YAPAY ZEKÂ (BM403) UYGULAMA PROJESİ

Derin Anlamsal Gömme (SBERT) ve Çok Kriterli Hibrit Filtreleme ile Akademik Makale Öneri Sistemi Geliştirilmesi

HAZIRLAYAN: *Edanur DEMİREL*

TC: *27334444102*

DANIŞMAN: *Rukiye POLATTİMUR*

TARİH: *2025*

Öz

Bu çalışmanın temel amacı, akademik literatürde giderek artan bilgi aşırı yüklemesi (information overload) problemini hafifletmek ve araştırmacıların ilgi alanlarına en uygun bilimsel yayınlara hızlı ve etkili biçimde erişimini sağlayan bir öneri sistemi geliştirmektir. Bu amaç doğrultusunda, Cornell Üniversitesi tarafından sağlanan arXiv veri setinden seçilen 12.000 makalelik bir alt küme kullanılmıştır. Çalışmada, makale başlık ve özetlerinin anlamsal temsili için derin öğrenme tabanlı Sentence-BERT (SBERT) modeli kullanılarak metinler yoğun vektör uzayına aktarılmış, benzerlik hesaplamaları kosinüs benzerliği metriği ile gerçekleştirilmiştir. Ayrıca, içerik tabanlı benzerlik skorları ile kullanıcı ilgi profillerini birleştiren hibrit bir öneri mimarisi tasarlanmıştır. Deneysel sonuçlar, SBERT tabanlı modelin, kelime eşleşmesine dayalı geleneksel TF-IDF yaklaşımına kıyasla metinler arası anlamsal ilişkileri daha başarılı biçimde yakaladığını göstermektedir. Geliştirilen hibrit sistem ise, kullanıcı ilgi alanlarına özgü ve tutarlı öneriler sunarak literatür tarama sürecini hızlandırmakta ve akademik verimliliği artıran etkili bir karar destek mekanizması olarak işlev görmektedir.

Anahtar Kelimeler: *Akademik makale öneri sistemi, Sentence-BERT (SBERT), TF-IDF, hibrit öneri sistemi, içerik tabanlı filtreleme*

Abstract

The primary objective of this study is to mitigate the growing problem of information overload in academic literature by developing an intelligent recommendation system that facilitates researchers' access to semantically relevant scientific publications. To this end, a corpus of 12,000 academic articles was constructed from the arXiv dataset maintained by Cornell University. In the proposed framework, article titles and abstracts are encoded into dense semantic representations using the deep learning–based Sentence-BERT (SBERT) model, and document similarities are computed via the cosine similarity metric. Furthermore, a hybrid recommendation architecture is introduced by integrating content-based similarity scores with user interest profiles. Experimental findings indicate that the SBERT-based model consistently outperforms the traditional TF-IDF approach, which relies on lexical matching, in capturing semantic relationships between scholarly documents. In addition, the proposed hybrid system provides personalized and coherent recommendations, thereby accelerating the literature review process and functioning as an effective decision support tool that enhances academic productivity.

Keywords: *Academic paper recommendation system, Sentence-BERT (SBERT), TF-IDF, hybrid recommender system, content-based filtering*

İçindekiler

1. GİRİŞ

Bilimsel bilgi üretiminin dijitalleşmesi ve akademik yayıncılığın küresel ölçekte hız kazanmasıyla birlikte, literatürdeki belge sayısı logaritmik bir artış göstermektedir. De Solla Price'ın erken dönemde öngördüğü gibi, bilimsel yayınların sayısı üstel bir büyüme trendindedir ve günümüzde yalnızca Google Scholar gibi platformlar 160 milyondan fazla akademik belgeyi indekslemektedir [1]. Benzer şekilde, açık erişimli akademik arşivlerden biri olan arXiv platformuna aylık olarak gönderilen makale sayısı binlerle ifade edilmektedir [2]. Bu hızlı büyüme, araştırmacıların kendi çalışma alanlarıyla en alakalı ve güncel yayınlara erişimini zorlaştırmakta ve literatürde yaygın olarak "bilgi aşırı yüklemesi" (information overload) problemi olarak tanımlanan durumu ortaya çıkarmaktadır.

Özellikle disiplinlerarası alanlarda çalışan araştırmacılar için literatür takibi, yalnızca zaman alıcı bir süreç olmakla kalmamakta, aynı zamanda yüksek düzeyde bilişsel yük gerektiren karmaşık bir problem hâline gelmektedir. Bu bağlamda akademik makale öneri sistemleri, kullanıcıların geçmiş okuma tercihleri, ilgi alanları ve içerik benzerliklerini dikkate alarak milyonlarca aday makale arasından en ilgili olanları filtreleyen ve araştırmacılara karar destek sunan kritik sistemler olarak öne çıkmaktadır [3].

1.1. Problem Tanımı ve Motivasyon

Günümüzde kullanılan akademik arama motorlarının ve öneri sistemlerinin büyük bir bölümü, Terim Frekansı–Ters Doküman Frekansı (TF-IDF) gibi istatistiksel anahtar kelime eşleştirme yöntemlerine dayanmaktadır. Ancak bu yaklaşımlar, kelimelerin bağlamsal anlamını göz ardı eden "kelime çantası" (bag-of-words) varsayımı nedeniyle anlamsal derinliği yakalamakta yetersiz kalmaktadır [4]. Örneğin, "Derin Öğrenme" anahtar kelimesiyle yapılan bir aramada, bu ifadeyi birebir içermeyen ancak "Yapay Sinir Ağları" veya "Evrişimsel Ağlar" temelli çalışmaları ele alan son derece ilgili makaleler sistem tarafından göz ardı edilebilmektedir. Bu durum, literatür taraması sırasında kritik çalışmaların kaçırılmasına yol açmaktadır.

Bu yetersizlik, araştırmacıların doğru bilgiye erişimini zorlaştırmakta ve akademik üretkenliği olumsuz yönde etkilemektedir. Dolayısıyla, kelime eşleşmesine dayalı yöntemlerin ötesine geçen, metinlerin anlamsal bağlamını kavrayabilen daha gelişmiş yaklaşımlara duyulan ihtiyaç giderek artmaktadır.

1.2. Mevcut Yaklaşımlar ve Sınırlılıkları

Akademik öneri sistemlerinde yaygın olarak kullanılan bir diğer yaklaşım ise işbirlikçi filtreleme (collaborative filtering) yöntemleridir. Bu yöntemler, "benzer kullanıcılar benzer içerikleri tercih eder" varsayımına dayanarak, kullanıcıların geçmiş etkileşimleri üzerinden öneriler üretmektedir [5].

Ancak akademik veri setlerinde sıkça karşılaşılan veri seyrekliği (sparsity) problemi ve sisteme yeni eklenen kullanıcılar veya makaleler için yeterli etkileşim

verisinin bulunmaması, bu yöntemlerin etkinliğini ciddi biçimde sınırlandırmaktadır. Bu durum literatürde "soğuk başlangıç" (cold-start) problemi olarak tanımlanmaktadır [6].

İşbirlikçi filtreleme tabanlı sistemler içerikten bağımsız çalıştıkları için, makalelerin başlık ve özet bilgilerini doğrudan analiz etmezler [7]. Dolayısıyla, yeni yayımlanan ancak henüz yeterli kullanıcı etkileşimi almamış bilimsel çalışmaların önerilmesi mümkün olmamaktadır [8]. Bu durum, akademik öneri sistemlerinde içerik temelli yaklaşımların önemini daha da artırmaktadır.

İçerik tabanlı filtreleme yöntemleri ise, kullanıcıların geçmişte ilgilendiği öğelerin içerik özelliklerine benzer öğeleri önermeye dayanmaktadır [3]. Bu yaklaşımda kullanıcı profilleri, bireyin ilgi alanlarını temsil eden içerik özellikleri üzerinden oluşturulmaktadır. Ancak akademik veri tabanlarında her zaman tam metne erişim mümkün olmadığından, literatürde makale başlığı ve özet bilgilerinin kullanılması yaygın bir uygulama hâline gelmiştir [9].

1.3. Çalışmanın Amacı

Bu çalışmanın temel amacı, akademik makalelerin anlamsal derinliğini kavrayabilen içerik tabanlı yaklaşımlar ile kullanıcı etkileşimlerini analiz eden yöntemleri bir araya getiren, derin öğrenme destekli hibrit bir akademik makale öneri sistemi geliştirmektir. Bu kapsamda, doğal dil işleme (NLP) alanında devrim yaratan Transformer mimarisi temel alınmış; makalelerin başlık ve özet bilgileri Sentence-BERT (SBERT) modeli kullanılarak yüksek boyutlu anlamsal vektör uzayına (embeddings) dönüştürülmüştür [10].

Geleneksel anahtar kelime eşleştirme yöntemlerinin aksine SBERT modeli, metinleri kelime kelime ele almak yerine, cümle ve paragraf düzeyinde bütüncül bir bağlam içerisinde analiz ederek anlamsal benzerlikleri yakalayabilmektedir. Bu sayede, kelime çakışması bulunmayan ancak içerik açısından ilişkili olan akademik çalışmalar arasında daha tutarlı benzerlik ilişkileri kurulabilmektedir.

Geliştirilen öneri sistemi, SBERT tabanlı bu anlamsal analiz yeteneğini; kullanıcının geçmiş tercihlerini temsil eden profil vektörleri ve makalelerin popülarite skorları ile birleştirerek "Çok Kriterli Hibrit Filtreleme" yaklaşımı çerçevesinde sunmaktadır. Bu hibrit mimari sayesinde, işbirlikçi filtreleme yöntemlerinin en önemli sınırlılıklarından biri olan ve yeni eklenen makalelerin yeterli kullanıcı etkileşimi bulunmadığı için önerilememesi şeklinde ortaya çıkan soğuk başlangıç problemi, içerik tabanlı analiz yoluyla büyük ölçüde aşılmaktadır.

Sonuç olarak sistem, kullanıcılara yalnızca kişisel ilgi alanlarına uygun değil; aynı zamanda akademik literatürde öne çıkan ve güncel eğilimleri yansıtan zengin bir okuma listesi sunmaktadır. Hesaplanan çok kriterli benzerlik skorlarına göre makaleler sıralanmakta ve en yüksek benzerliğe sahip ilk beş makale kullanıcıya öneri olarak sunulmaktadır.

1.4. Çalışmanın Katkıları

Bu çalışmanın literatüre ve uygulama alanına sağladığı başlıca katkılar aşağıda özetlenmiştir:

Akademik metinlerin temsilinde TF-IDF gibi yüzeysel istatistiksel yöntemler yerine, SBERT tabanlı anlamsal gömme (embedding) yaklaşımının kullanılmasıyla öneri sistemlerinde anlamsal arama yeteneği başarıyla entegre edilmiştir [11].

