoscience*, 16(9), 3849-3853.
- Bansal S., (2021). Netflix Movies and TV Shows, <https://www.kaggle.com/shivamb/netflix-shows>
- Başer, B. Ö., Yangın, M., Sarıdaş, E. S. (2021). Makine Öğrenmesi Teknikleriyle Diyabet Hastalığının Sınıflandırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 25(1), 112-120.
- Bennett, J., & Lanning, S. (2007). The netflix prize. In *Proceedings of KDD cup and workshop* (Vol. 2007, p. 35).
- Beyaz, S., & Yaylı, Ş. B. (2021). Ortopedi ve Travmatolojide Yapay Zeka Uygulamaları: Artificial Intelligence Applications in Orthopaedics & Traumatology. *Sağlık Bilimlerinde Yapay Zeka Dergisi (Journal of Artificial Intelligence in Health Sciences)*, 1(1), 12-15.
- Bulut E. (2020, Kasım 18). Doğal dil işleme nedir? Doğal dil işleme kursları vb. <https://turkiyeyapayzeka.com/dogal-dil-isleme/dogal-dil-isleme-nedir-dogal-dil-isleme-kurslari-vb/>
- Cahyani, D. E., & Patasik, I. (2021). Performance comparison of TF-IDF and Word2Vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, 10(5), 2780-2788.
- Cai, X., Hu, Z., Zhao, P., Zhang, W., & Chen, J. (2020). A hybrid recommendation system with many-objective evolutionary algorithm. *Expert Systems with Applications*, 159, 113648.
- Carrasco, S. S., & Rosillo, R. C. (2021). Word embeddings, cosine similarity and deep learning for identification of professions & occupations in health-related social media. In *Proceedings of the Sixth Social Media Mining for Health (# SMM4H) Workshop and Shared Task* (pp. 74-76).
- Chaipornkaew, P., & Banditwattanawong, T. (2021). A recommendation model based on user behaviors on commercial websites using TF-IDF, KMeans, and Apriori algorithms. In *International Conference on Computing and Information Technology* (pp. 55-65). Springer, Cham.
- Chen, J., Guo, Z., & Hu, J. (2021). Ring-Regularized Cosine Similarity Learning for Fine-Grained Face Verification. *Pattern Recognition Letters*, 148, 68-74.

- El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In *machine learning in radiation oncology* (pp. 3-11). Springer, Cham.
- Felfernig, A., Tran, T. N. T., & Le, V. M. (2021). Recommender Systems Beyond E-Commerce: Presence and Future. In *Consumer Happiness: Multiple Perspectives* (pp. 203-230). Springer, Singapore.
- Gasparetti, F., Sansonetti, G., & Micarelli, A. (2021). Community detection in social recommender systems: a survey. *Applied Intelligence*, 51(6), 3975-3995.
- Gürsakal, N. (2017). *Makine öğrenmesi ve derin öğrenme*. Dora Basım Yayın Dağıtım, Bursa.
- Jalilifard, A., Caridá, V. F., Mansano, A. F., Cristo, R. S., & da Fonseca, F. P. C. (2021). Semantic sensitive TF-IDF to determine word relevance in documents. In *Advances in Computing and Network Communications* (pp. 327-337). Springer, Singapore.
- Jiang, Z., Gao, B., He, Y., Han, Y., Doyle, P., & Zhu, Q. (2021). Text classification using novel term weighting scheme-based improved TF-IDF for Internet media reports. *Mathematical Problems in Engineering*, 2021.
- Kılıç, I. (2021). Sağlık Bilimlerinde Yapay Zeka Ve Makine Öğrenmesi. <https://healthintechno.com/2021/02/22/saglik-bilimlerinde-yapay-zeka-ve-makine-ogrenmesi/>
- Korkmaz, T., Çetinkaya, A., Aydın, H., & Barışkan, M. A. (2021). Analysis of whether news on the Internet is real or fake by using deep learning methods and the TF-IDF algorithm. *International Advanced Researches and Engineering Journal*, 5(1), 31-41.
- Kumaş, E. (2021). Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 2(2), 1-5.
- Liao, J., Huang, Y., Wang, H., & Li, M. (2021, June). Matching Ontologies with Word2Vec Model Based on Cosine Similarity. In *The International Conference on Artificial Intelligence and Computer Vision* (pp. 367-374). Springer, Cham.
