

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

AKADEMİK HUKUK MAKALELERİNDE ATIF ÖNERİSİ

YÜKSEK LİSANS TEZİ

Doğukan ARSLAN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

HAZİRAN 2023

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

AKADEMİK HUKUK MAKALELERİNDE ATIF ÖNERİSİ

YÜKSEK LİSANS TEZİ

**Doğukan ARSLAN
(504201515)**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Gülşen ERYİĞİT

HAZİRAN 2023

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

CITATION RECOMMENDATION ON SCHOLARLY LEGAL ARTICLES

M.Sc. THESIS

**Doğukan ARSLAN
(504201515)**

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Doç. Dr. Gülşen ERYİĞİT

JUNE 2023

İTÜ, Lisansüstü Eğitim Enstitüsü'nün 504201515 numaralı Yüksek Lisans Öğrencisi Dođukan ARSLAN, ilgili yönetmeliklerin belirlediđi gerekli tüm şartları yerine getirdikten sonra hazırladıđı "AKADEMİK HUKUK MAKALELERİNDE ATIF ÖNERİSİ" başlıklı tezini ařađıda imzaları olan jüri önünde başarı ile sunmuřtur.

Tez Danıřmanı : **Doç. Dr. Gülřen ERYİĐİT**
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Doç. Dr. Cüneyd TANTUĐ**
İstanbul Teknik Üniversitesi

Prof. Dr. Banu DİRİ
Yıldız Teknik Üniversitesi

Teslim Tarihi : **26 Mayıs 2023**
Savunma Tarihi : **22 Haziran 2023**





Daima yanımnda olan eşime, aileme ve dostlarıma,



ÖNSÖZ

Öncelikle lisans eğitimimden bu yana iyi bir arařtırmacı, öğretmen, yönetici ve hepsinden önemlisi iyi bir insan olarak kendime örnek aldığım değerli danışmanım Doç. Dr. Gülşen ERYİĞİT'e en içten teşekkürlerimi sunarım.

Ayrıca araştırma sürecinde büyük katkıları sunan Saadet Sena ERDOĞAN'a minnettarım. İTÜ Doğal Dil İşleme Araştırma Grubu ile Dil Teknolojileri ve Sosyal Robotik Laboratuvarında beraber çalışma fırsatı edindiğim bütün çalışma arkadaşlarıma destekleri için şükranlarımı sunuyorum.

En önemlisi, bana her zaman desteklerini hissettiren aileme, annem Yeliz ARSLAN'a, babam Şafak ARSLAN'a, kardeşim İrem ARSLAN'a ve Şıla ile Maggie'ye sonsuz teşekkürlerimi iletiyorum.

Ve destekleri, fikirleri, yardımları olmadan bu çalışmanın asla gerçekleşmeyeceği sevgili eşim Emel'e; teşekkürlerin en büyüğünü sunarım.

Haziran 2023

Doğukan ARSLAN

İÇİNDEKİLER

	<u>Sayfa</u>
ÖNSÖZ	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SEMBOLLER	xv
ÇİZELGE LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxv
1. GİRİŞ	1
1.1 Atıf Önerisi	2
1.2 Hukuk Dilinin Ayırt Edici Özellikleri	4
1.3 Tezin Katkısı	5
1.4 Tezin Bölümleri	6
2. ATIF ÖNERME ARAŞTIRMALARI	7
2.1 Yöntemler	7
2.1.1 İşbirlikçi filtreleme	7
2.1.2 Çizge temelli filtreleme	8
2.1.3 İçerik temelli filtreleme	9
2.2 Yaklaşımlar	10
2.2.1 Yerel atıf önerme	10
2.2.2 Küresel atıf önerme	11
2.2.3 Veri kümeleri	12
2.3 Hukuki Atıf Önerme	12
3. HUKUKİ ATIF ÖNERME	15
3.1 Yöntem	15
3.2 Veri Kümesi	20
3.3 Deney Düzenegi	20
3.4 Değerlendirme Ölçütleri	22
3.5 Sonuçlar	23
4. TARTIŞMA	27
4.1 BM25	27
4.1.1 BM25 türevleri	28
4.1.2 BM25 türevlerinin karşılaştırılması	30
4.2 Hukuki Dokümanlarla Eğitilen Dil Modelleri	32
4.3 Bilimsel Makalelerle Eğitilen Dil Modelleri	33
4.4 İki Aşamalı Bilgi Getirimi	34
4.5 Sonuç	34
5. SONUÇ VE ÖNERİLER	37
KAYNAKLAR	39
EKLER	51
EK A1	53
EK A2	54
EK B1	55
ÖZGEÇMİŞ	57



KISALTMALAR

DDİ	: Dođal Dil İřleme
BM25	: Best Matching-25
ds	: Doküman Sıklığı
MAP	: Ortalama Kesinliklerin Ortalaması
MRR	: Sıralamaların Terslerinin Ortalaması





SEMBOLLER

Q	: Sorgu dokümanı
q_i	: Sorgu dokümanında yer alan her bir kelime
D	: Sorgu ile benzerliği karşılaştırılan doküman
 D 	: Sorgu ile benzerliği karşılaştırılan dokümanın uzunluğu
 D _{ort}	: Sorgu ile benzerliği karşılaştırılan dokümanların ortalama uzunluğu





ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 3.1 : BM25 ve ön eğitilmiş modellerin ilk 10 makaleyi getirme performansı.....	23
Çizelge 3.2 : İnce ayar yapılmış modellerin ilk 10 makaleyi getirme performansı.	24
Çizelge 3.3 : BM25 tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ön eğitilmiş modellerin performansı.	25
Çizelge 3.4 : BM25 tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ince ayar yapılmış modellerin performansı.	25
Çizelge 4.1 : BM25+, Okapi BM25 ve BM25L modellerinin derlenen veri kümesindeki başarımları.....	30
Çizelge 4.2 : BM25 türevleri tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ön eğitilmiş modellerin performansı.	31
Çizelge 4.3 : BM25 türevleri tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ince ayar yapılmış modellerin performansı.....	32
Çizelge B.1 : Tez kapsamında kullanılan kavramların İngilizce literatürdeki karşılıkları.	55



ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 1.1 : Atıf Önerisi sistemlerinin genel yapısı.	2
Şekil 2.1 : İşbirlikçi filtrelemede benzer makalelere atıfta bulunan makaleler benzer olarak ele alınması.	8
Şekil 2.2 : Çizge temelli filtrelemede öneri için makalelerin, yayınladıkları dergilerin ve içerdikleri anahtar kelimelerin kullanılması.	9
Şekil 2.3 : Yerel atıf önerme modellerinde atıf bağlamından yararlanılması (Şekildeki alıntı Bölüm 2.1.2’den alınmıştır).	11
Şekil 3.1 : Birinci deney düzeneğinde önceden eğitilmiş modellerden elde edilen doküman temsillerinin, yine aynı şekilde elde edilen sorgu temsili ile olan kosinüs benzerliklerine göre sıralanması.	16
Şekil 3.2 : İkinci deney düzeneğinde ince ayar yapılmış modellerden elde edilen doküman temsillerinin, yine aynı şekilde elde edilen sorgu temsili ile olan kosinüs benzerliklerine göre sıralanması.	16
Şekil 3.3 : Üçüncü deney düzeneğinde BM25 tarafından getirilen 10 makalenin önceden eğitilmiş modellerle yeniden sıralanması.	16
Şekil 3.4 : Dördüncü deney düzeneğinde BM25 tarafından getirilen 10 makalenin ince ayar yapılmış modellerle yeniden sıralanması.	17
Şekil 3.5 : LawArXiv’da yer alan makalelerin bazı temel hukuk konularına göre dağılımı.	21
Şekil 3.6 : Eğitimlerde makalelerin öz kısımlarının ön işlemde geçirilerek kullanılması ve veri kümesinin üçte birlik kısmının test için ayrılması.	21
Şekil A.1 : Kanada Yüksek Mahkemesi’nin Örnek Kararı.	53
Şekil A.2 : Türkiye Cumhuriyeti Yargıtay’ının Örnek Kararı.	54



AKADEMİK HUKUK MAKALELERİNDE ATIF ÖNERİSİ

ÖZET

Hukuk ve Doğal Dil İşleme çalışmalarının kesişiminde, hukuki metinlerin anlaşılması, işlenmesi, yorumlanması ve üretilmesi gibi konulara odaklanan "Hukuki DDİ" çalışmaları yer alır ve bu çalışmalar farklı hukuki metin türleri üzerinde çeşitli alt görevlere odaklanmaktadır. Bu çalışmalardan biri de Atıf Öneri görevidir.

Atıf Önerisi, bilimsel makalelerde belirli bir metin için potansiyel atıfların belirlenmesi çalışmalarını kapsar. Ancak, bu görevdeki çalışmalarda, veri kümelerinin alan bazında yeterince kapsayıcı olmaması ve alanlara dengesiz dağılması gibi sorunlar genellikle ihmal edilmektedir. Son zamanlarda yapılan bir çalışmada, bu sorunlar ele alınmış ve farklı alanları kapsayan yeni bir veri kümesi oluşturulmuştur. Ancak, hukuk gibi bazı temel alanlar hala bu tür çalışmaların dışında kalmaktadır. Bu nedenle, Atıf Önerisi gibi alt görevlerde bile, büyük veri kümeleriyle eğitilen dil modelleri, alan bazında eksiklikler gösterebilmektedir.