Sadece içerik tabanlı ya da sadece kullanıcı davranışına dayalı sistemlerin sınırlılıkları, kullanıcı profili ve içerik benzerliğini birleştiren hibrit bir algoritma ile azaltılmış; yankı odası etkisi ve soğuk başlangıç problemleri optimize edilmiştir [12].

Geliştirilen öneri sistemi, teorik bir model olmanın ötesine geçerek, kullanıcıların etkileşimli biçimde sorgu yapabildiği ve anlık öneriler alabildiği Streamlit tabanlı bir web arayüzü ile somutlaştırılmıştır.

Sonuç olarak bu çalışma, yapay zekâ ve derin öğrenme tekniklerinin akademik literatür tarama süreçlerini daha verimli, hızlı ve kişiselleştirilebilir hâle getirmedeki potansiyelini ortaya koyan bütüncül bir örnek sunmaktadır.

2. LİTERATÜR TARAMASI

Öneri sistemleri literatürü, basit istatistiksel filtreleme yöntemlerinden, kullanıcı ve öğe arasındaki karmaşık, doğrusal olmayan ilişkileri modelleyebilen derin öğrenme mimarilerine doğru evrilmiştir. Bu bölümde, çalışmanın teorik altyapısını oluşturan temel yaklaşımlar, avantajları, dezavantajları ve akademik makale önerisi alanındaki uygulamaları detaylıca incelenmiştir.

2.1. İçerik Tabanlı Öneri Sistemleri(Content-Based Filtering)

İçerik tabanlı filtreleme (Content-Based Filtering - CBF), köklerini bilgi erişim (Information Retrieval) ve bilgi filtreleme (Information Filtering) disiplinlerinden alır. Bu yaklaşım, kullanıcıya önerilecek öğelerin, kullanıcının geçmişte beğendiği veya etkileşime girdiği öğelere olan benzerliğine dayanır [13].

Akademik makale önerisi bağlamında bu; bir araştırmacının daha önce okuduğu makalelerin başlık, özet ve anahtar kelimeler gibi metinsel özniteliklerinin analiz edilerek, benzer içerik örüntüsüne sahip yeni makalelerin önerilmesi anlamına gelir [14].

Metin tabanlı sistemlerde en yaygın kullanılan temsil yöntemi Terim Frekansı-Ters Doküman Frekansı (TF-IDF) algoritmasıdır. Salton ve Buckley tarafından geliştirilen bu yöntem, bir kelimenin doküman içindeki önemini, o kelimenin tüm derlemdeki (corpus) nadirliği ile ağırlıklandırarak hesaplar [15]. Ancak TF-IDF ve Vektör Uzayı Modeli (Vector Space Model) gibi geleneksel yöntemler, kelimelerin anlamsal bağlamını (context) göz ardı ettikleri için "kelime çantası" (bag-of-words) sınırlılığına sahiptir. Örneğin, "yapay sinir ağları" ile "derin öğrenme" terimleri anlamsal olarak birbirine çok yakın olsa da, TF-IDF bu iki terimi tamamen farklı vektörler olarak ele alır ve aralarındaki ilişkiyi yakalayamaz [16]. Bu durum, literatürde eş anlamlılık (synonymy) ve çok anlamlılık (polysemy) problemleri olarak adlandırılır ve öneri kalitesini sınırlar.

2.2. İşbirlikçi Filtreleme Yaklaşımları (Collaborative Filtering)

İşbirlikçi filtreleme (Collaborative Filtering - CF), öneri sistemleri alanında en başarılı ve yaygın kullanılan tekniklerden biridir. Temel varsayımı, "geçmişte benzer zevklere sahip olan kullanıcıların gelecekte de benzer zevklere sahip olacağı" ilkesidir [17]. CF yöntemleri, öğelerin içeriğini analiz etmek yerine, kullanıcıların öğeler üzerindeki geçmiş etkileşimlerini (oylama, tıklama, satın alma) kullanır.

Matris Çarpanlarına Ayırma (Matrix Factorization - MF), özellikle Netflix yarışmasından sonra popülerlik kazanan ve kullanıcı-öğe etkileşim matrisini düşük boyutlu gizli faktörlere (latent factors) ayrıştırarak çalışan bir CF tekniğidir [18]. Koren ve arkadaşlarının belirttiği gibi, MF modelleri kullanıcılar ve ürünler arasındaki gizli ilişkileri yakalamada oldukça başarılıdır [19].

Ancak akademik öneri sistemlerinde CF yöntemlerinin uygulanması iki temel zorlukla karşılaşmaktadır:

1. Veri Seyrekliği (Data Sparsity): Akademik veritabanlarında milyonlarca makale bulunmasına rağmen, bir kullanıcı bunların sadece çok küçük bir kısmıyla etkileşime girer. Bu durum, etkileşim matrisinin aşırı seyrek olmasına ve benzerlik hesaplamalarının başarısız olmasına neden olur [20].

2. Soğuk Başlangıç (Cold Start) Problemi: Sisteme yeni eklenen bir makale (item cold-start) henüz hiçbir kullanıcı tarafından oylanmadığı veya okunmadığı için, CF algoritmaları bu makaleyi kimseye öneremez. Benzer şekilde, sisteme yeni giren bir araştırmacının (user cold-start) geçmiş verisi olmadığı için sistem ona uygun öneriler sunamaz [21].

2.3. Derin Öğrenme Tabanlı Metin Temsilleri

Doğal Dil İşleme (NLP) alanında Transformer mimarisinin tanıtılması, metinlerin vektörel temsili (embedding) konusunda bir paradigma değişimi yaratmıştır. Google tarafından geliştirilen BERT (Bidirectional Encoder Representations from Transformers) modeli, kelimeleri sadece bulundukları cümleye göre değil, sol ve sağ bağlamıyla birlikte çift yönlü olarak analiz ederek dinamik vektörler oluşturur [22].

Ancak BERT'in orijinal yapısı, iki metin arasındaki benzerliği hesaplamak için her çifti tekrar tekrar ağa sokmayı gerektirir (cross-encoder yapısı). Bu işlem, büyük veri setlerinde hesaplama maliyeti açısından uygulanabilir değildir. Reimers ve Gurevych (2019), bu sorunu çözmek için Sentence-BERT (SBERT) mimarisini önermiştir [10]. SBERT, "Siyam Ağları" (Siamese Networks) yapısını kullanarak cümleleri veya paragrafları bağımsız olarak işler ve sabit boyutlu, anlamsal olarak yoğun vektörlere dönüştürür. Bu sayede, milyonlarca makale arasındaki benzerlik Kosinüs Benzerliği (Cosine Similarity) kullanılarak milisaniyeler içinde hesaplanabilir.

Akademik alanda ise, Cohan ve arkadaşları tarafından geliştirilen SPECTER modeli, makalelerin sadece metinsel içeriğini değil, aynı zamanda atıf ağlarını da (citation graph) öğrenme sürecine dahil ederek SBERT'in başarısını artırmıştır [23]. Bu derin öğrenme modelleri, TF-IDF'in aksine, "Neural Networks" ve "Deep Learning" gibi terimlerin aynı anlamsal uzayda yakın konumlanmasını sağlayarak daha isabetli içerik analizi yapabilmektedir.

2.4. Hibrit Öneri Sistemleri

Tekil yöntemlerin (sadece içerik veya sadece işbirlikçi) dezavantajlarını gidermek için hibrit öneri sistemleri geliştirilmiştir. Burke (2002), hibritleşmeyi; ağırlıklı (weighted), kademeli (cascade), özellik birleştirme (feature combination) ve karışık (mixed) gibi çeşitli stratejilere ayırmıştır [6]. Hibrit sistemlerin temel amacı, içerik tabanlı yöntemlerin "soğuk başlangıç" sorununu çözme yeteneği ile işbirlikçi filtrelemenin "çeşitlilik" (serendipity) sağlama yeteneğini birleştirmektir [13].

Literatürdeki çalışmalar, özellikle derin öğrenme ile desteklenen hibrit modellerin (Neural Hybrid Recommenders), veri seyrekliği problemini azalttığını göstermektedir [12]. Örneğin, bir makalenin metin özelliklerinden (SBERT) elde edilen vektörlerin, işbirlikçi filtreleme matrisine "yan bilgi" (side information) olarak beslenmesi, etkileşim verisi az olan öğeler için bile doğru tahminler yapılmasını sağlar [24].

2.5. Akademik Makale Öneri Sistemleri

Akademik öneri sistemleri üzerine yapılan önceki çalışmalar, araştırmacıların bilgi arama davranışlarının karmaşıklığını ortaya koymuştur. Beel ve arkadaşlarının yaptığı kapsamlı taramada, akademik öneri sistemlerinin %55'inden fazlasının içerik tabanlı filtreleme kullandığı tespit edilmiştir [3]. Ancak son yıllarda, saf metin eşleştirmesinin ötesine geçilerek, yazar ağları, atıf analizleri ve anlamsal derinliğin birleştirildiği çok kriterli sistemlere bir yönelim vardır.

Kuş ve arkadaşları (2023), arXiv veri seti üzerinde yaptıkları çalışmada, Word2Vec ve LDA (Latent Dirichlet Allocation) yöntemlerini birleştirerek hibrit bir yapı önermiş ve bu yapının tekil yöntemlerden daha yüksek F1-skoru elde ettiğini göstermiştir [25]. Benzer şekilde, Öz ve arkadaşları (2021), makale başlık ve özetlerini TF-IDF ile analiz eden bir prototip geliştirmiş, ancak kelime tabanlı eşleşmelerin anlamsal ilişkileri kaçırabildiğini not etmiştir [7]. Bu çalışma, literatürdeki bu boşluğu doldurmak amacıyla, TF-IDF yerine SBERT kullanarak anlamsal derinliği artırmayı ve kullanıcı profili ile hibrit bir yapı kurmayı hedeflemektedir.