- Marcinčuk, M., Gniewkowski, M., Walkowiak, T., & Będkowski, M. (2021). Text document clustering: Wordnet vs. TF-IDF vs. word embeddings. In *Proceedings of the 11th Global Wordnet Conference* (pp. 207-214).
- Martineau, J. C., & Finin, T. (2009). Delta tfidf: An improved feature space for sentiment analysis. In *Third international AAAI conference on weblogs and social media*.
- Mee, A., Homapour, E., Chiclana, F., & Engel, O. (2021). Sentiment analysis using TF-IDF weighting of UK MPs' tweets on Brexit. *Knowledge-Based Systems*, 107238.
- Meltem, I., & Çamurcu, A. Y. (2011). Web belgeleri kümelemede benzerlik ve uzaklık ölçütleri başarılarının karşılaştırılması. *Marmara Fen Bilimleri Dergisi*, 20(1), 35-49.
- Mohammed, M., & Omar, N. (2020). Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. *PloS one*, 15(3), e0230442.
- Monti, D., Rizzo, G., & Morisio, M. (2021). A systematic literature review of multicriteria recommender systems. *Artificial Intelligence Review*, 54, 427-468.
- Nguyen, H. V., & Bai, L. (2010, November). Cosine similarity metric learning for face verification. In *Asian conference on computer vision* (pp. 709-720). Springer, Berlin, Heidelberg.
- NLTK, (2021). *Natural Language Toolkit*. <https://www.nltk.org/>
- Önden, A., & Çallı, M. K. (2019). İnovatif Ürün Özelliklerinin Tüketicilerin Tutumları Üzerindeki Etkileri: Sosyal Medya Paylaşımları Üzerine Bir Çalışma. *AJIT-e: Bilişim Teknolojileri Online Dergisi*, 10(36), 53-72.

- Öztürk, A., Durak, Ü., & Badıllı, F. (2020). Twitter verilerinden doğal dil işleme ve makine öğrenmesi ile hastalık tespiti. *Konya Mühendislik Bilimleri Dergisi*, 8(4), 839-852.
- Paltoglou, G., & Thelwall, M. (2010). A study of information retrieval weighting schemes for sentiment analysis. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 1386-1395).
- Polat, H., & Körpe, M. (2018). TBMM Genel Kurul Tutanaklarından Yakın Anlamalı Kavramların Çıkarılması. *Bilişim Teknolojileri Dergisi*, 11(3), 235-244.
- Rani, U., & Bidhan, K. (2021). Comparative assessment of extractive summarization: textrank tf-idf and lda. *Journal of Scientific Research*, 65(1), 304-311.
- Sjarif, N. N. A., Azmi, N. F. M., Chuprat, S., Sarkan, H. M., Yahya, Y., & Sam, S. M. (2019). SMS spam message detection using term frequency-inverse document frequency and random forest algorithm. *Procedia Computer Science*, 161, 509-515.
- Tata, S., & Patel, J. M. (2007). Estimating the selectivity of tf-idf based cosine similarity predicates. *ACM Sigmod Record*, 36(2), 7-12.
- Thakkar, A., & Chaudhari, K. (2020). Predicting stock trend using an integrated term frequency-inverse document frequency-based feature weight matrix with neural networks. *Applied Soft Computing*, 96, 106684.
- Thomas, B., & John, A. K. (2021, February). Machine Learning Techniques for Recommender Systems–A Comparative Case Analysis. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1085, No. 1, p. 012011). IOP Publishing.
- Wang, S., Cao, L., Wang, Y., Sheng, Q. Z., Orgun, M. A., & Lian, D. (2021). A survey on session-based recommender systems. *ACM Computing Surveys (CSUR)*, 54(7), 1-38.
- Xia, P., Zhang, L., & Li, F. (2015). Learning similarity with cosine similarity ensemble. *Information Sciences*, 307, 39-52.
- Yücebaşı, S. C. (2019). MovieANN: A Hybrid Approach to Movie Recommender Systems Using Multi Layer Artificial Neural Networks. *Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 5(2), 214-232.
- Yumusak, S., Dogdu, E., & Kodaz, H. (2018). Classification of Linked Data Sources Using Semantic Scoring. *Leice Transactions on Information and Systems*, 101(1), 99-107.
- Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & Intelligent Systems*, 7(1), 439-457.
- Zheng, L., Idrissi, K., Garcia, C., Duffner, S., & Baskurt, A. (2015, May). Triangular similarity metric learning for face verification. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Vol. 1, pp. 1-7). IEEE.