Hukuki Doğal Dil İşleme bağlamında Atıf Önerisi, çoğunlukla mahkeme kararları gibi bilimsel olmayan hukuki metinlerden, var olan argümanları gerekçelendirmek için çeşitli atıfların elde edilmesini amaçlar. Hukuk sistemleri, Ortak Hukuk ve Kıta Avrupası Hukuk sistemi olmak üzere iki ana kategoriye ayrılabilir. Ortak Hukuk sistemine sahip ülkelerde, kararların sonuçları geçmiş davaların incelenmesiyle belirlenir ve bu nedenle kararlar arasında çok sayıda atıf bulunurken, Kıta Avrupası Hukuk sistemine sahip ülkelerde karar verme süreci daha çok olgusal kanıtlar ve ilgili kanun maddelerine dayanır. Bu da kararların kanunlara ve tüzüklere daha fazla atıf içermesine yol açar. Her iki sistemde de hukuk uygulayıcıları için emsal kararları bulmak önemlidir, ancak bu süreç zaman alıcı olabilir. Türkiye’de Yargıtay tarafından yayınlanan 7 milyondan fazla karar bulunmaktadır ve avukatlar, ilgili içtihatları aramak için önemli miktarda zaman harcamaktadır. Hukuki Atıf Önerisi görevinin halihazırdaki önemi ve faydaları, akademik hukuk metinlerinin gereken ilgiyi görmemesi ve görev kapsamına alınmamasıyla sonuçlanmıştır. Bununla birlikte, bilimsel makalelerden otomatik olarak atıf bilgisi çıkarılarak elde edilecek olan işaretli veri ile, etiketli veri oluşturmanın maliyetli olduğu Hukuki Doğal Dil İşleme görevleri için önemli bir kaynak oluşturulabilir. Bu yaklaşım, Atıf Önerisi görevinin yanı sıra emsal karar bulma, hukuki belge benzerliği ve hukuki karar tahmini gibi diğer görevlerde de etkili olabilir. Bu şekilde, akademik hukuk metinleri daha verimli bir şekilde kullanılarak daha iyi performans gösteren dil modelleri geliştirilebilir. Ayrıca, diğer bilimsel alanlardan farklı dilbilimsel özelliklere sahip olan hukuki metinler için özel bir ilgi gerekir. Geleneksel Atıf Önerisi görevinden ayrılan Hukuki Atıf Önerisi, bu özellikleri anlayabilen ve etkili atıf önerileri sunabilen dil modellerine ihtiyaç duyar.

Bilimsel yayıncılığın hızlı genişlemesiyle birlikte, atıfların güvenilirliği ve kalitesiyle ilgili endişeler ortaya çıkmış ve Atıf Önerme görevi zaman içinde önem kazanmıştır. Bu görev kapsamında işbirlikçi filtreleme, çizge temelli filtreleme ve içerik temelli filtreleme gibi yöntemler kullanılmaktadır. Farklı metin türleri, haberlerden patentlere ve yargı kararlarına kadar, Atıf Önerme görevinde kullanılmıştır. Görev, önerinin kapsamına bağlı olarak da genellikle yerel ve küresel olmak üzere iki ana kategoriye ayrılır. Çeşitli akademik makale veri kümeleri, Atıf Önerme tekniklerinin geliştirilmesi ve test edilmesi için kullanılmıştır. Atıf Öneri yöntemleri, akademik olmayan hukuki metinleri (mahkeme kararları, tüzükler, atıfta bulunan yasalar vb.) tespit etmek amacıyla hukuk alanına uyarlanmaktadır. Bu uyarlamalar, Hukuki Atıf Önerme görevi adı altında gerçekleştirilmektedir.

Tez kapsamında Hukuki Atıf Öneri görevi için, akademik hukuki makalelerden oluşan bir veri kümesi toplanmıştır. Bu veri kümesi, Atıf Önerme ve ilgili görevlerde iyi performans gösteren veya hukuk alanında eğitilmiş toplamda yedi farklı modelin test edildiği dört farklı deney düzeninde kullanılmıştır. Gerçekleştirilen deneylerde, yedi farklı model için dört farklı deney düzeni kullanılarak, önceden eğitilmiş modellerin doğrudan kullanılması, modellere ince ayar yapılması ve BM25 ile ilgili makalelerin çekilmesiyle birlikte yeniden sıralanması üzerinde çalışmalar yapılmıştır. Benimsenen iki aşamalı yaklaşım, dil modellerinin hantallığını azaltmak için BM25 gibi daha hızlı ancak daha az doğruluk gösteren modelleri kullanarak makale örneklerini hızlı bir şekilde seçmeyi amaçlar. Bu yaklaşım, bilgi getiri çalışmaları sistem etkinliğini artırmak için sıkça kullanılır. İlk aşamada, hızlı modellerle ilgili belgelerin örneklerini alırken, daha sonra yavaş ancak daha doğru olan modellerle bu aday makaleler yeniden sıralanır.

İngilizce hukuki atıf önerme görevi için LawArXiv adlı hukuki bilimsel makaleler veritabanından makaleler indirilmiştir. Bu veritabanı, 1366 bilimsel hukuki makaleye sahip olan ve çeşitli hukuki konuları kapsayan bir kaynaktır. Makalelerin atıf yapılan kaynakları elde etmek için Google Scholar kullanılmış ve 10 binden fazla atıf içeren makale elde edilmiştir. Elde edilen makalelerin öz kısmı pdfplumber adlı bir Python paketi ile çıkarılmış, ardından başarılı bir şekilde çıkarılan İngilizce makaleler seçilmiştir. Ön işleme adımlarıyla makaleler düzenlenmiş ve öz kısımları çıkarılmıştır. Deneylerde 719 LawArXiv makalesi ve 8,887 atıf içeren 10,111 atıf bağlantısı içeren bir veri kümesi kullanılmıştır. Makalelerin öz kısımları, benzer içerik temelli küresel atıf önerme çalışmalarıyla uyumlu bir şekilde, ince ayar, temsil elde etme ve test aşamalarında girdi olarak kullanılmıştır. Veri kümesi, eğitim ve test olarak ayrılmış olup, verilerin %70'i eğitimde kullanılmış ve kalan %30'u test için ayrılmıştır. İnce ayar aşamasında üçlü kayıp fonksiyonu kullanılmıştır. Bu fonksiyon referans girdiyi (çapa) pozitif bir girdiyle (benzer) ve çapayla eşleşmeyen negatif bir girdiyle karşılaştırır. İnce ayar ve temsil elde etme adımlarından sonra, belge temsil vektörleri vektör uzayında benzerliklerine göre sıralanmıştır. Tüm eğitim ve test süreçlerinde Sentence-Transformers çerçevesi kullanılmıştır. Deneylerin sonuçları, bilgi getiri çalışmaları yaygın olarak kullanılan üç farklı metrik olan Mean Average Precision (MAP) (Ortalama Kesinliklerin Ortalaması), Recall (Duyarlılık) ve Mean Reciprocal Rank (MRR) (Sıralamaların Terslerinin Ortalaması) kullanılarak sunulmuştur. Bu metrikler, bir makalenin ortalama olarak 14 atıf bağlantısına sahip olduğu göz önüne alınarak, getirilen ilk 10 belge için (n=10) raporlanmıştır.

Önceden eğitilmiş çeşitli modeller ve derlenen veri kümesi eğitilmiş BM25 modelinin karşılaştırması, SciBERT'in diğer modellere kıyasla en düşük performansı gösterdiği, Law2Vec ve LegalBERT gibi hukuki derlemelerle eğitilen modellerin atıf önerme görevinde başarısız olduğu, SGPT'nin ise SPECTER ve SciBERT'ten daha iyi performans gösterdiği ancak BM25'in en başarılı model olarak öne çıktığı sonucunu ortaya koymuştur. Bu sonuçlar, literatürdeki bilimsel alan temelli Atıf Önerme çalışmalarıyla da uyumludur. Önceden eğitilmiş modellere ince ayar yapıldığında elde edilen sonuçlar incelendiğinde, modellerin genel olarak benzer performans sergilediği ancak BM25'i geçemediği görülmektedir. Bununla birlikte, ince ayarlı LegalBERT modelinin performansının önemli ölçüde arttığı, modelin göreve aşinalığının alan bilgisiyle birleşmesinin performansı artırdığı gözlemlenmiştir. En başarılı modeller arasında SciNCL ve SciBERT öne çıkmaktadır, SciBERT'in performansındaki sıçrama dikkat çekicidir. Önceden eğitilmiş modellerin sıralama yeteneklerini BM25'in geri getirme kapasitesiyle birleştiren deneylerin sonuçları önceden eğitilmiş modellerin BM25'in performansını artıramadığını gösterse de, SciNCL'nin tartışmasız olarak en başarılı model olduğunu ortaya koymaktadır. BM25 ile getirilen makalelerin ince ayarlı modellerle yeniden sıralanması sonucunda, tüm ince ayarlı modellerin BM25'in performansını artırdığı gözlemlenmekte olup, SciNCL'in diğer deneylerle uyumlu olarak en başarılı model olduğu görülmektedir (0.30 MAP@10).

Bu çalışmada, İngilizce Hukuki Atıf Önerisi veri kümesi oluşturulmuş ve Atıf Önerisi görevinde başarılı modeller ile alana özel eğitilmiş modellerin performansları karşılaştırılmıştır. Ayrıca, iki aşamalı bilgi getirme yöntemi kullanılmıştır. Sonuçlar, öne sürülen hipotezlerin doğruluğunu desteklemektedir. Dil modellerinin Hukuki Atıf Önerisi görevinde başarılı olabilmesi için akademik hukuk makalelerine yer verilmesi gerektiği ortaya çıkmıştır. Aynı şekilde, hukuki dokümanlarla eğitilen modellerin daha kapsayıcı olabilmesi için akademik hukuk makalelerinin de eğitim veri kümesinde bulunması gerektiği gösterilmiştir. İki aşamalı bilgi getirme yöntemi, büyük dil modellerinin ve BM25'in en iyi yönlerini birleştirerek genel performansı artırmaktadır. BM25 ile SciNCL'in birlikte kullanılması, Hukuki Atıf Önerisi görevinde en başarılı sonuçları vermektedir. Gelecek çalışmalar açısından, iki aşamalı bilgi getirme yöntemi önemli bir araştırma alanıdır. Ayrıca, elde edilen Hukuki Atıf Önerisi modelinin farklı hukuki görevlere uygulanması ve başarımlarının test edilmesi önemlidir. Veri kümesinin boyutunu artırmak için çeşitli çalışmalar da yapılabilir. Özellikle veri kümesi büyüdükçe, BM25'in hızı ve performansı daha iyi değerlendirilebilir.