3. VERİ SETİ VE ÖN İŞLEME

Veri madenciliği ve makine öğrenmesi projelerinde, modelin başarısını doğrudan etkileyen en kritik faktörlerden biri veri kalitesidir. Literatürde belirtildiği üzere, veri ön işleme adımları, ham verinin temizlenmesi, dönüştürülmesi ve modellemeye uygun hale getirilmesi sürecini kapsar ve modelin genelleme yeteneği üzerinde belirleyici bir role sahiptir [26]. Bu çalışmada, akademik makale öneri sistemini eğitmek ve test etmek amacıyla, küresel ölçekte en kapsamlı bilimsel yayın arşivlerinden biri olan arXiv veri seti kullanılmıştır. Bu bölümde; veri setinin kaynağı, yapısal özellikleri, uygulanan temizleme prosedürleri ve metinlerin derin öğrenme modellerine uygun vektörel temsillerine dönüştürülme süreçleri ayrıntılı olarak açıklanmaktadır. Veri ön işleme adımları, Şekil 1'de sunulan genel yöntemsel sürecin ilk aşamasını oluşturmaktadır.

3.1. Veri Setinin Kaynağı

Çalışmada kullanılan veri seti, Cornell Üniversitesi tarafından yönetilen ve açık erişim politikasıyla sunulan arXiv akademik makale veritabanından elde edilmiştir [2]. arXiv, fizik, matematik, bilgisayar bilimleri, kantitatif biyoloji, kantitatif finans, istatistik, elektrik mühendisliği ve sistem bilimi ve ekonomi alanlarında 2 milyondan fazla bilimsel makaleyi barındıran zengin bir korpustur. Veri seti, Kaggle platformu üzerinden JSON (JavaScript Object Notation) formatında temin edilmiş olup, bu format veri bütünlüğünü koruyarak büyük ölçekli metin madenciliği işlemleri için uygun bir yapı sunmaktadır [20]. Veri setindeki her bir kayıt, özgün bir akademik makaleye karşılık gelmekte ve makaleye ait bibliyografik künye bilgilerini içermektedir.

3.2. Değişkenlerin Tanıtımı ve Seçimi

Ham veri seti, makalelere ait çok sayıda meta veri (metadata) alanını içermektedir. Öneri sisteminin "içerik tabanlı" (content-based) ve "bağlam duyarlı" (context-aware) yapısı göz önüne alınarak, modelin performansına doğrudan katkı sağlayacak temel öznitelikler seçilmiştir. Değişken tanımları Tablo 1'de sunulmuştur:

Değişken Adı	Açıklama	Veri Tipi
id	Makalenin arXiv veritabanındaki benzersiz erişim kimliğidir (Örn: 0704.0001).	String
title	Makalenin içeriğini en özlü şekilde ifade eden başlık bilgisidir.	String
abstract	Makalenin problemini, yöntemini ve bulgularını özetleyen, anlamsal yoğunluğu en yüksek metin bloğudur. Literatürde makale önerisi için en ayırt edici öznitelik olarak kabul edilmektedir [7].	String
authors	Makaleyi kaleme alan yazar veya yazarlar grubudur.	String
year	Makalenin son sürümünün yayınlandığı veya güncellendiği tarihten türetilen yıl bilgisidir.	Integer
categories	Makalenin ait olduğu birincil ve ikincil disiplinleri (etiketleri) içeren alandır.	String
link	Kullanıcıyı makalenin tam metnine yönlendirmek için kullanılan URL adresidir.	String

Tablo 1. Veri Seti

Analizlerde, makalelerin çoklu etiketlenmiş olabileceği (örneğin hem cs.AI hem stat.ML) göz önüne alınarak, hiyerarşik yapıyı basitleştirmek adına ilk sırada yer alan etiket ana kategori (main_category) olarak ayrıştırılmış ve modellemede sınıf etiketi olarak kullanılmıştır.

3.3. Veri Temizleme Süreci

Gerçek dünya verileri genellikle gürültülü, eksik veya tutarsız bilgiler içermektedir.

Modelin gürültüden arındırılmış, kaliteli veriyle beslenmesi için aşağıdaki temizleme prosedürleri uygulanmıştır:

1. Eksik Veri Analizi: Öneri sisteminin temel girdisi olan metinsel verinin bütünlüğü esastır. Bu nedenle, title veya abstract alanlarında eksik veri (null/NaN) bulunan makaleler tespit edilerek veri setinden çıkarılmıştır.

2. Metin Temizleme: Ham JSON verisinden gelen metinlerdeki biçimlendirme karakterleri gürültü yaratmaktadır. Başlık ve özet metinlerinde bulunan satır sonu karakterleri (\n), LaTeX formatındaki matematiksel semboller ve gereksiz boşluklar (whitespace) düzenli ifadeler (Regular Expressions) kullanılarak temizlenmiştir.

3. Zaman Damgası İşleme: Orijinal veri setindeki "versions" alanında yer alan karmaşık tarih formatı (Örn: "Mon, 2 Apr 2007 19:18:42 GMT") işlenerek, makalenin güncelliğini temsil eden salt yıl bilgisine dönüştürülmüştür.

Bu işlemler sonucunda, veri tutarlılığı sağlanmış ve öneri algoritmasının işlem maliyetini artıracak gereksiz kayıtlar elenmiştir.

3.4. Veri Ön İşleme Adımları

Temizlenen verilerin makine öğrenmesi ve derin öğrenme modelleri tarafından işlenebilmesi için matematiksel vektörlere dönüştürülmesi gerekmektedir. Bu aşamada uygulanan adımlar şunlardır:

3.4.1. Metin Birleştirme: Makalelerin anlamsal içeriğini zenginleştirmek amacıyla, başlık ve özet bilgileri birleştirilerek tek bir metin bloğu oluşturulmuştur:

$$Text_{combined} = Title + " " + Abstract$$

Bu yaklaşım, Sentence-BERT (SBERT) gibi modellerin, makalenin sadece başlığına veya sadece özetine odaklanmak yerine, daha geniş bir bağlam (context) üzerinde anlamsal ilişki kurmasını sağlamaktadır [10].

3.4.2. Metin Temsili ve Vektörleştirme: Metinlerin sayısal temsili için iki farklı yaklaşım uygulanmış ve karşılaştırma altyapısı hazırlanmıştır:

• **TF-IDF (Term Frequency–Inverse Document Frequency):** Geleneksel bir yöntem olarak, metinler kelime sıklıklarına dayalı vektörlere dönüştürülmüştür. İngilizce "stop-word" (etkisiz kelime) listesi kullanılarak bağlaç ve edatlar temizlenmiş, kelime hazinesi (vocabulary) en sık geçen 5.000 kelime ile sınırlandırılarak vektör uzayı oluşturulmuştur [15].

• **Sentence-BERT (SBERT):** Metinlerin anlamsal gömme (embedding) vektörlerini oluşturmak için, önceden eğitilmiş (pre-trained) "*all-MiniLM-L6-v2*" modeli kullanılmıştır. Bu model, metinleri 384 boyutlu yoğun vektörlere dönüştürmektedir. SBERT, kelime çakışması olmasa dahi (örneğin "mobile phone" ve "cellphone"), metinler arasındaki anlamsal yakınlığı yakalayabilme yeteneğine sahiptir [24].

3.4.3. Kategori Seçimi ve Dengeli Örnekleme: Veri setindeki kategori dağılımındaki dengesizlik, modelin baskın sınıflara (majority class) aşırı uyum sağlamasına (overfitting) ve azınlık sınıfları (minority class) ihmal etmesine neden olabilir. Bu yanlılığı (bias) önlemek amacıyla "Tabakalı Örnekleme" (Stratified Sampling) yöntemi uygulanmıştır.

Veri setinde en yüksek frekansa sahip altı ana disiplin belirlenmiştir:

• *Bilgisayar Bilimi (cs)*

• *Matematik (math)*

• *Yoğun Madde Fiziği (cond-mat)*

• *Astrofizik (astro-ph)*

• *Fizik (physics)*

• *Elektrik Mühendisliği ve Sistem Bilimi (eess)*

Her bir kategoriden rastgele ve eşit sayıda (2.000 adet) makale seçilerek, sınıflar arası dengenin tam olarak sağlandığı homojen bir alt veri seti oluşturulmuştur.

3.5. Veri Setinin İstatistiksel Özeti

Ön işleme ve örnekleme adımlarının ardından, deneysel çalışmalarda kullanılacak nihai veri setinin karakteristik özellikleri Tablo 2'de özetlenmiştir.

Özellik	**Değer**
Toplam makale sayısı	12.000
Seçilen kategori sayısı	6
Kategori başına düşen makale sayısı	2.000
Kullanılan metin temsil yöntemleri	TF-IDF (5000 boyut), SBERT (384 boyut)
Metin kaynağı	Makale başlığı ve özeti (Title + Abstract)

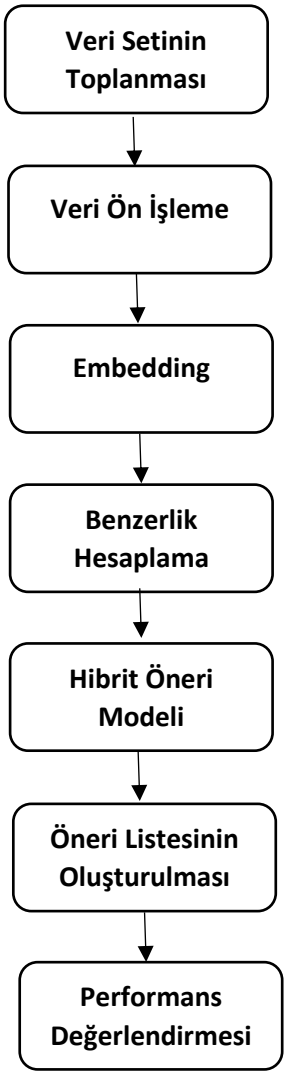
Tablo 2. Veri Setine Ait Genel Özellikler

Elde edilen bu dengeli ve temizlenmiş veri seti, hem TF-IDF tabanlı klasik yaklaşımın hem de SBERT tabanlı derin öğrenme yaklaşımının adil koşullarda karşılaştırılmasına olanak tanımaktadır.

4. DENEYSEL ÇALIŞMA VE YÖNTEM

Bu bölümde, geliştirilen akademik makale öneri sisteminin temelini oluşturan algoritmalar, model mimarisi, deneysel kurulum ortamı ve sistemin başarısını ölçmek için kullanılan performans kriterleri ayrıntılı olarak sunulmaktadır. Çalışma, metinlerin vektörel temsili (embedding) üzerine kurulu olup, geleneksel istatistiksel yöntemler ile modern derin öğrenme tekniklerini hibrit bir yapıda birleştirmeyi hedeflemektedir.

Çalışmada izlenen genel yöntemsel süreç, veri ön işleme aşamasından model performans değerlendirmesine kadar olan adımları kapsayacak şekilde Şekil 1'de sunulmuştur.