CITATION RECOMMENDATION ON SCHOLARLY LEGAL ARTICLES

SUMMARY

At the intersection of law and natural language processing is the study of "Legal NLP", which focuses on the understanding, processing, interpretation, and generation of legal texts, and performs various subtasks on different types of legal texts. One of these tasks is Citation Recommendation.

Citation Recommendation is a method for identifying potential citations for a given text in scholarly articles. However, studies in this area often neglect problems such as domain-inclusive data sets and uneven distribution across domains. In a recent study, these problems were addressed and a new dataset covering different domains was created. However, some core domains, such as law, are still excluded from such studies. Therefore, even in subtasks such as citation recommendation, language models trained on large datasets may show domain-specific deficiencies. In the context of Legal Natural Language Processing, Citation Recommendation is a method that aims to extract various citations from non-scientific legal texts, such as court decisions, to justify existing arguments.

Legal systems can be divided into two main categories: Common Law and Civil Law. In countries with a Common Law system, the outcomes of judgments are determined by examining past cases and therefore there are a large number of citations between judgments, whereas in countries with a Civil Law system, the decision-making process is based more on factual evidence and relevant articles of law, resulting in judgments that contain more citations to laws and statutes. In both systems, legal practitioners need to find precedents, but this process can be time-consuming. In Turkey, there are more than 7 million judgments published by the Court of Cassation. Lawyers spend a lot of time searching for relevant case law.

The current importance and utility of the Legal Citation Recommendation task has resulted in academic legal texts not receiving attention and not being included in the scope of the task. However, by automatically extracting citation information from scholarly articles, labeled data could be a valuable resource for Legal Natural Language Processing tasks where the generation of labeled data is costly. In addition to the Citation Recommendation task, this approach can be effective in other tasks such as case retrieval, legal document similarity, and legal decision prediction. In this way, scholarly legal documents can be used more efficiently to develop more powerful language models. Legal texts, which have different linguistic features than other scientific fields, require special attention. Legal Citation Recommendation, which is different from the traditional Citation Recommendation task, requires language models that understand these features to provide effective recommendations.

With the rapid expansion of scholarly publishing, concerns about the reliability and quality of citations have emerged and the task of Citation Recommendation has gained importance. Methods such as collaborative filtering, graph-based filtering and content-based filtering are used in this task. Different text types have been used in the Citation Recommendation task, ranging from news articles to patents and judicial decisions. Depending on the scope of the recommendation, the task is usually divided into two main categories: local and global. Various scholarly article datasets have been utilized to develop and test Citation Recommendation techniques.

The Citation Recommendation methods are adapted to the legal domain to identify non-academic legal texts (court decisions, statutes, cited laws, etc.). These adaptations are performed under the name of the Legal Citation Recommendation. In this thesis, a dataset of scholarly legal articles was collected for the Legal Citation Recommendation task. This dataset was used in four different experimental settings in which a total of seven different models, either trained in the legal domain or performing well on the Citation Recommendation and related tasks, were tested.

For the seven different models, four different experimental setups were used: direct use of pre-trained models, fine-tuning, and ranking them, as well as retrieving articles related to BM25. Adopted two-stage approach aims to quickly select article instances using faster but less accurate models such as BM25 to reduce language models' cumbersomeness. This approach is often used in knowledge retrieval studies to improve system efficiency. In the first stage, it retrieves samples of relevant documents with fast models, and then re-ranks these candidate articles with slower but more accurate models.

For the English Legal Citation Recommendation task, articles were downloaded from a database of legal scholarly articles called LawArXiv. This database contains 1366 scholarly legal articles, covering a variety of legal topics. Google Scholar was used to obtain the cited sources of the articles and more than 10 thousand cited articles were obtained. The context of the articles was extracted with a Python package called pdfplumber, and then the successfully extracted English articles were selected. The experiments used a dataset of 719 LawArXiv articles and 10,111 citation links containing 8,887 citations. The abstracts of the articles were used as input in the fine-tuning, representation retrieval and testing phases, in line with similar content-based Global Citation Recommendation studies. The dataset was split into training and testing, with 70% of the data used for training and the remaining 30% for testing. In the fine tuning phase, a triple loss function was applied, which compares the reference input (anchor) with a positive input and a negative input that does not match the anchor. After the fine-tuning and representation extraction steps, the document representation vectors are ranked according to their similarity in the vector space. The Sentence-Transformers framework was used for all training and testing phases. The results of the experiments are presented using Mean Average Precision (MAP), Recall and Mean Reciprocal Rank (MRR), three different metrics commonly used in information retrieval studies. These metrics are reported for the first 10 retrieved documents (n=10), considering that on average an article has 14 citation links.

A comparison of various pre-trained models and the BM25 model trained on the corpus dataset reveals that SciBERT performs the worst compared to the other models,

models trained on legal corpora such as Law2Vec and LegalBERT perform poorly in the Citation Recommendation task, SGPT outperforms SPECTER and SciBERT, but BM25 stands out as the most successful model. These results are in line with scientific domain-based citation recommendation studies in the literature. When the results obtained by fine-tuning the pre-trained models are analyzed, it is seen that the models perform similarly in general but do not outperform BM25. However, it was observed that the fine-tuned LegalBERT model improved significantly, and that the model's familiarity with the task combined with domain knowledge improved performance. Among the most successful models, SciNCL and SciBERT stand out, with SciBERT's performance jump being remarkable. The results of experiments combining pre-trained models' ranking capabilities with BM25's retrieval capacity show that pre-trained models cannot improve BM25's performance, but SciNCL is arguably the most successful model. Re-ranking the articles retrieved by BM25 with fine-tuned models shows that all fine-tuned models improve BM25 performance, with SciNCL being the most successful model (0.30 MAP@10), in line with the other experiments.

In this study, an English Legal Citation Recommendation dataset is created. The performance of language models trained on scholarly articles and language models trained in the legal domain is compared on the Citation Recommendation task. Furthermore, a two-stage information retrieval method is used. Also, the results support the hypotheses. In order for the language models to be successful in the Legal Citation Recommendation task, scholarly law articles should be included. Likewise, it has been shown that scholarly law articles should also be included in the training dataset for the models trained with legal documents to be more inclusive. The two-stage information retrieval method improves overall performance by combining the most powerful aspects of large language models and BM25. The combination of BM25 with SciNCL gives the most reliable result on the Legal Citation Recommendation task. In terms of future work, the two-stage information retrieval method is a significant area of research. Furthermore, it is important to apply the obtained Legal Citation Recommendation models to different legal tasks and test its performance. Various studies can also be conducted to increase the dataset size. As the dataset grows, BM25's performance can be better evaluated.

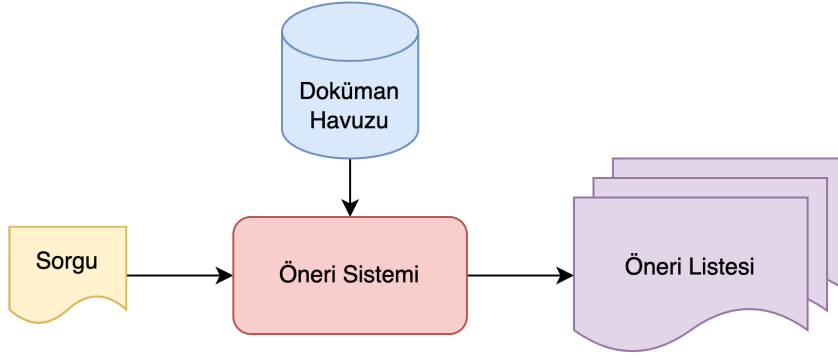


1. GİRİŞ

Hukuk, çeşitli doğal dil işleme (DDİ) tekniklerinin uzun süredir uygulandığı bir araştırma alanıdır [1]. Bu iki alanın kesişimindeki çalışmalar başta mahkeme kararları [2], tüzükler [19], makaleler [3], yönetmelikler [20], sözleşmeler [22] ve patentler [21] olmak üzere çeşitli hukuki metinlerin anlanması, işlenmesi, yorumlanması ve üretilmesi gibi konulara odaklanan "Hukuki DDİ" etiketi altında gruplandırılır. Bu belge türleri, hukuki metin ilişkilendirme [24], hukuki gerekçelendirme [25], hukuki soru cevap [23], hukuki karar tahmini [26], tüzük bulma [27] ve içtihat bulma [28] gibi çok sayıda alt görevde kullanılmaktadır.

Hukuki metin ilişkilendirme, incelemeye tabii tuttuğu iki belgeden birinin diğerini içerip içermediğini yahut bu iki metnin ilişkili olup olmadığını bulmayı amaçlar [24]. Örneğin, bir hukuki dokümanın içerdiği yasa, tüzük ve mevzuat maddelerinin tespit edilmesi için kullanılabilir. Hukuki gerekçelendirme görevi ise herhangi bir durum için hukuki sorunların analiz edilip gerekli argümanların oluşturulması ve hukuki kararların gerekçelendirilmesi gibi bir dizi alt görevden oluşur [25]. Hukuki soru cevap görevinde ise doğal dil biçiminde sorulan hukuki bir sorunun çeşitli kaynaklardan (yasalar, mevzuat, mahkeme kararları vs.) yararlanılarak cevaplanması hedeflenir [23]. Hukuki karar tahmininde, bir davadaki mevcut bilgilere (kanıtlar, ifadeler, önceki kararlar vs.) dayanarak, karar veya hüküm tahmin edilmeye çalışılır [26]. Tüzük ve içtihat bulma görevleri ise var olan bir hukuki dokümana benzer, bu dokümanla ilişkili tüzüğü veya mahkeme kararını bulmayı içerir [27,28]. Hukuki DDİ görevleri genel olarak birbiriyle oldukça ilişkili olmakla beraber amaç ve konu edilen hukuki doküman türü yönünden çok çeşitlilik göstermektedir. Hukuki Atıf Önerisi görevi de bunlardan biridir.

Aşağıda, Atıf Önerisi görevi ve hukuk alanına olan yansımaları tartışılmış ardından tezin cevap aradığı sorular, bu soruların çözümlenmesinin hangi açılardan önemli



Şekil 1.1 : Atıf Önerisi sistemlerinin genel yapısı.

olduğu ve tezin literatüre olan katkısı açıklanmıştır. Tezin bu kısmı diğer bölümlerin kısaca açıklanıp listelenmesi ile sonlanmaktadır.

1.1 Atıf Önerisi

Atıf Önerisi görevi, bilimsel makalelerde, özellikle argümanları gerekçelendirmek veya sunulan kavramları netleştirmek amacıyla, belirli bir metin için var olan aday metin havuzundan potansiyel atıfların belirlenmesi olarak tanımlanabilir (Şekil 1.1). Ağırlıklı olarak bu görevdeki çalışmalar, modellerin performansını etkileyebilecek olan, veri kümelerindeki makalelerin ait oldukları alanların yeterince kapsayıcı olmaması ve dengesiz dağılması gibi konuları ihmal etmektedir. Bu sorun yakın zamanda [41]'de ele alınmış ve çeşitli modellerin alan düzeyinde değerlendirilmesinin yanı sıra farklı bilimsel alanları kapsayan ve bu görev için sunulacak modellerin alan düzeyinde değerlendirilmesine yardımcı olacak yeni bir veri kümesi oluşturulmuştur. Buna rağmen hukuk gibi bazı temel alanlar bu ve benzeri çalışmaların kapsamı dışında kalmıştır. Atıf Önerisi görevi dahil birçok alt görevde kullanılmak üzere, bilimsel makalelerden oluşan büyük veri kümeleri ile eğitilen dil modellerinde dahi durum bu şekildedir.

Hukuki Doğal Dil İşleme bağlamında Atıf Önerisi temel olarak, bilimsel olmayan hukuki makalelerden, çoğunlukla mahkeme kararları için, bu kararı gerekçelendirecek argümanları önceden verilen mahkeme kararlarından elde edebilmek için kullanılır [11,51,104]. Dünya genelindeki hukuk sistemleri Ortak (Anglo-Sakson) Hukuk Sistemi ve Kıta (Kara) Avrupası Hukuk Sistemi olarak ikiye ayrılabilir. İngiltere,

Amerika Birleşik Devletleri ve Kanada gibi ülkelerde uygulanan Ortak Hukuk Sisteminde, bir kararın sonucu, karşılaştırılabilir geçmiş davalar incelenerek belirlenir. Bu nedenle, verilen kararlar önceki davalara birçok atıf içerir (bkz. Ek A'daki örnek Kanada Yüksek Mahkemesi kararı). Öte yandan, Fransa, Almanya, Japonya ve Türkiye gibi ülkelerde uygulanmakta olan Kıta Avrupası Hukukunda, yargının karar verme süreci esas olarak hem olgusal kanıtların hem de ilgili kanun maddelerinin dikkate alınmasını içerir [5], bu da mahkeme kararlarının kanunlara ve tüzüklere bolca atıf içermesiyle sonuçlanır (bkz. Ek A'deki örnek Türkiye Cumhuriyeti Yargıtay kararı). Her iki sistemde de hukuk uygulayıcıları için hukuki argümanlarını güçlendirebilecek emsal kararlar bulmak önemlidir fakat arama yapılan veri devasa boyutlara ulaşmış durumdadır. Türkiye'de, hem ceza hem de hukuk davalarında alt mahkemeler tarafından verilen kararları inceleme yetkisine sahip olan Yargıtay tarafından bugüne kadar tek başına 7 milyondan fazla karar yayınlanmıştır. Karar sayısı çok fazla olduğu için doğru kararları bulmak zaman almaktadır. Bir avukat ortalama olarak haftada yaklaşık 15 saatini, yani çalışma saatlerinin yaklaşık %30'unu ilgili içtihatları aramak için harcamaktadır [4].

Hukuki Atıf Önerisi görevinin mahkeme kararları düzeyinde bu denli önemli olması, akademik hukuk metinlerinin şu ana kadar gerekli ilgiliyi görmemesi ve görev kapsamına alınmamasına neden olmuştur. Bu doğrultuda, hukuk alanı için eğitilen dil modellerinin eğitim kümelerinde sıklıkla akademik hukuk makalelerinin dışında kalan mahkeme kararları, tüzükler, yönetmelikler vb. kullanıldığı görülmektedir. Halbuki bilimsel makalelerden atıf bilgisi çıkarılarak, otomatik ve hızlı bir biçimde etiketli veri oluşturmak mümkün olduğundan, bilimsel hukuk makaleleri üzerinde eğitilen modeller sadece Atıf Önerisi görevi için değil, emsal karar bulma [105], hukuki belge benzerliği [106] ve hukuki karar tahmini [107] dahil olmak üzere etiketli veri elde etmenin maaliyetli olduğu çeşitli hukuki doğal dil işleme görevleri için önemli bir kaynak oluşturma potansiyeline sahiptir. Bu kapsamda ele alındığında, bu tezin cevap bulmaya çalıştığı araştırma soruları şu şekilde özetlenebilir:

- Halihazırda bilimsel makalelerle önceden eğitilmiş dil modelleri, Hukuki Atıf Önerisi görevinde ne kadar başarılıdır? Modellere Atıf Önerisi görevi için ince

ayar yapılması başarımlarını ne derece artırmaktadır? Eğitim kümelerinde hukuki metinlere yer vermemeleri başarımlarını etkilemekte midir?

- Halihazırda hukuki metinlerle önceden eğitilmiş dil modelleri, Hukuki Atıf Önerisi görevinde ne kadar başarılıdır? Modellere Atıf Önerisi görevi için ince ayar yapılması başarımlarını ne derece artırmaktadır? Bu modeller akademik dille yazılmış hukuk metinlerinin temsili konusunda da başarılı sonuçlar verebilmekte midir?

1.2 Hukuk Dilinin Ayırt Edici Özellikleri

Hukuki Atıf Önerisi görevinin, daha önce sıralanan potansiyel kullanım alanları sıralandıktan sonra, geleneksel Atıf Önerisi görevinden hangi açılardan ayrıştığı ve neden özel bir ilgiyi hak ettiği sorusu da gündeme gelmektedir. Akademik hukuk makaleleri de dahil olmak üzere hukuki metinler, diğer birçok bilimsel alandan farklı olarak, kendine has birtakım dilbilimsel özellik barındırırlar. Aşağıda da detaylı incelenen bu özellikler, Hukuki Atıf Önerisi görevine geleneksel Atıf Önerisi görevi ile benzer şekilde yaklaşım yaklaşımamayacağı konusunda soru işaretleri oluşturmaktadır. Bu açıdan bakıldığında tez, bu soru işaretlerini de gidermeye çalışmaktadır.

Hukuki metinler incelendiğinde genel olarak şu özelliklerin hukuki metinlerdeki dili, günlük konuşma dilinden ayırdığı söylenebilir:

Yabancı sözcük kullanımı: Hukuki metinlerde sık sık yabancı dillerden alınan sözcüklere rastlanılmaktadır. İngilizce için bu, günlük konuşma dilinde karşılaşılmayan Fransızca ve Latince ekseriyetli ifadeler olurken Türkçe hukuki metinlerde yabancı sözcükler genellikle Arapça ve Farsça olarak karşımıza çıkmaktadır [108,109]. Bu ifadeler yabancı dillerden o dile yerleşen kelimeler değil, o dillerden doğrudan alınıp hukuki metinlerde özel olarak kullanılan ifadelerdir. Örneğin İngilizce için bunlara örnek olarak *certiorari* (Lat., kesinleştirmek), *procedendo* (Lat., devam etmek) ve *voir dire* (Fr., doğruyu konuşmak) verilebilir. Bu tür sözcükler akademik hukuk makalelerinde de sıkça yer almaktadır. Benzer bir şekilde yabancı bir dilden geçmeyen

fakat dilde sadeleşme sonucu günlük hayatta artık sık kullanılmayan bazı kelimeler de hukuki metinlerde yer alabilmektedir [110].

Sözcüklerin anlamı: Hukuki metinlerde sözcükler günlük konuşma dilindeki ilk anlamlarının dışında sıklıkla kullanılabilir [110]. Örneğin ilk anlamı "cümle" olan "*sentence*" kelimesi hukuki metinlerde "ceza, hüküm" anlamlarında kullanılmaktadır.

Hukuki terimler: Diğer bilimsel alanlarda olduğu gibi hukukta da sıklıkla o alanda anlam kazanan terimler mevcuttur ve bu terimler günlük konuşma dilinde pek yer almamaktadır.

Cümle yapısı: Hukuki metinler, birden fazla bağlaçla birbirine bağlanmış çok sayıda yan cümleden meydana gelen, zarf tümleçlerinin bolca kullanıldığı, uzun cümleler içerir [110]. Uzunluk hukuki metnin türüne göre değişmekle beraber [109], sadece İngilizce için değil Türkçe ve diğer birçok dil için de bu durum geçerlidir. Buna ek olarak cümlelerde kelime yapısı günlük konuşma dilinden farklı bir hal almakta, kişisiz ve edilgen yapıları cümleler kurulmakta, cümledeki kelimelerin sırası normalden farklı olabilmektedir. Öte yandan dilbilgisi açısından da hukuki metinlerdeki cümlelerin kendine has bir yapıyı takip ettiği söylenebilir.