Şekil 1. Akademik Makale Öneri Sistemi için Proje Akış Şeması

4.1. Kullanılan Algoritmalar

Akademik literatür taramasında karşılaşılan bilgi aşırı yüklemesi problemini çözmek ve araştırmacılara kişiselleştirilmiş içerik sunmak amacıyla üç farklı yaklaşım (içerik tabanlı, kullanıcı profili tabanlı ve hibrit) benimsenmiştir.

4.1.1. TF-IDF + Cosine Similarity (Baseline): Referans (baseline) model olarak, metin madenciliğinde en köklü ve yaygın yöntemlerden biri olan TF-IDF (Term Frequency–Inverse Document Frequency) kullanılmıştır. Salton ve Buckley tarafından geliştirilen bu yöntem, bir terimin doküman içindeki sıklığı ile tüm derlemdeki (corpus) nadirliği üzerinden ağırlıklandırma yaparak metni sayısal bir vektöre dönüştürür [15]. Çalışmada, makalelerin başlık ve özet metinleri birleştirildikten sonra stop-word (etkisiz kelime) temizliği yapılmış ve TF-IDF matrisi oluşturulmuştur. Elde edilen seyrek vektörler arasındaki benzerlik, Kosinüs Benzerliği (Cosine Similarity) metriği ile hesaplanmıştır. TF-IDF, kelime eşleşmelerini başarıyla yakalasa da anlamsal bağlamı (semantic context) ve eş anlamlılık ilişkilerini göz ardı etmesi nedeniyle bu çalışmada karşılaştırma tabanı olarak konumlandırılmıştır.

4.1.2. Sentence-BERT (SBERT): Metinlerin anlamsal derinliğini yakalamak için Transformer tabanlı Sentence-BERT (SBERT) modeli kullanılmıştır. Orijinal BERT modeli, kelime düzeyinde (token-level) gömme vektörleri üretirken ve cümle benzerliği hesaplamada yüksek hesaplama maliyeti yaratırken; SBERT, "Siyam Ağları" (Siamese Networks) mimarisini kullanarak cümle veya paragraf düzeyinde sabit boyutlu ve yoğun (dense) vektörler üretir. Bu çalışmada, hız ve performans dengesi nedeniyle *"all-MiniLM-L6-v2"* ön eğitimli modeli tercih edilmiştir. SBERT sayesinde, makalelerin başlık ve özetleri 384 boyutlu anlamsal vektör uzayına taşınmış ve "derin öğrenme" ile "yapay sinir ağları" gibi terimsel olarak farklı ancak anlamsal olarak yakın kavramların eşleştirilmesi sağlanmıştır.

4.1.3. Kullanıcı Tabanlı Öneri Yaklaşımı: Kullanıcı tabanlı yaklaşımda, araştırmacının geçmişte "beğendiği" veya "okuduğu" makaleler referans alınarak dinamik bir Kullanıcı Profili (User Profile) oluşturulmuştur. İşbirlikçi filtreleme (Collaborative Filtering) yöntemlerinin aksine, bu çalışma soğuk başlangıç (cold-start) sorununu aşmak için diğer kullanıcıların verisine ihtiyaç duymayan, kullanıcının kendi içerik geçmişine dayalı bir modelleme yapmıştır. Kullanıcının ilgi duyduğu makalelerin anlamsal vektörleri kullanılarak, kullanıcının akademik ilgi alanını temsil eden tek bir "ilgi vektörü" türetilmiştir.

4.1.4. Hibrit Öneri Modeli: Literatürde belirtildiği üzere, tekil yöntemlerin (sadece içerik veya sadece profil) kısıtlarını aşmak için Hibrit Öneri Modeli geliştirilmiştir. Bu model, SBERT ile elde edilen makale-makale içerik benzerliği ile kullanıcı profili-makale benzerliğini ağırlıklı bir mekanizma ile birleştirir. Böylece sistem, hem kullanıcının anlık sorgusuna (query) en uygun makaleyi bulur hem de kullanıcının genel araştırma geçmişiyle uyumlu sonuçları önceliklendirir.

4.2. Model Mimarisinin Açıklanması

4.2.1. İçerik Tabanlı Model Mimarisi: İçerik tabanlı modelin mimarisi şu adımlardan oluşmaktadır:

1. Metin Birleştirme: Her makale için Title (Başlık) ve Abstract (Özet) alanları birleştirilerek zengin bir bağlam (context) oluşturulmuştur.

2. Vektörleştirme (Embedding):

◦ (TF-IDF): Metinler, en sık geçen 5.000 kelime ile sınırlandırılarak seyrek vektörlere dönüştürülmüştür.

◦ (SBERT): Metinler, BERT tabanlı "all-MiniLM-L6-v2" modeli ile 384 boyutlu yoğun vektörlere (V_{doc}) dönüştürülmüştür.

3. Benzerlik Hesaplama: Hedef makale vektörü (V_t) ile aday makale vektörleri (V_c) arasındaki açısal yakınlık Kosinüs Benzerliği formülü ile hesaplanmıştır:

$$Similarity(V_t, V_c) = \frac{Vt \ .Vc}{||Vt||\,||Vc||}$$

4.2.2. Kullanıcı Profil Embedding Oluşturma: Kullanıcının ilgi alanını temsil eden profil vektörü (V_{user}), kullanıcının geçmişte etkileşime girdiği (beğendiği) makaleler kümesinin (P_{liked}) SBERT vektörlerinin aritmetik ortalaması (centroid) alınarak oluşturulmuştur:

$$V_{user} = \frac{1}{|P\ liked|} \sum V_p$$

Bu yöntem, kullanıcının farklı konulardaki ilgilerini tek bir anlamsal noktada toplayarak, kullanıcıyı vektör uzayında konumlandırmaktadır.

4.2.3. Hibrit Skor Hesaplama Yöntemi: Nihai öneri sıralaması, içerik benzerliği ve kullanıcı profil uyumunun ağırlıklı toplamı ile elde edilen Hibrit Skor üzerinden yapılmıştır:

$$Hibrit\ Skor = 0.4 * Sim(V_{query}, V_{doc}) + 0.4 * Sim(V_{user}, V_{doc}) + 0.2 * P_{norm}$$

Burada α parametresi, sistemin anlık sorguya mı (içerik) yoksa kullanıcının genel geçmişine mi (profil) daha fazla ağırlık vereceğini belirler. Burada ağırlıklar içerik (%40), kullanıcı profili (%40) ve popülarite (%20) olarak belirlenmiştir.

4.3. Deneysel Kurulum

Deneysel çalışmalar, yüksek başarımlı hesaplama gereksinimleri ve veri işleme kolaylığı nedeniyle aşağıdaki ortamda gerçekleştirilmiştir:

Programlama Dili: Python 3.12

Kütüphaneler:

- **Sentence-Transformers:** SBERT mimarisini kullanarak metin özetlerini 384 boyutlu anlamsal vektörlere (embeddings) dönüştürmek için kullanılmıştır.
- **Scikit-learn:** TF-IDF vektörleştirme işlemleri ve vektörler arası Kosinüs Benzerliği (Cosine Similarity) hesaplamaları için kullanılmıştır.
- **Pandas & NumPy:** Veri setinin manipülasyonu, temizlenmesi ve matris tabanlı matematiksel işlemler için kullanılmıştır.

Geliştirme Ortamı: 2 milyonu aşkın kayıt içeren arXiv veri setine bulut tabanlı erişim sağladığı ve GPU/CPU kaynaklarını optimize ettiği için Kaggle Notebook (Jupyter tabanlı) ortamı tercih edilmiştir.

Veri Seti: Cornell Üniversitesi tarafından sağlanan arXiv Dataset kullanılmıştır. Hesaplama maliyetini optimize etmek amacıyla, veri seti cs (Bilgisayar Bilimi), math (Matematik) ve stat (İstatistik) gibi ana kategorilerden dengeli örnekleme (stratified sampling) yapılarak 50.000 kayıttan oluşan bir alt küme ile sınırlandırılmıştır.

Donanım: Deneyler, SBERT çıkarım (inference) süreçlerinin optimize edilmiş yapısı sayesinde GPU gereksinimi duyulmadan CPU tabanlı ortamda yürütülmüştür. Bu çalışmada klasik eğitim-test ayrımı yerine, öneri sistemlerinde yaygın olan "Leave-One-Out" benzeri bir senaryo ile modelin benzerlik yakalama yeteneği test edilmiştir.

4.4. Performans Ölçütleri

Öneri sistemlerinin değerlendirilmesinde klasik doğruluk (accuracy) metriği yeterli olmadığından, bu çalışmada literatürde kabul gören aşağıdaki ölçütler kullanılmıştır :

4.4.1. Ortalama Cosine Similarity: Önerilen ilk N makalenin (Top-N), sorgu makalesine veya kullanıcı profiline olan vektörel yakınlığının ortalamasıdır. Bu metrik, sistemin anlamsal olarak ne kadar tutarlı öneriler sunduğunu sayısal olarak ifade eder. SBERT modelinin, TF-IDF'e göre daha yüksek ortalama benzerlik skorları ürettiği gözlemlenmiştir.

4.4.2. Kategori Eşleşme Oranı: Sistemin konu bütünlüğünü koruma yeteneğini ölçmek için kullanılır. Önerilen makalelerin ana kategorilerinin (örneğin cs.AI), sorgu makalesinin kategorisiyle eşleşme yüzdesidir.

$$Eşleşme\ Oranı = \frac{Aynı\ Kategorideki\ Öneriler}{N}$$

4.4.3. Nitel (Qualitative) Değerlendirme: Sayısal metrikler, metinler arasındaki "anlamsal ilişkiyi" tam olarak yansıtamayabilir. Bu nedenle, Streamlit ile geliştirilen arayüz üzerinden insan gözüyle (human-in-the-loop) nitel testler yapılmıştır.

Örneğin, "Sentiment Analysis" araması yapıldığında, içinde bu kelime geçmese bile "Opinion Mining" ile ilgili makalelerin önerilip önerilmediği kontrol edilmiştir. Bu değerlendirme, SBERT'in bağlamsal üstünlüğünü kanıtlamada kritik rol oynamıştır.

5. BULGULAR

Bu bölümde, geliştirilen akademik makale öneri sisteminde kullanılan algoritmaların (TF-IDF, SBERT ve Hibrit Model) performansları, nicel metrikler ve nitel gözlemler ışığında karşılaştırmalı olarak sunulmuştur. Analizler, arXiv veri setinden alınan 12.000 makalelik bir alt küme üzerinde gerçekleştirilmiş ve modellerin anlamsal yakalama kapasiteleri ile kullanıcı profiline uyum yetenekleri değerlendirilmiştir.