1.3 Tezin Katkısı

Yukarıda verilen bilgiler ışığında tez kapsamında aşağıdaki çalışmalar yapılmıştır:

- Akademik hukuk makalelerinden oluşan ilk İngilizce veri kümesi derlenmiştir.
- Bilimsel makalelerle ve hukuki metinlerle eğitilmiş en gelişmiş dil modellerinin derlenen veri kümesi üzerinde Atıf Önerisi görevi için başarıları ölçülmüştür.
- Bu modeller ince ayar yapılarak hukuk alanına ve Atıf Önerisi görevine uyarlanmıştır.

1.4 Tezin Bölümleri

Tezin bölümleri aşağıdaki şekilde düzenlenmiştir:

- *Bölüm 2: Atıf Önerme Araştırmaları*

Bu bölümde, Atıf Önerisi görevine ilişkin literatürün mevcut durumu, görevin kategorizasyonu ve ilgili çalışmalar da dahil olmak üzere görev genel hatlarıyla özetlenmektedir.

- *Bölüm 3: Hukuki Atıf Önerme*

Bu bölümde, Hukuki Atıf Önerisi görevi için toplanan veri kümesi, kullanılan yöntemler, deney düzeneği ve elde edilen sonuçlar açıklanmaktadır.

- *Bölüm 4: Tartışma*

Bu bölümde deneyler sonucu elde edilen sonuçların detaylı analizine yer verilmektedir.

- *Bölüm 5: Sonuç ve Öneriler*

Bu bölümde ulaşılan bulgular toparlanmış ve gelecek çalışmalar için önerilere yer verilmiştir.

2. ATIF ÖNERME ARAŞTIRMALARI

Bu bölümde, Atıf Önerme görevinin literatürdeki genel durumu, bu görevin kavramsal kategorizasyonu ile birlikte özetlenmekte ve bu görev üzerine yapılan çalışmalar sunulmaktadır.

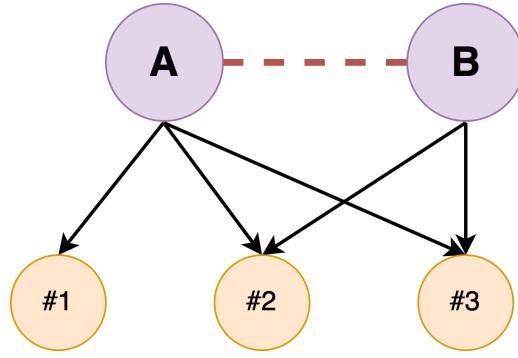
Atıf Önerme, bir dizi aday arasından belirli bir metin için uygun atıfları belirlemeyi amaçlayan bir bilgi erişim görevidir. Özellikle kavramları netleştirmek veya argümanları desteklemek için kullanışlıdır. Bilimsel yayıncılığın hızlı bir şekilde genişlemesiyle paralel olarak [6], son zamanlarda yayınlanan makalelerin geçmişte yayınlanan makalelerden daha fazla alıntı içermesi, bu makalelerin atıflarının güvenilirliği ve kalitesiyle ilgili birtakım endişeler ortaya çıkarmıştır [7]. Bu nedenle, atıf önerme görevi daha fazla çalışılan bir konu haline gelmiştir. Zaman içinde haberler [8], Vikipedi makaleleri [9], patentler [10] ve yargı kararları [11] gibi çok çeşitli metin türleri bu görev için kullanılmıştır.

2.1 Yöntemler

Atıf Önerme görevinde kullanılan yöntemler kabaca işbirlikçi filtreleme, çizge temelli filtreleme ve içerik temelli filtreleme olarak kategorize edilebilir.

2.1.1 İşbirlikçi filtreleme

Bilgi erişim görevlerinde işbirlikçi filtreleme algoritmaları, farklı kullanıcıların tercihlerini eşleştirerek öneride bulunmaya çalışır. Bu yaklaşım, özellikle veri sınırlı olduğunda veya yeni eklenen bir öğenin kullanıcılardan çok az geri bildirim aldığı veya hiç olmadığı (ilk oylayan problemi) veya bilinmeyen bir kullanıcı sisteme girdiğinde (soğuk başlangıç problemi) veya kötü niyetli bir kullanıcı algoritmayı bozmayı amaçladığında (şilin saldırısı) kırılabilir [12].



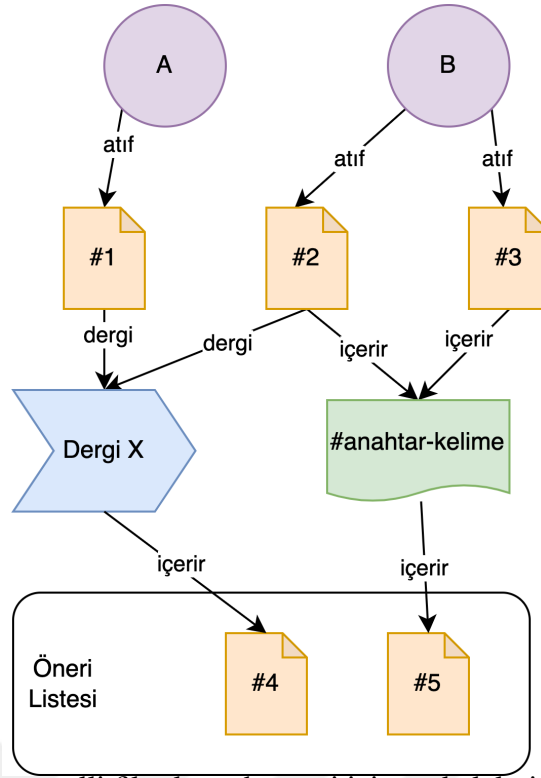
Şekil 2.1 : İşbirlikçi filtrelemede benzer makalelere atıfta bulunan makaleler benzer olarak ele alınması.

Atıf önerme bağlamında, atıfta bulunan makaleler ve atıfta bulunulan makaleler, geleneksel tavsiye sistemlerindeki kullanıcılar ve öğeler arasındaki ilişkiye benzer şekilde ele alınır. Bu nedenle, iki atıfta bulunan makale (kullanıcı) arasındaki benzerlik, atıfta bulunulan ortak makaleler (kullanıcı tercihleri) aracılığıyla belirlenir [13]–[15,66] (Şekil 2.1).

[13] tarafından önerilen içerik temelli işbirlikçi filtreleme yaklaşımı, makaleler arasındaki benzerlikleri hesaplamak için makalelerin atıf bağlamlarından türetilen ikili temsillerini kullanır. İçerik temelli işbirlikçi filtreleme yaklaşımının arkasındaki mantık, "aynı makalelere atıfta bulunan makalelerin benzer kabul edildiği" varsayımına dayanmaktadır. [15] makaleler arasında daha derin bağlantılar çıkarmayı amaçlayan iki seviyeli makale-atıf ilişkileri yaklaşımını önerir. [66]'de, bir sıralamayı öğrenme yöntemi, öğe temelli bir işbirlikçi filtreleme yaklaşımı ile birleştirilmiştir. Yazarlar bunun işbirlikçi filtrelemenin bazı eksikliklerini giderdiğini savunmaktadır. [14], her iki yaklaşımın avantajlarını birleştiren bir atıf öneri sistemi oluşturmak için hem işbirlikçi filtreleme hem de ağ temsili öğrenme tekniklerini kullanır.

2.1.2 Çizge temelli filtreleme

Çizge temelli filtrelemede çizge ve yazarlar, makaleler, mekanlar vb. arasındaki ilişkiler kullanılarak tavsiye sistemi modelleyen algoritmalar kullanılır (Şekil 2.2). Çizge temelli filtreleme yöntemlerinde ortaya çıkabilecek zorluklar arasında hesaplama karmaşıklığı [12] ve ağ içindeki eski düğümlere karşı potansiyel önyargı [16] yer almaktadır. Çoğu çizge temelli filtreleme çalışmasında, kullanılan yaklaşımlar hibrit



Şekil 2.2 : Çizge temelli filtrelemede öneri için makalelerin, yayımlandıkları dergilerin ve içerdikleri anahtar kelimelerin kullanılması.

bir yöntem oluşturmak için bir işbirlikçi filtreleme veya içerik temelli filtreleme yaklaşımı ile birleştirilir [17].

Yeni bir ortak atıf tabanlı örnekleme stratejisi kullanan bir çizge temelli filtreleme yaklaşımı [67] tarafından sunulmuştur ve çizge temelli filtrelemenin Atıf Önerme görevi için sağlam bir çözüm olduğu sonucuna varmışlardır. [68] bilgi çizgelerini genel Atıf Önerisi bağlamında kullanmaktadır.

2.1.3 İçerik temelli filtreleme

Atıf öneri görevinde kullanılan içerik temelli filtreleme algoritmaları, ilgili makaleleri önermek için doğrudan makalelerin başlık, özet, cümle veya anahtar kelimeler gibi bilgilendirici özelliklerini kullanır. Bu çalışmada olduğu gibi, birçok tavsiye sistemi içerik temelli bir yaklaşım kullanmaktadır [18] çünkü kullanılan bu özellikleri elde etmek nispeten basit bir işlemdir.

Meta veri (örneğin, yazar) kullanımı, istenmeyen bir yan etki olarak, kendi kendine atıflara öncelik verebilir [29]. Yine de bu durum, atıf önerme modelinin performansını artırır ve yaygın olarak kullanılır [30,31].

Atıf Öneri görevinde kullanılan içerik temelli filtreleme algoritmaları için, [29] tarafından meta veri eksikliğinden etkilenmeyen sağlam bir içerik temelli filtreleme yaklaşımı önerilmiştir. [69], yazarın çalışmalarına yapılan atıfların sıklığı, başlıklar, özetler ve anahtar kelimeler dahil olmak üzere makalelerden çeşitli metinsel, meta veri ve meta yol tabanlı özellikleri çıkaran yeni bir derin sinir ağı yöntemi sunmaktadır. Ayrıca, [70] tarafından geleneksel bir veri madenciliği yöntemini içeriğe dayalı benzerlik bulmayı amaçlayan bir derin öğrenme modeliyle birleştiren hibrit bir yaklaşım sunulmuştur.

2.2 Yaklaşımlar

Atıf önerme araştırmaları, önerinin kapsamına göre iki geniş kategoriye ayrılabilir: yerel ve küresel.

2.2.1 Yerel atıf önerme

Sorgu metninin bir cümle veya küçük bir pasaj gibi belirli bir bölümü yerel atıf önerme görevinin birincil odak noktasıdır. Bağlama duyarlı atıf önerme olarak da adlandırılır. Örneğin, [32]'de yerel bağlamı elde etmek için, belgedeki atıf göstergesi ([#] veya "XXX et al." gibi ifadeler) bir [CIT] belirteci ile değiştirilir ve değiştirilen atıf göstergesinin çevresindeki 200 karakterlik bir pencere metinden seçilir.

[73] tarafından geliştirilen, dikkat mekanizması tabanlı bir kodlayıcı-kod çözücü modeli atıf bağlamı ile birlikte yazar ve dergi bilgilerini de kullanmaktadır. Yerel Atıf Önerme görevi [32]'de ön-geri ve yeniden sıralama olarak ikiye ayrılmıştır ve daha iyi bir hız-doğruluk dengesi için yeniden sıralayıcı olarak SciBERT ile hiyerarşik dikkat mekanizmalı kodlayıcı kullanılmıştır. [71], cümle temsilleri elde etmek için başlık ve alıntı bağlamını birleştirerek, GloVe ve ELMo temsilleri ile başlatılan BiGRU ve dikkat mekanizması içeren bir model kullanır.



Şekil 2.3 : Yerel atıf önerme modellerinde atıf bağlamından yararlanılması (Şekildeki alıntı Bölüm 2.1.2’den alınmıştır).

Ayrıca, [72]’ye göre, Yerel Atıf Önerme görevi için şu üç faktör çok önemlidir: ön-getirme modelinin performansı, ön-getirme modelinin doğru bir şekilde aldığı atıf bağlamlarının seçilen bir alt kümesiyle yeniden sıralayıcı modeli eğitmek ve etkili bir negatif örnekleme stratejisi.

2.2.2 Küresel atıf önerme

Yerel Atıf Önerme görevinin aksine, Küresel Atıf Önerme çalışmalarının ana odağı belgenin tamamı [33] veya öz bölümüdür [34]. Buna ek olarak, başlık [35], yazar bilgileri [36], dergi [37] ve anahtar ifadeler [38] gibi çeşitli meta veri bilgileri de kullanılabilir.

Küresel Atıf Önerme görevi için [74]’da özellik regresyonu ve konu öğrenme yöntemleri ile farklı atıf özellikleri ve atıf bağlamları kullanılmıştır. [75], ağ seyrekliği gibi sorunların etkisini en aza indirirken atıflar, dergi, konu, yazarlar ve anahtar kelimeler gibi bilgileri kullanarak kişiselleştirilmiş bir atıf önerme modeli önermiştir. [76] tarafından ortaya atılan bir Küresel Atıf Önerme yaklaşımı, makale temsillerini ve yazar temsillerini birleştirmekte ve otomatik kodlayıcı kullanmaktadır.

Bu çalışmalara ek olarak, atıfların daha gelişmiş temsillerini elde etmek için hem yerel bağlamdan hem de küresel özelliklerden yararlanan çalışmalar da mevcuttur [39,77].

2.2.3 Veri kümeleri

Özellikle Atıf Önerme tekniklerini geliştirmek ve test etmek için çok sayıda akademik makale veri kümesi toplanmıştır. Veri kümelerinin hiçbiri akademik hukuk makalelerini içermese de, CORE [114], S2ORC [40], CiteSeer [115] ve MDCR [41] birden fazla disiplinden makaleler içermekte iken; PubMed [116] ve RELISH [42] gibi diğer bazı veri kümeleri tıp gibi yoğun olarak araştırılan alanlara odaklanmakta ve atıf bağlantıları, meta veriler gibi bilgiler sağlamaktadır. Ancak, DBLP [43], ACL-AAN [44], ACL-ARC [45], arXiv CS [46], Scholarly Dataset [47], ve unarXiv [48] gibi veri kümeleri ise ağırlıklı olarak bilgisayar bilimleri veya ilgili alanlardaki makaleleri içermektedir.

Türkçe için [49] tarafından yürütülen bir çalışma sonucunda 7.000 makalelik bir derlem oluşturulmuştur. Derlem, hukuk da dahil olmak üzere 35 farklı konu alanından 200 makale içermektedir. Ancak, herhangi bir atıf bilgisi sağlanmamıştır.

2.3 Hukuki Atıf Önerme

Atıf Önerme yöntemleri, hukuk alanına, Hukuki Atıf Önerme görevi adı altında mahkeme kararları, tüzükler ve atıfta bulunulacak yasalar gibi akademik olmayan hukuki metinleri tespit etmek amacıyla uyarlanmaktadır.

Hukuki Atıf Önerme alanındaki ilk çalışmalar incelendiğinde, başlangıçta içerik temelli filtreleme yöntemlerinin hukuki öneri sistemleri oluşturmak için kullanıldığı görülmektedir [50]. Buna ek olarak, Hukuki Atıf Önerme yöntemleri, alıntı ağları ve tıklama akışı analizi kullanarak, hukuki belgeleri önerme görevi için modeller oluşturmak amacıyla kullanılır [51]. [52]'de, hukuki belgeler bir çizgedeki düğümler olarak temsil edilir ve atıf bağlantıları kenarlar olarak gösterilir. Burada atıflar kullanılarak hukuki belge önerme görevi için çoklu çizge tabanlı yöntemlerin uygulanması amaçlanmaktadır. [11]'deki çalışma, alıntı yapılacak uygun hukuki belgeleri bulmayı amaçlayan çeşitli metrikler üzerinde farklı içerik temelli filtreleme ve işbirlikçi filtreleme yöntemlerinin uygulanması yoluyla hukuki atıf önerme görevinde sinir ağı modellerinin geleneksel modellere göre üstünlüğünü göstermektedir. Ayrıca, [53]'de

uygun atıf bağlantılarını ve ilgili mahkeme kararlarını bulmak için hukuki bilgi çizgeleri kullanılmaktadır.

Bu çalışmalar genel olarak değerlendirildiğinde görülmektedir ki, hukuki atıf önerme çalışmaları temelde akademik olmayan hukuki metinler üzerinden ilerlemekte ve akademik hukuki makalelere gereken önem verilmemektedir.





3. HUKUKİ ATIF ÖNERME

Bu bölümde, hukuk alanında Atıf Önerme görevi için derlenen veri kümesi, uygulanan yöntemler, deney düzeneği ve elde edilen sonuçlardan bahsedilmektedir.

Hukuki Atıf Önerme görevi için, sadece akademik hukuki makalelerden oluşan özel bir veri kümesi toplanmıştır. Ardından, Atıf Önerme ve Atıf Önerme ile ilgili görevlerde iyi performans gösteren veya hukuk alanında eğitilmiş toplamda yedi farklı model dört farklı deney düzeneğinde test edilmiştir. Deneyleri tekrarlamak için gereken veri kümesi, kod ve ince ayar sonucu elde edilen modeller GitHub'ta¹ paylaşılmıştır. Aşağıdaki bölümlerde Atıf Önerme görevinin hukuki makalelerdeki uygulaması açıklanmaktadır.

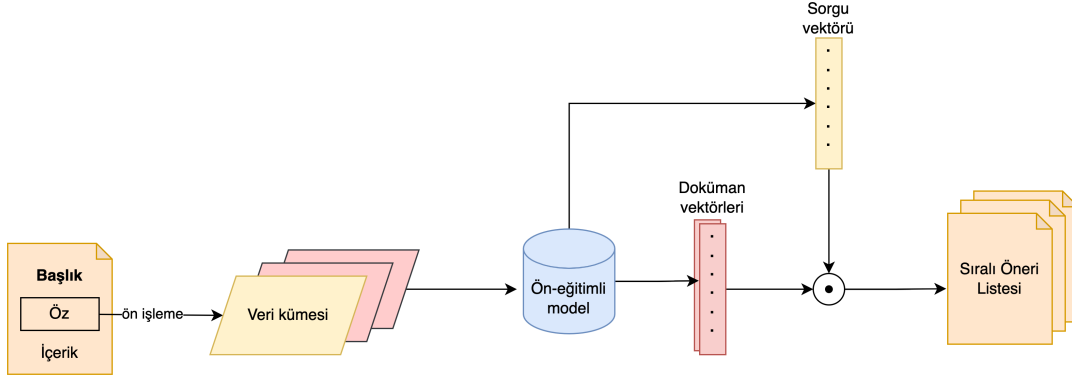
3.1 Yöntem

Gerçekleştirilen deneylerde, aşağıda açıklanan yedi model için dört farklı deney düzeneği bulunmaktadır. Deneylerde:

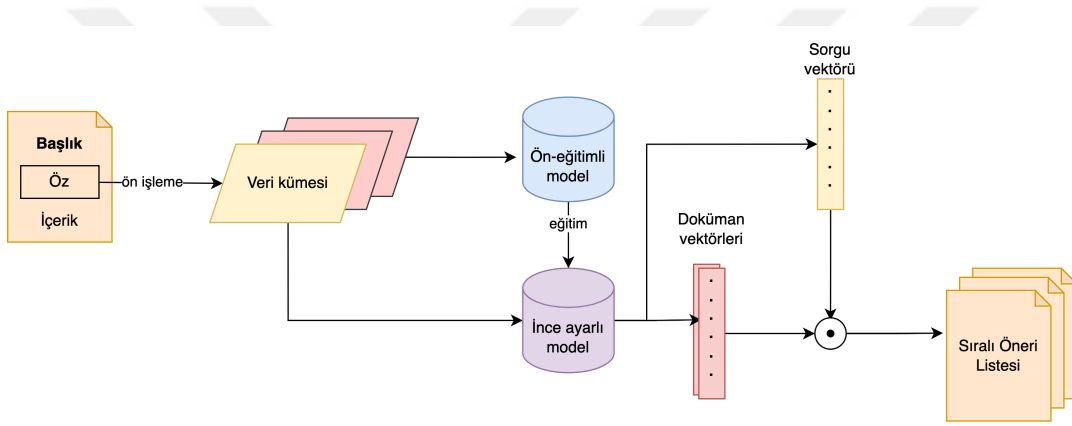
- (i) önceden eğitilmiş modeller doğrudan kullanılır (Şekil 3.1),
- (ii) önceden eğitilmiş modeller, Atıf Önerme görevi için toplanan veri seti ile ince ayardan geçirilir (Şekil 3.2),
- (iii) ilk olarak BM25 ile ilgili makaleler çekilir, daha sonra çekilen makaleleri yeniden sıralamak için önceden eğitilmiş modeller kullanılır (Şekil 3.3),
- (iv) ilk olarak BM25 ile ilgili makaleler alınır, daha sonra alınan makaleleri yeniden sıralamak için ince ayarlı modeller kullanılır (Şekil 3.4).