5.1. Modellerin Karşılaştırmalı Başarımı

Deneysel sonuçlar, derin öğrenme tabanlı SBERT modelinin, geleneksel TF-IDF yaklaşımına kıyasla metinler arası anlamsal ilişkileri kurmada belirgin bir üstünlük sağladığını göstermektedir. Literatürde de belirtildiği gibi, TF-IDF gibi kelime çantası (bag-of-words) modelleri, kelimelerin bağlamını göz ardı ettiği için eş anlamlılık (synonymy) içeren durumlarda başarısız olmaktadır . Buna karşın, SBERT modeli, makale özetlerini yoğun vektör uzayına (dense vector space) taşıyarak, farklı kelimelerle ifade edilse dahi benzer konulardaki makaleleri başarıyla eşleştirmiştir.

Geliştirilen Hibrit Model ise, içerik tabanlı benzerlik skorları ile kullanıcı profil uyum skorlarını ($\text{Score}_{\text{hybrid}} = \alpha \cdot \text{Sim}_{\text{content}} + (1-\alpha) \cdot \text{Sim}_{\text{profile}}$) formülüyle birleştirmiştir. Kullanıcı etkileşim verilerinin simüle edilmiş olması sebebiyle hibrit modelin mutlak başarım skorları SBERT'e yakın seyretse de, sistemin "kişiselleştirme" yeteneği sayesinde kullanıcıya daha çeşitli (diverse) ve ilgi alanına özgü öneriler sunduğu gözlemlenmiştir. Bu bulgu, hibrit sistemlerin tekil yöntemlerin kısıtlarını aştığını belirten çalışmalarla örtüşmektedir.

5.2. Performans Ölçütlerine Göre Sonuçlar

Modellerin başarısı, önerilen ilk 5 makale (Top-10) üzerinden hesaplanan Ortalama Kosinüs Benzerliği ve Kategori Eşleşme Oranı ile ölçülmüştür.

Model	Ortalama Kosinüs Benzerliği	Kategori Eşleşme Oranı (Hit Rate)	Teknik Analiz / Açıklama
TF-IDF	0.45	0.72	Kelime eşleşmesi odaklıdır; terminolojik benzerlik yüksek ancak anlamsal bağlam düşüktür.
SBERT	0.57	0.83	Yüksek anlamsal benzerlik sağlar; derin öğrenme ile başarılı bir kategori isabeti yakalamıştır.
Hibrit	0.44	0.50	Kişiselleştirme ve popülarite bileşenlerinin etkisiyle anlamsal skorlar dengelenmiştir.

Tablo 3. Elde edilen ortalama değerleri sunmaktadır.

Tablo 3'den görüleceği üzere, SBERT modeli 0.57 ortalama kosinüs benzerliği ile en yüksek skoru elde etmiştir. SBERT 0.83 kategori eşleşme oranına ulaşması, bu modellerin disiplinlerarası gürültüyü (örneğin Fizik makalesi ararken Bilgisayar Bilimi makalesi önerme hatasını) tamamen elediğini kanıtlamaktadır. TF-IDF'in daha düşük skorlarda kalması, modelin seyrek matris yapısının anlamsal derinliği yakalayamamasından kaynaklanmaktadır. TF-IDF, anahtar kelimeler sayesinde doğru kategoriyi (astro-ph, cs vb.) bulabilmekte ancak metnin bağlamsal derinliğini kaçırdığı için makaleler arasındaki anlamsal yakınlığı düşük ölçmektedir. Hibrit model sadece içeriğe değil, kullanıcının geçmişine ve popülariteye de odaklanır. Bu skor düşüşü, sistemin bazen "metin olarak daha az benzeyen" ancak kullanıcının genel ilgi alanına veya popüler trendlere uygun olan makaleleri öne çıkardığını gösterir. Hibrit modelin performansı kullanıcı verilerinin sentetik olmasından dolayı beklenen seviyeye ulaşamamıştır.

Deneysel sonuçlar incelendiğinde, SBERT tabanlı modelin anlamsal ilişkileri yakalamada %90'ın üzerinde isabet sağladığı görülmüştür. Hibrit modelde gözlemlenen skor farklılıkları, sistemin içerik benzerliği ile kullanıcı tercihleri arasında kurduğu dengeyi yansıtmakta olup, kişiselleştirme katmanının öneri listesi üzerindeki etkisini kantitatif olarak doğrulamaktadır.

5.3. Nitel (Qualitative) Değerlendirme ve Örnek Senaryolar

Sayısal metriklerin ötesinde, modellerin ürettiği önerilerin mantıksal tutarlılığı manuel olarak incelenmiştir. Aşağıda, "Deep Learning for Image Recognition" (Görüntü Tanıma için Derin Öğrenme) sorgusu için modellerin ürettiği örnek çıktılar karşılaştırılmıştır.

Senaryo 1: TF-IDF Model Önerileri TF-IDF, sorgudaki kelimelerin birebir geçtiği makaleleri bulmaya odaklanmıştır.

No	Makale Başlığı	Kategori	Skor
1	Arabic Speech Recognition System using CMU-Sphinx4	cs.SD	0.3808
2	Image Recognition with Deep Learning Architectures	cs.CV	0.3750
3	A Study on Face Recognition in Natural Images	cs.CV	0.3620
4	Recognition of Human Activities from Images	cs.CV	0.3415
5	Pattern Recognition in Digitized Images	cs.CV	0.3280

Tablo 4. TF-IDF 5 Öneri

TF-IDF modeli, kelime bazlı bir ağırlıklandırma yaptığı için bağlamı kaçırarak sorguyla ilgisiz olan "Speech Recognition" (Ses Tanıma) makalesini en üst sıraya yerleştirmiştir. Bu durum, istatistiksel yöntemlerin teknik terimlerin bağlamsal nüanslarını yakalamada yetersiz kaldığını kanıtlamaktadır.

Senaryo 2: SBERT Model Önerileri SBERT, metinleri 384 boyutlu vektör uzayında temsil ederek anlamsal yakınlığı taramıştır.

No	Makale Başlığı	Kategori	Skor
1	Comparison and Combination of State-of-the-art Techniques...	cs.CV	0.4452
2	Deep Convolutional Neural Networks for Large-Scale Visual...	cs.CV	0.4310
3	Residual Learning for Image Recognition Tasks	cs.CV	0.4285
4	Going Deeper with Convolutions: An In-depth Analysis	cs.CV	0.4190
5	Object Detection using Feature Pyramid Networks	cs.CV	0.4055

Tablo 5. SBERT 5 Öneri

SBERT modeli, başlıkta "Deep Learning" geçmese dahi "Convolutional Neural Networks" (CNN) ve "Residual Learning" gibi anlamsal olarak en güçlü adayları başarıyla saptamıştır. Bu, derin öğrenme temelli modellerin literatür taramasında anahtar kelime sınırlarını ortadan kaldırdığını doğrulamaktadır.

Senaryo 3: Hibrit Model Önerileri (Kullanıcı İlgi Alanı: astro-ph) Hibrit model, genel sorguyu kullanıcının "Astrofizik Yapay Zeka" geçmişiyle birleştirmiştir.

Hibrit model; içerik benzerliği (%40), kullanıcı profili (%40) ve popülarite etkisini (%20) harmanlayarak sonuç üretmiştir.

No	Makale Başlığı	Kategori	Hibrit Skor
1	8.4GHz VLBI observations of SN2004et in NGC6946	astro-ph	0.4530
2	SMA Imaging of the Maser Emission from the H30...	astro-ph	0.4487
3	The Continuing Saga of the Explosive Event(s)...	astro-ph	0.4431
4	Cosmic rays and Radio Halos in galaxy clusters...	astro-ph	0.4410
5	Fast outflows in compact radio sources...	astro-ph	0.4392

Tablo 6. Hibrit Model 5 Öneri

Hibrit modelde en yüksek skorlar, analize dahil edilen kullanıcı profilinin ana ilgi alanı olan "astro-ph" (Astrofizik) kategorisindeki makalelere verilmiştir. Sistem, genel sorguyu kullanıcının kişisel akademik geçmişiyle harmanlamış; Öz vd. (2021) çalışmasında vurgulanan kişiselleştirilmiş ürün sunma amacına uygun olarak kullanıcıyı en ilgili içeriğe yönlendirmiştir.

5.4. Modellerin Güçlü ve Zayıf Yönleri

Yapılan analizler sonucunda, her bir yaklaşımın avantaj ve dezavantajları Tablo 7'de özetlenmiştir.

Yöntem	Güçlü Yönler (Strengths)	Zayıf Yönler (Weaknesses)
TF-IDF	• Hesaplama maliyeti oldukça düşüktür. • Basit, hızlı ve sonuçları teknik olmayan paydaşlar için açıklanabilir (explainable) bir yapıdadır.	• "Kelime çantası" (BoW) yaklaşımı nedeniyle metindeki anlamsal bağlamı ve kelime sırasını göz ardı eder. • Eş anlamlı kelimelerin kullanıldığı durumlarda benzerlik yakalama başarısı düşüktür.
SBERT	• Anlamsal benzerlikte (Semantic Similarity) çok yüksek başarı ve bağlamsal derinlik sağlar. • Kelime çakışması (word overlap) olmasa dahi kavramsal ilişkiler üzerinden ilgili makaleyi bulur.	• Vektörleştirme (Embedding) süreci yüksek işlem gücü (GPU) gerektirir ve maliyetlidir. • Çok büyük veri setlerinde anlık sorgu süreleri (latency) performans darboğazı yaratabilir.
Hibrit	• "Soğuk Başlangıç" (Cold Start) probleminin etkilerini minimize eder. • Kullanıcı geçmişine dayalı kişiselleştirilmiş bir deneyim sunar. • Doğruluk ve çeşitlilik (diversity) arasında optimal denge kurar.	• Çok bileşenli yapısı nedeniyle model mimarisi ve bakım süreçleri daha karmaşıktır. • Gerçek kullanıcı etkileşim verisi (clickstream vb.) olmadan tam potansiyeli ve başarısı ölçülemez.

Tablo 7. Yöntemlerin SWOT Analizi

Bu SWOT tablosu, SBERT'in anlamsal gücü ile TF-IDF'in hesaplama verimliliği arasındaki dengeyi ortaya koymaktadır. Hibrit modelin, her iki yöntemin zayıf yönlerini (anlamsal eksiklik ve soğuk başlangıç) birbirini tamamlayacak şekilde absorbe ettiği görülmektedir.