Bu dört farklı deney düzeneğinin seçilmesinin nedenleri şu şekilde sıralanabilir: bilimsel makalelerle önceden eğitilmiş Atıf Önerme modellerinin performansını hukuk

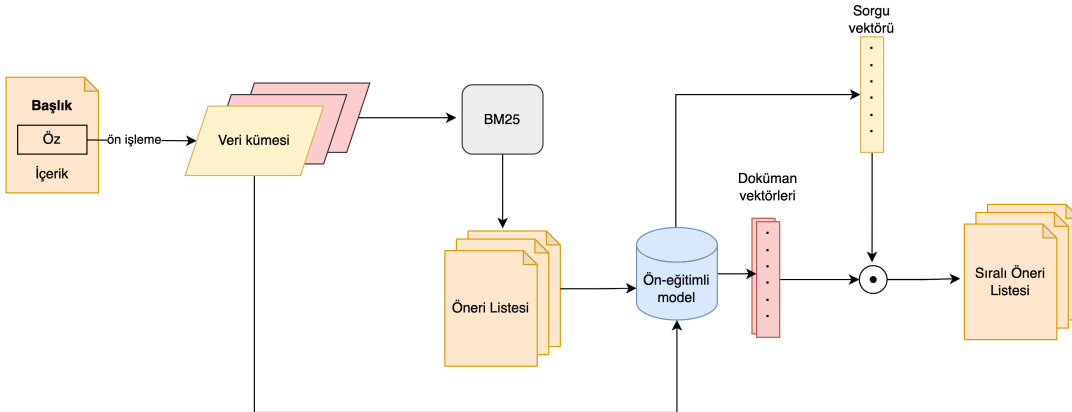
¹<https://github.com/dgknrsln/LegalCitationRecommendation>



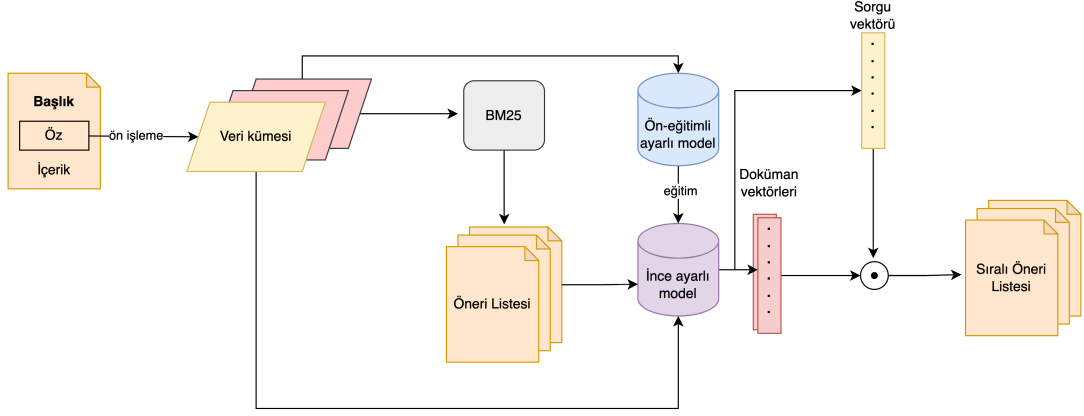
Şekil 3.1 : Birinci deney düzeneğinde önceden eğitilmiş modellerden elde edilen doküman temsillerinin, yine aynı şekilde elde edilen sorgu temsili ile olan kosinüs benzerliklerine göre sıralanması.



Şekil 3.2 : İkinci deney düzeneğinde ince ayar yapılmış modellerden elde edilen doküman temsillerinin, yine aynı şekilde elde edilen sorgu temsili ile olan kosinüs benzerliklerine göre sıralanması.



Şekil 3.3 : Üçüncü deney düzeneğinde BM25 tarafından getirilen 10 makalenin önceden eğitilmiş modellerle yeniden sıralanması.



Şekil 3.4 : Dördüncü deney düzeneğinde BM25 tarafından getirilen 10 makalenin ince ayar yapılmış modellerle yeniden sıralanması.

alanında önceden eğitilmiş dil modelleriyle karşılaştırmak, kullanılan modellere ince ayar yaparak alan bilgisi eklemenin etkisini gözlemlemek ve BM25 ile ön-getirmenin hukuki Atıf Önerme görevi için dil modellerinin performansını artırıp artırmadığını araştırmaktır.

Buna ek olarak, iki aşamalı yaklaşım, BM25 kullanarak bir makale örneklemini hızlı bir şekilde seçerek dil modellerinin hantallığını hafifletmeyi amaçlamaktadır. Bilgi getirmiş çalışmalarında bu iki aşamalı yaklaşım, sistem etkinliğini artırmak için sıklıkla kullanılmaktadır. Yaklaşım, ilk aşamada BM25 gibi daha hızlı olan ancak daha az doğruluk gösteren modellerin, ilgili belgelerin bir ön örneklemini almak için kullanılmasını içerir. Daha sonra, BERT gibi daha yavaş ancak yüksek doğruluk gösteren modellerle, ön-getirimi yapılan aday makaleler yeniden sıralanır.

Aşağıda, kullanılan modeller tarihsel gelişim sırasına göre ve bir önceki modelde eksik olan ve mevcut modelde iyileştirilen hususlara atıfta bulunularak sunulmaktadır.

BM25 Best Matching-25 [84] bir sorguya dayalı olarak ilgili belgeleri bulmak için kullanılan güçlü bir bilgi getirme yöntemidir. Deneyler sırasında, diğer versiyonlara nazaran daha başarılı olduğu tespit edilen BM25+ versiyonu rank-bm25 Python paketi aracılığıyla kullanılmıştır (BM25 varyantlarının derlenen veri kümesi üzerinde detaylı değerlendirmesi için bkz., Bölüm 4.1). BM25+, kelimeleri q_i olan bir Q sorgusu ve bir D belgesi için aşağıdaki gibi formüle edilebilir:

$$skor_{benzerlik}(D, Q) = \sum_n^{i=1} ds(q_i)^{-1} \cdot \left(\frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{|D|_{ort}})} + \delta \right) \quad (3.1)$$

Burada, $ds(q_i)^{-1}$ bir sorgu teriminin (q_i) doküman sıklığının tersi ve $f(q_i, D)$ ise q_i 'nin belgedeki (D) sıklığıdır. Belge uzunluğu ($|D|$) ve ortalama belge uzunluğu ($|D|_{ort}$) da BM25'teki hesaplamaya katkıda bulunur.

Bu noktada BM25+'ın BM25'dan farkı eklenen δ parametresidir. Geleneksel BM25 formülünde, Q sorgusunda yer alan q_i terimini içeren uzun dokümanlar, bu terimi içermeyen kısa dokümanlarla benzer skorlar elde edebilmektedir. Bu haksız değerlendirmenin önüne geçebilmek için δ parametresi kullanılmaktadır.

Law2Vec Law2Vec adı verilen iki ayrı Word2Vec [85] modeli hukuk alanı için [86] tarafından eğitilmiştir. Biri 100 boyutlu, diğeri ise 200 boyutludur. Bu tezdeki deneyler için ikincisi kullanılmıştır. Law2Vec eğitimi sırasında, farklı ülkelerden mevzuat, mahkeme kararları ve tüzükler gibi farklı türde İngilizce hukuki belgeler kullanılmıştır. Eğitim veri seti yaklaşık 120 bin hukuki belgeden oluşmaktadır. [87], yasal benzerlik görevleri için Law2Vec'in kullanılan diğer yöntemlerle benzer performans gösterdiğini belirtmektedir.

Bir dil modeli olarak Word2Vec, ilgili kelimelerin birbirine daha yakın olması amacıyla kelimeleri vektörel uzayda konumlandırır. Bağlamsal bilgileri yakalamak için her kelime çevresiyle birlikte işlenir. Bu tezde, Law2Vec, bilimsel hukuk makaleleri için temsil elde etmek amacıyla kullanılmıştır. Veri kümesi alana özgü olduğundan, Law2Vec, kullanılan diğer dil modellerinden daha iyi temsil sağlayabilir.

SciBERT SciBERT [89], bilimsel makalelerle eğitilen bir BERT (Bidirectional Encoder Representations from Transformers) [88] türevidir. 1.14 milyon bilgisayar bilimi ve biyomedikal makalesi ile eğitilmiştir. Bu nedenle SciBERT, belge sınıflandırma, varlık ismi tanıma, belge özetleme ve Atıf Önerme gibi bilimsel makalelere odaklanan alt görevlere uygulanabilir durumdadır.