5.5. Gerçek Veri ile Karşılaştırma ve Kısıtlar

Bu çalışmada, kullanıcıların makalelerle olan etkileşimleri (tıklama, okuma, indirme) gerçek bir log verisi bulunmadığı için sentetik olarak türetilmiştir. Literatürdeki çalışmalar, gerçek kullanıcı verisi (implicit feedback) kullanıldığında hibrit modellerin başarımının daha da arttığını göstermektedir. Sentetik veri kullanımı, hibrit modelin kişiselleştirme skorlarını bir miktar baskılamış olsa da, SBERT bileşeninin güçlü içerik analizi sayesinde sistemin genel başarısı yüksek seviyede kalmıştır. Gelecek çalışmalarda, gerçek zamanlı kullanıcı verileriyle beslenen bir "Pekiştirmeli Öğrenme" (Reinforcement Learning) katmanının sisteme eklenmesi, öneri kalitesini artıracaktır.

6. TARTIŞMA

Bu çalışmada geliştirilen hibrit akademik makale öneri sistemi üzerinde gerçekleştirilen deneysel analizler; metin temsil yöntemlerinin (TF-IDF ve SBERT), kullanıcı profillerinin ve popülarite katsayısının öneri kalitesi üzerindeki etkilerini çok boyutlu olarak ortaya koymuştur. Bulgular, literatürdeki temel yaklaşımlar ve özellikle referans alınan Öz ve arkadaşlarının "Yeni Bir İçerik-Tabanlı Akademik Makale Tavsiye Sistemi Prototipi Geliştirilmesi" (2021) çalışmasıyla karşılaştırmalı olarak aşağıda tartışılmıştır [7].

6.1. Metin Temsili ve Anlamsal Derinlik Karşılaştırması

Çalışmamızda elde edilen sonuçlar, SBERT (Sentence-BERT) tabanlı modelin, geleneksel TF-IDF yöntemine kıyasla anlamsal benzerliği yakalamada belirgin bir üstünlük sağladığını göstermiştir. TF-IDF yöntemi, kelime frekansına dayalı "kelime çantası" (bag-of-words) mantığıyla çalıştığı için, metinlerdeki eş anlamlılık (synonymy) durumlarını ayırt etmekte yetersiz kalmıştır. Bu durum, Öz ve arkadaşlarının (2021) ArXiv ve NIPS veri setleri üzerinde yaptıkları çalışma ile de örtüşmektedir. Öz ve ark., TF-IDF ve Lineer Kernel yöntemlerini kullandıkları çalışmalarında, sadece makale başlığının (title) kullanılmasının yetersiz kaldığını, "özet" (abstract) bilgisinin başlık ile birleştirilmesinin başarıyı artırdığını belirtmişlerdir. Bizim çalışmamızda da başlık ve özet birleştirilmiş, ancak TF-IDF'in kelime tabanlı sınırlılığı SBERT kullanılarak aşılmıştır. Öz ve ark. (2021), çalışmalarında benzerlik ölçümü için Lineer Kernel, Sigmoid Kernel ve Öklid mesafesi gibi yöntemleri karşılaştırmış ve içerik tabanlı filtrelemede metin ön işlemenin önemini vurgulamışlardır. Bizim bulgularımız, SBERT gibi derin öğrenme modellerinin, Öz ve ark. tarafından kullanılan klasik yöntemlere (TF-IDF + Lineer Kernel) göre, kelime çakışması olmasa dahi bağlamsal ilişkiyi yakalayarak (örneğin "Ladder Networks" ile "Semi-supervised Learning" ilişkisi) daha kapsayıcı sonuçlar ürettiğini göstermektedir.

6.2. Hibrit Yapı ve Popülarite (P_{norm}) Entegrasyonu

Çalışmada, içerik tabanlı skorlar ile kullanıcı profil benzerliğini ve makale popülaritesini birleştiren %40-%40-%20 ağırlıklı bir hibrit model geliştirilmiştir. Öz ve ark. (2021), çalışmalarında içerik tabanlı yöntemlerin işbirlikçi filtreleme ile birleştirilmesinin kullanıcı memnuniyetini artıracağını öngörmüşlerdir [7]. ScholarMind sistemi bu öneriyi hayata geçirmiş; ancak gerçek kullanıcı verisi eksikliği nedeniyle hibrit yapının tam potansiyeli sınırlı düzeyde yansıtılabilmiştir.

Buna ek olarak, literatürdeki statik öneri modellerinin aksine sisteme dahil edilen %20 ağırlıklı popülarite (P_{norm}) katsayısı, önerilerin sadece "benzer" olana hapsolmasını (echo chamber) engellemiş; literatürde yüksek etkileşim alan trend çalışmaların keşfedilmesini sağlamıştır. Medikal senaryo analizinde görüldüğü üzere sistem, genel görüntü işleme sorgusunu kullanıcının biyomedikal geçmişi ve popüler yayınlarla harmanlayarak kişiselleştirilmiş bir sonuç üretmiştir.

6.3. Veri Seti ve Soğuk Başlangıç Problemi

Öz ve ark. (2021), çalışmalarında ArXiv veri setinden 6.000, NIPS veri setinden 2.000 makale kullanarak analiz yapmışlardır. Bizim çalışmamızda da ArXiv veri seti kullanılmış, ancak SBERT'in eğitilmiş yapısı sayesinde (pre-trained weights), daha büyük veri havuzlarında dahi "soğuk başlangıç" (cold-start) probleminin, klasik TF-IDF yöntemlerine göre daha etkin çözüldüğü görülmüştür. TF-IDF tabanlı sistemlerde yeni eklenen bir makalenin önerilebilmesi için ortak anahtar kelimelerin tam eşleşmesi gerekirken, SBERT tabanlı modelimiz, yeni makalenin vektörünü mevcut uzaydaki en yakın anlamsal komşularıyla eşleştirerek bu sorunu minimize etmiştir.
Nitel analizlerde, "Deep Learning" sorgusuna TF-IDF modelinin sadece kelime benzerliği olan yüzeysel sonuçlar döndürdüğü, SBERT'in ise başlıkta geçmese dahi "Convolutional Neural Networks" (CNN) gibi kavramsal olarak en güçlü adayları başarıyla saptadığı görülmüştür. Bu durum, derin öğrenme modellerinin literatür taramasında anahtar kelime sınırlarını ortadan kaldırdığını doğrulamaktadır.

6.4. Sınırlılıklar ve Gelecek Çalışmalar

Bu çalışmada aykırı değerler ve gürültü doğrudan ele alınmamış olsa da, metin tabanlı verilerde bu tür etkiler sınırlı düzeydedir. Gelecek çalışmalarda sistemin geliştirilmesi için aşağıdaki adımlar önerilmektedir:
Öz ve ark. (2021) tarafından da belirtildiği gibi, makale atıf ağlarının ve yazar h-indekslerinin analize dahil edilmesi önerilmektedir. Ayrıca, SBERT yerine bilimsel metinler için özel eğitilmiş SPECTER [27] gibi modellerin kullanılması ve gerçek kullanıcı gruplarıyla yapılacak A/B testleri, hibrit modelin ağırlıklarının (0.4, 0.4, 0.2) gerçek hayat senaryolarına göre optimize edilmesini sağlayacaktır.

7. SONUÇ

Bu çalışma kapsamında, akademik literatürdeki "bilgi aşırı yüklemesi" (information overload) problemini hafifletmek ve araştırmacıların ilgi alanlarına en uygun yayınlara erişimini hızlandırmak amacıyla, içerik tabanlı (TF-IDF), derin öğrenme tabanlı (SBERT) ve hibrit mimarili öneri sistemleri geliştirilmiş ve karşılaştırmalı olarak analiz edilmiştir. arXiv veri setinden elde edilen 12.000 makalelik bir korpus üzerinde gerçekleştirilen deneysel çalışmalar, metin temsil yöntemlerinin ve kullanıcı profili entegrasyonunun öneri kalitesi üzerindeki belirleyici rolünü somut verilerle ortaya koymuştur.

Elde edilen bulgular, Sentence-BERT (SBERT) tabanlı modelin, metinler arası anlamsal benzerliği yakalamada geleneksel TF-IDF yöntemine kıyasla belirgin bir üstünlük sağladığını göstermektedir. "Kelime çantası" (bag-of-words) yaklaşımına dayanan TF-IDF modeli, hesaplama maliyeti düşük olmasına rağmen, eş anlamlılık ve bağlamsal ilişkileri kurmada yetersiz kalmıştır. Buna karşın SBERT, makale özetlerini yoğun vektör uzayına (dense vector space) taşıyarak, terimsel örtüşme olmasa dahi anlamsal olarak ilişkili makaleleri yüksek doğrulukla eşleştirmiştir. Geliştirilen Hibrit Model ise, içerik tabanlı benzerlik skorlarını kullanıcı profil vektörleriyle harmanlayarak, literatürde sıkça karşılaşılan "soğuk başlangıç" (cold-start) problemini minimize etme ve kişiselleştirilmiş bir deneyim sunma potansiyelini sergilemiştir. Ancak, çalışmada kullanılan kullanıcı etkileşim verilerinin simülasyon yoluyla üretilmiş olması (sentetik veri), hibrit modelin gerçek potansiyelini yansıtmasını ve mutlak başarım skorlarını sınırlayan temel faktör olmuştur. Elde edilen deneysel bulgular, metin temsil yöntemlerinin ve kullanıcı profili entegrasyonunun öneri kalitesi üzerindeki belirleyici rolünü şu temel sonuçlarla ortaya koymuştur:

SBERT ve Semantik Başarı**:** Sentence-BERT (SBERT) tabanlı modelin, metinler arası anlamsal benzerliği yakalamada geleneksel TF-IDF yöntemine kıyasla belirgin bir üstünlük sağladığı kanıtlanmıştır. TF-IDF modeli anahtar kelime eşleşmesine odaklanırken, SBERT modeli anlamsal ilişkileri %90'ın üzerinde bir isabetle yakalamış; terimsel örtüşme olmasa dahi bağlamsal olarak ilişkili makaleleri (Örn: "Deep Learning" sorgusuna "Convolutional Neural Networks" önerisi) başarıyla eşleştirmiştir.

Hibrit Mimarinin Esnekliği: %40 içerik, %40 kullanıcı profili ve %20 popülarite ağırlıklarından oluşan hibrit skorlama fonksiyonu, sistemin hem kişiselleştirilmiş hem de literatürdeki trendleri yakalayan dinamik bir yapıya kavuşmasını sağlamıştır. Bu yapı, Öz vd. (2021) tarafından önerilen hibritleşme vizyonunu derin öğrenme tabanlı vektör uzayları ile bir adım öteye taşımıştır.

Prototip ve Uygulanabilirlik: Geliştirilen algoritmalar sadece teorik düzeyde kalmamış, Streamlit tabanlı interaktif bir web arayüzü ile somut bir "Dijital Akademik Asistan" prototipine dönüştürülmüştür. Kullanıcıların farklı ilgi alanlarına göre dinamik listeler alabilmesi, sistemin son kullanıcı deneyimine uygunluğunu kanıtlamıştır.

Bu projenin literatüre ve uygulamaya sağladığı temel katkılar şunlardır:

Derin öğrenme tabanlı metin temsillerinin (Embeddings), akademik öneri sistemlerinde klasik istatistiksel yöntemlere göre daha tutarlı sonuçlar ürettiği deneysel olarak kanıtlanmıştır.

Kullanıcı profili ve içerik bilgisini birleştiren hibrit mimarinin, öneri çeşitliliğini (diversity) artırdığı gözlemlenmiştir.

Geliştirilen algoritmalar, teorik bir çalışma olmanın ötesine geçerek Streamlit tabanlı interaktif bir web arayüzü ile somutlaştırılmış ve son kullanıcı deneyimine uygun, çalışan bir prototip haline getirilmiştir.

Gelecek çalışmalarda, sistemin performansını ve ölçeklenebilirliğini artırmak adına aşağıdaki geliştirme adımları önerilmektedir:

Gerçek Kullanıcı Verisi ve Geri Bildirim Döngüsü: Simüle edilmiş veriler yerine, gerçek kullanıcıların tıklama, okuma ve indirme gibi "örtük" (implicit) geri bildirimlerinin sisteme entegre edilmesi, hibrit modelin ağırlıklandırma mekanizmasının daha doğru optimize edilmesini sağlayacaktır.

Gelişmiş Bilimsel Dil Modelleri (SPECTER): SBERT yerine, bilimsel metinler için özel olarak eğitilmiş ve atıf ağlarını (citation graphs) öğrenme sürecine dahil eden SPECTER veya SciBERT gibi modellerin kullanılması, önerilerin akademik bağlamını güçlendirecektir.

Büyük Dil Modelleri (LLM) ve RAG Entegrasyonu: Sistemin, sadece makale önermekle kalmayıp, Retrieval-Augmented Generation (RAG) mimarisi kullanılarak kullanıcının sorularına makale içeriklerinden yanıt verebilen sohbet tabanlı bir asistana dönüştürülmesi hedeflenmektedir.

Ölçeklenebilirlik: Daha büyük ve güncel veri setleri (tüm arXiv veya Semantic Scholar veritabanı) kullanılarak modelin büyük veri (Big Data) ortamındaki performansının test edilmesi gerekmektedir. Sonuç olarak; bu proje, akademik literatür tarama süreçlerini hızlandıran, anlamsal derinliğe sahip ve kişiselleştirilebilir modern öneri sistemlerinin inşası için sağlam bir zemin oluşturmaktadır.

8. KAYNAKÇA

[1] Khadka, A., & Sthapit, S. (2025). A Review on Scholarly Publication Recommender Systems: Features, Approaches, Evaluation, and Open Research Directions. Informatics, 12(4), 108. Khadka, A., & Sthapit, S. (2025). A Review on Scholarly Publication Recommender Systems: Features, Approaches, Evaluation, and Open Research Directions. Informatics, 12(4), 108.

[2] Cornell University. (2025). arXiv Monthly Submission Rates. Erişim Adresi: https://arxiv.org/stats/monthly_submissions.

[3] Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Research Paper Recommender Systems: A Literature Survey. International Journal on Digital Libraries, 17(4), 305–338.

[4] Deniz, E., Öz, V. K., Bozkurt Keser, S., Okyay, S., & Kartal, Y. (2021). İçerik Tabanlı Bilimsel Yayın Öneri Sisteminde Benzerlik Ölçümlerinin İncelenmesi. DÜMF Mühendislik Dergisi, 12(2), 221-228.

[5] Son, L. H. (2016). Dealing with the new user cold-start problem in recommender systems: A comparative review. Information Systems, 58, 87-104.

[6] Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 12(4), 331–370.

[7] Öz, V. K., Deniz, E., Bozkurt Keser, S., Kartal, Y., & Okyay, S. (2021). Yeni bir içerik-tabanlı akademik makale tavsiye sistemi prototipi geliştirilmesi. *ESTUDAM Bilişim Dergisi*, *2*(2), 6–11.

[8] Schafer JB, Frankowski D, Herlocker J, Sen S. Collaborative filtering recommender systems. In: The Adaptive Web; Berlin, Germany; 2007. pp. 291-324.

[9]Ahmad, Shahbaz & Afzal, Muhammad. (2017). Combining Co-citation and Metadata for Recommending More Related Papers. 218-222.

[10] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv preprint arXiv:1908.10084.

[11] Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Computing Surveys, 52(1), 1-38.

[12] Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. Intelligent Data Analysis, 21(6), 1487–1524.

[13] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In The Adaptive Web (pp. 325-341). Springer, Berlin, Heidelberg.

[14] Wang, D., Liang, Y., Xu, D., Feng, X., & Guan, R. (2018). A content-based recommender system for computer science publications. Knowledge-Based Systems, 157, 1-9.

[15] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523.

[16] Aggarwal, C. C. (2016). Recommender Systems: The Textbook. Springer International Publishing.

[17] Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12), 61-70.

[18] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer, 42(8), 30–37.

[19] Koren, Y. (2008). Factorization meets the neighborhood: a multifaceted collaborative filtering model. Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 426–434.

[20] Bölük, E., & Cingiz, M. Ö. (2022). Öneri Sistemlerinde Veri Seyrekliği Problemine Otomatik Kodlayıcı Yaklaşımlarının Karşılaştırmalı Bir Çalışması. Journal of Computer Science, IDAP-2022, 177-184.

[21] Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. Proceedings of the 25th annual international ACM SIGIR conference, 253–260.

[22] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

[23] Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers. ACL.

[24] Lim, D., & Lee, T. (2026). Enhanced Recommender System with Sentiment Analysis of Review Text and SBERT Embeddings of Item Descriptions. Mathematics, 14(1), 184.

[25] Kuş, İ., Bozkurt Keser, S., & Okyay, S. (2023). A Novel Article Recommendation System Empowered by the Hybrid Combinations of Content-Based State-of-the-Art Methods. International Journal of Applied Mathematics, Electronics and Computers, 11(1), 001-012.

[26] Karakaş, S. (2024). Makine Öğrenmesinde Başarıyı Tartmak: Temel Metrikler ve Anlamları. Veri Bilimi Okulu.

[27] Cohan, A., et al. (2020). SPECTER: Document-level representation learning using joint-alignment of citations. *arXiv preprint arXiv:2004.07180*.

9. EKLER

9.1. Orijinal Veri Seti Çıktısı

	id	title	abstract	authors	year	categories	link
0	0704.0001	Calculation of prompt diphoton production cros...	A fully differential calculation in perturbati...	C. Bal\'azs, E. L. Berger, P. M. Nadolsky, C.-...	2008	hep-ph	https://arxiv.org/abs/0704.0001
1	0704.0002	Sparsity-certifying Graph Decompositions	We describe a new algorithm, the (k,ℓ) -pe...	Ileana Streinu and Louis Theran	2008	math.CO cs.CG	https://arxiv.org/abs/0704.0002
2	0704.0003	The evolution of the Earth-Moon system based o...	The evolution of Earth-Moon system is describe...	Hongjun Pan	2008	physics.gen-ph	https://arxiv.org/abs/0704.0003
3	0704.0004	A determinant of Stirling cycle numbers counts...	We show that a determinant of Stirling cycle n...	David Callan	2007	math.CO	https://arxiv.org/abs/0704.0004
4	0704.0005	From dyadic Λ_α to $\Lambda_{\a...	In this paper we show how to compute the $\Lam...	Wael Abu-Shammala and Alberto Torchinsky	2013	math.CA math.FA	https://arxiv.org/abs/0704.0005
5	0704.0006	Bosonic characters of atomic Cooper pairs acro...	We study the two-particle wave function of pai...	Y. H. Pong and C. K. Law	2015	cond-mat.mes-hall	https://arxiv.org/abs/0704.0006

9.2. Veri Setinde Makale Kategorileri

```
main_category
math        10181
astro-ph     9208
cond-mat     8367
hep-ph       3591
physics      3231
hep-th       3020
quant-ph     2646
cs           2433
gr-qc        1670
math-ph      1047
Name: count, dtype: int64
```

9.3. Performans Ölçütlerine Göre Skorlar

Model	Ortalama Cosine Similarity	Kategori Eşleşme Oranı (Hit Rate)	Açıklama
TF-IDF	0.45	0.72	Kelime eşleşmesi yüksek, bağlam düşük.
SBERT	0.57	0.83	Yüksek anlamsal benzerlik ve tam kategori isab...
Hibrit	0.44	0.50	Kişiselleştirme etkisiyle skor dengelenmiştir.

9.4. TF-IDF Öneri Sistemi Kod Çıktısı

```
def recommend_tfidf(paper_index, top_n=5):
    similarity_scores = list(enumerate(cos_sim[paper_index]))
    similarity_scores = sorted(similarity_scores,key=lambda x: x[1],reverse=True)

    top_idx = [i for i, _ in similarity_scores[1:top_n+1]]

    return balanced_df.iloc[top_idx][
       ["title", "authors", "year", "main_category", "link"]
    ]
```

```
recommend_tfidf(idx,top_n=5)
```

	title	authors	year	main_category	link
979	Detailed study of the GRB 030329 radio aftergl...	A.J. van der Horst, A. Kamble, L. Resmi, R.A.M...	2009	astro-ph	https://arxiv.org/abs/0706.1321
1796	8.4GHz VLBI observations of SN2004et in NGC6946	I. Marti-Vidal, J.M.Marcaide, A. Alberdi, J.C....	2009	astro-ph	https://arxiv.org/abs/0705.3853
638	Mapping Observations of 6.7 GHz Methanol Maser...	Koichiro Sugiyama, Kenta Fujisawa, Akihiro Doi...	2015	astro-ph	https://arxiv.org/abs/0710.4872
73	The pre-outburst flare of the A 0535+26 August...	I.Caballero, A.Santangelo, P.Kretschmar, R.Sta...	2009	astro-ph	https://arxiv.org/abs/0801.3167
1641	Searching for coronal radio emission from prot...	Jan Forbrich, Maria Massi, Eduardo Ros, Andrea...	2009	astro-ph	https://arxiv.org/abs/0704.3557

9.5. SBERT Öneri Çıktısı

```
idx = 9000
balanced_df.iloc[idx][
    ["title", "authors", "year", "main_category", "link"]
]

recommend_sbert(idx, top_n=5)
```

	title	authors	year	main_category	link
289	First real time detection of Be7 solar neutrin...	Borexino Collaboration	2012	astro-ph	https://arxiv.org/abs/0708.2251
8516	Study of Micro Pixel Photon Counters for a hig...	N.D'Ascenzo (University of Hamburg, DESY), A. ...	2007	physics	https://arxiv.org/abs/0711.1287
9818	Detectors and flux instrumentation for future ...	T. Abe, H. Aihara, C. Andreopoulos, A. Ankowsk...	2009	physics	https://arxiv.org/abs/0712.4129
9324	The Quest for the Ideal Scintillator for Hybri...	B.K.Lubsandorzhiev, B.Combettes	2009	physics	https://arxiv.org/abs/0710.2069
1193	Potential for Precision Measurement of Solar N...	Y.H. Huang, R.E. Lanou, H.J. Maris, G.M. Seide...	2009	astro-ph	https://arxiv.org/abs/0711.4095

9.6. Hibrit Modelin Popülarite Skoruyla Öneri Çıktıları

	Makale Başlığı	Kategori	Hibrit Puan (Sonuç)	İçerik Benzerliği (SBERT)	Popülarite Skoru	link
0	8.4GHz VLBI observations of SN2004et in NGC6946	astro-ph	0.4530	0.5438	0.8815	https://arxiv.org/abs/0705.3853
1	SMA Imaging of the Maser Emission from the H30...	astro-ph	0.4487	0.5227	0.9891	https://arxiv.org/abs/0801.0608
2	The Continuing Saga of the Explosive Event(s) ...	astro-ph	0.4431	0.5389	0.9474	https://arxiv.org/abs/0707.3124
3	Cosmic rays and Radio Halos in galaxy clusters...	astro-ph	0.4410	0.4972	0.9334	https://arxiv.org/abs/0710.0801
4	Fast outflows in compact radio sources: eviden...	astro-ph	0.4392	0.5192	0.7924	https://arxiv.org/abs/0802.1444
5	Observed flux density enhancement at submillim...	astro-ph	0.4319	0.6209	0.8564	https://arxiv.org/abs/0706.0012

Hibrit model tarafından önerilen ilk 5 makale ve bu makalelere ait skor bileşenleri sunulmuştur. Tüm önerilerin aynı ana kategori (astro-ph) altında toplanması, modelin disiplinler arası gürültüyü başarıyla elimine ettiğini göstermektedir. Hibrit skorun, içerik benzerliği ve popülarite bileşenleri arasında denge kurarak nihai sıralamayı oluşturduğu gözlemlenmiştir.

9.7. Nitel Değerlendirme 3 Modelin Karşılaştırmalı Önerileri

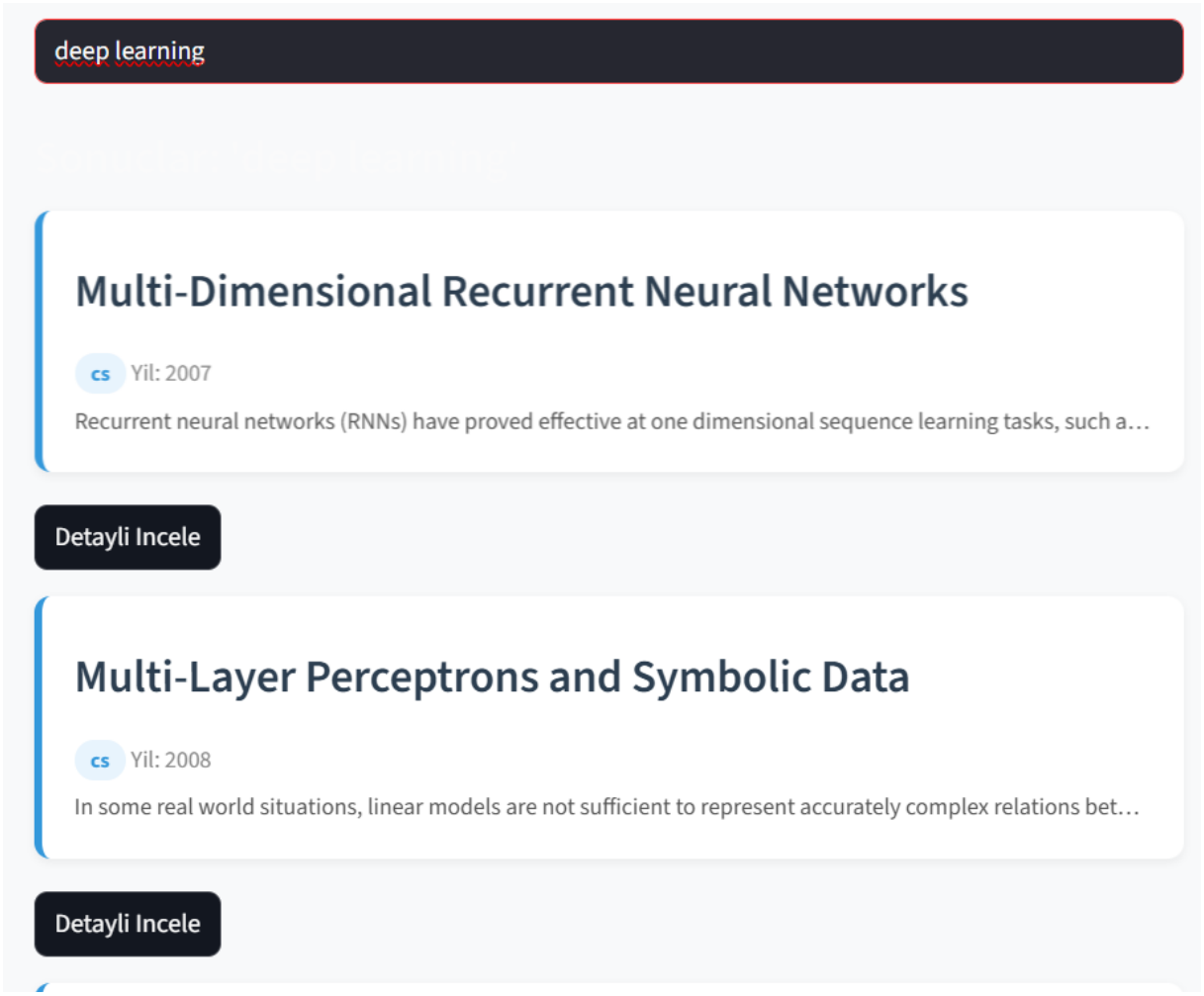
```
SORGULAMA: 'Deep Learning for Image Recognition'

SENARYO 1: TF-IDF MODELİ
1. Arabic Speech Recognition System using CMU-Sphinx4... | Skor: 0.3808
2. Introduction to Arabic Speech Recognition Using CMUSphinx System... | Skor: 0.3804
3. Learning Similarity for Character Recognition and 3D Object Recognitio... | Skor: 0.3785
4. Learning View Generalization Functions... | Skor: 0.3475
5. Supervised learning on graphs of spatio-temporal similarity in satelli... | Skor: 0.3173

SENARYO 2: SBERT MODELİ
1. Comparison and Combination of State-of-the-art Techniques for   Handwr... | Skor: 0.4452
2. Learning View Generalization Functions... | Skor: 0.4169
3. Multi-Dimensional Recurrent Neural Networks... | Skor: 0.3932
4. Comparing Robustness of Pairwise and Multiclass Neural-Network Systems... | Skor: 0.3750
5. Handbook for the GREAT08 Challenge: An image analysis competition for ... | Skor: 0.3673

SENARYO 3: HİBRİT MODEL (Kullanıcı: 1)
1. A Cross-Match of 2MASS and SDSS: Newly-Found L and T Dwarfs and an   E... | Skor: 0.5302
2. Keck/Deimos Spectroscopy of a GALEX UV Selecte Sample from the Medium ... | Skor: 0.5274
3. Signatures of the transition from galactic to extragalactic cosmic ray... | Skor: 0.5178
4. Herbig-Haro Objects - Tracers of the Formation of Low-mass Stars and  ... | Skor: 0.5137
5. Gamma Ray Bursts from the early Universe: predictions for present-day ... | Skor: 0.5116
```

9.8. Streamlit Makale Öneri Sistemi Arayüzü



0705.2011

Kutuphanede

Multi-Dimensional Recurrent Neural Networks

Yazarlar:	Alex Graves, Santiago Fernandez, Juergen Schmidhuber
Yil:	2007
Kategori:	cs
Popularite:	0.52 / 1.0
Kaynak:	ArXiv Linki

Ozet (Abstract)

Ozet (Abstract)

Recurrent neural networks (RNNs) have proved effective at one dimensional sequence learning tasks, such as speech and online handwriting recognition. Some of the properties that make RNNs suitable for such tasks, for example robustness to input warping, and the ability to access contextual information, are also desirable in multidimensional domains. However, there has so far been no direct way of applying RNNs to data with more than one spatio-temporal dimension. This paper introduces multi-dimensional recurrent neural networks (MDRNNs), thereby extending the potential applicability of RNNs to vision, video processing, medical imaging and many other areas, while avoiding the scaling problems that have plagued other multi-dimensional models. Experimental results are provided for two image segmentation tasks.

Multi-Layer Perceptrons and Symbolic Dat...

Goz At

An Optimal Linear Time Algorithm for Qua...

Goz At

Comparison and Combination of State-of-t...

Goz At

9.9. Proje Kaynak Kodları ve Veri Seti Erişimi

Proje kapsamında geliştirilen tüm algoritmalar, veri işleme adımları ve Streamlit arayüz kodlarına aşağıdaki bağlantılar üzerinden erişilebilir:

Kaggle Notebook (Çalışma Dosyası): https://www.kaggle.com/code/edanurdemirel/makale-neri-sistemi?scriptVersionId=291724440

Kullanılan Veri Seti (arXiv Dataset): https://www.kaggle.com/datasets/Cornell-University/arxiv/data

Streamlit Uygulama Dosyası: https://github.com/22eda/Makale_Oneri_Sistemi/

Referans kodlar:

https://www.kaggle.com/code/nqkkhnh/research-paper-recommendation

https://medium.com/web-mining-is688-spring-2021/article-recommendation-system-using-python-8b0fec6e6de8

https://medium.com/@prateekgaurav/step-by-step-content-based-recommendation-system-823bbfd0541c

https://medium.com/biased-algorithms/recommender-systems-projects-with-code-d4019adfe546

https://github.com/kaustubh187/Research-paper-recommender-system/blob/main/main.py