SciBERT ile yapılan deneyler, ince ayarın, model mimarisinin göreve özgü değişiminden daha etkili olduğunu göstermektedir. Ayrıca, diğer birçok veri kümesi

arasında SciBERT, bilimsel makalelerden oluşan bir derlem olan SciVocab ile önceden eğitildiğinde en iyi performansı göstermektedir.

SPECTER BERT ve türevlerinin cümle temsilinin ötesine geçemediği iddiasından yola çıkarak, belge temsili elde etmek için geliştirilen dönüştürücü tabanlı bir dil modeli olan SPECTER (Scientific Paper Embeddings using Citation informed TransformerS) [90], [91] birçok alt görevde başarılı sonuçlar elde etmektedir. Yazarlar, SPECTER'in bu alt görevler için ince ayar gerektirmediğini iddia etmektedir. Ayrıca aynı yayında, bilimsel makalelerle ilgili görevler için SciDocs adlı bir kıyaslama ölçütü de sunmuşlardır. SPECTER'in SciDocs'un yedi farklı görevinde SciBERT de dahil olmak üzere çeşitli dil modellerinden daha iyi performans gösterdiğini belirtmişlerdir.

LegalBERT BERT dil modeli; modelin daha fazla ön eğitimi, sıfırdan eğitimi ve modele ince ayar yapılması olarak üç farklı şekilde hukuk alanına uyarlanmıştır [92]. 12 GB'lık eğitim verileri AB, Birleşik Krallık ve ABD'den sözleşmeler, mahkeme davaları ve çeşitli mevzuat belgelerini içermektedir. Sıfırdan eğitilmiş bir BERT modeli olan LEGAL-BERT-SC adlı bir LegalBERT türevi bu tez kapsamında kullanılmıştır.

SciNCL SPECTER'daki örnekleme stratejisi [93]'de eleştirilmiş ve önceki çalışmaların "bilimsel makalelerin, aralarında doğrudan atıf olmasa bile önemli benzerlikler gösterebileceği" varsayımını gözden kaçırdığı belirtilmiştir. SciNCL (Neighborhood Contrastive Learning for Scientific document representations with citation graph embeddings) "kontrollü en yakın komşu örnekleme" adı verilen bir yöntem kullanmaktadır. SciNCL başka bir BERT türevidir ve SciBERT'in ağırlıkları ile eğitime başlanılır. Yazarlar, SciNCL'nin Atıf Önerme dahil olmak üzere çeşitli alt görevlerde SPECTER'dan daha iyi performans gösterdiğini belirtmektedir.

SGPT GPT'nin (Generative Pre-trained Transformer) [95] SGPT [94] adı verilen bir türevi, anlamsal arama ve cümle temsili elde etme görevi için özel olarak eğitilmiştir.

GPT, dil modelleme amacıyla eğitilmiştir ve bu nedenle bir sonraki belirteç tahmin görevinde çok iyi performans gösterir. SGPT, anlamsal arama görevi için özel olarak tasarlanmış temsiller oluşturmak için GPT mimarisindeki kod çözücülerini kullanarak bir adım daha ileri gider. Bu nedenle, cümle temsillerini elde etmek için kullanılan diğer son teknoloji yöntemlerden daha iyi performans gösterir.

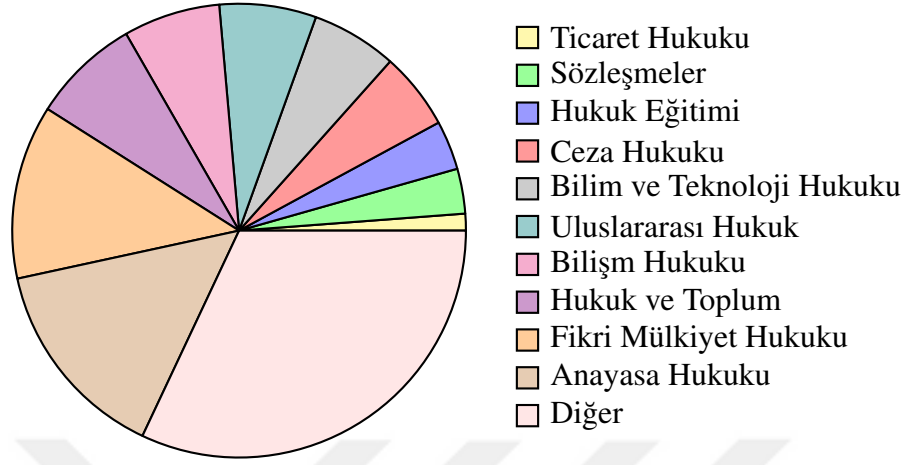
3.2 Veri Kümesi

İngilizce hukuki Atıf Önerme görevi için, yukarıda listelenen modelleri eğitmek üzere LawArXiv [117] adlı hukuki bilimsel makaleler veritabanındaki makaleler indirilmiştir. LawArXiv artık yeni makale kabul etmeyen açık kaynaklı bir veri havuzudur. Birçok hukuki konuyu kapsayan 1366 bilimsel hukuki makaleden oluşmaktadır. Makalelerin konularına göre dağılımı Şekil 3.5'de gösterilmektedir.

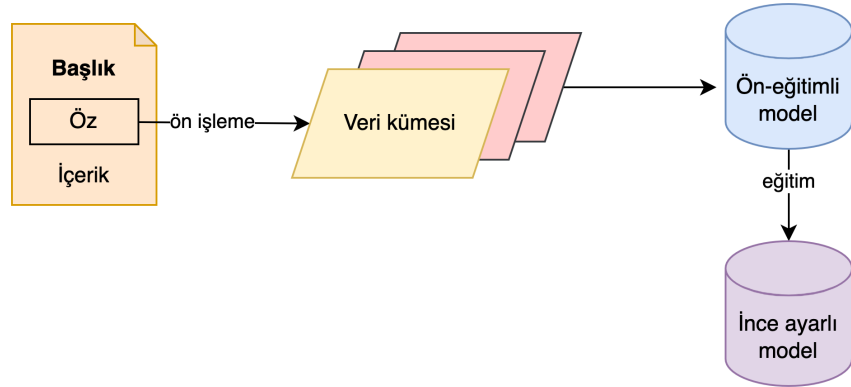
Makaleler elde edildikten sonra, atıfta bulunan makaleleri elde etmek için Google Scholar'da [118] makalelerin başlıkları ile sorgu yapılır. Bu amaçla SERP [119] adlı bir araç kullanılmıştır. Sonuç olarak, 10 binin üzerinde atıfta bulunan makale elde edilmiştir. Elde edilen makale PDF'lerinden, pdfplumber [120] adlı bir Python paketi kullanılarak bu makalelerin bağlamı çıkarılır. Ardından, başarılı bir şekilde çıkarılan İngilizce makaleler seçilir. Ön işleme adımında, ASCII olmayan karakterler kaldırılır ve makaledeki büyük harfler küçük harfe çevrilir. Eğitim aşamasında kullanılmak üzere, "abstract" (öz) anahtar kelimesi aranarak makalelerin öz kısımları çıkarılır. Deney aşamasında toplamda 719 LawArXiv makalesinden oluşan ve 8.887 atıfta bulunan makaleden elde edilen 10.111 atıf bağlantısı içeren bir veri kümesi kullanılmıştır.

3.3 Deney Düzenegi

Bahsedilen tüm yöntemler için makalelerin öz kısımları, benzer içerik temelli küresel Atıf Önerme çalışmalarıyla tutarlı olarak ince ayar, temsil elde etme ve test aşamalarında girdi olarak kullanılmıştır [41] (Şekil 3.6). Veri kümesi ayrı eğitim ve test kümelerine bölünmüş, verilerin %70'i eğitim için kullanılmış ve kalan %30'u test için ayrılmıştır. 3 devir süren ince ayar aşamasında, üçlü kayıp fonksiyonu



Şekil 3.5 : LawArXiv’da yer alan makalelerin bazı temel hukuk konularına göre dağılımı.



Şekil 3.6 : Eğitimlerde makalelerin öz kısımlarının ön işlemden geçirilerek kullanılması ve veri kümesinin üçte birlik kısmının test için ayrılması.

[96] kullanılmıştır. Üçlü kayıp, bir referans girdiyi (çapa olarak adlandırılır) çapaya benzeyen pozitif bir girdiyle ve çapayla eşleşmeyen negatif bir girdiyle karşılaştırır. Bir sorgu için (çapa olarak seçilir), atıfta bulunulan belgeler pozitif eşleşme olarak belirlenir ve atıfta bulunulmayan belgelerden rastgele biri negatif örnek olarak belirlenir.

İnce ayar ve temsil elde etme adımlarından sonra, üretilen belge temsil vektörleri, vektör uzayında iki vektör arasındaki açının kosinüsü kullanılarak hesaplanan, verilen sorgu ve diğer belgeler arasındaki benzerliğe göre sıralanır. Tüm eğitim ve test prosedürleri için Sentence-Transformers [98] uygulama çerçevesi kullanılmıştır.

3.4 Değerlendirme Ölçütleri

Söz konusu deneylerin sonuçları, bilgi getirimi çalışmalarında yaygın olarak kullanılan üç farklı metrik kullanılarak sunulmuştur: *Mean Average Precision (MAP)* (Ortalama Kesinliklerin Ortalaması), *Recall* (Duyarlılık), ve *Mean Reciprocal Rank (MRR)* (Sıralamaların Terslerinin Ortalaması). Bu metrikler, bir makale için ortalama 14 atıf bağlantısına sahip olduğumuzdan, getirilen ilk 10 belge (n=10) için raporlanmıştır.

MAP Ortalama kesinliklerin ortalaması, aşağıdaki formülde tanımlandığı gibi, bir dizi sorgu (N) genelinde ortalama kesinlik (AP) değerlerinin ortalaması alınarak hesaplanır:

$$MAP = \frac{1}{N} \sum_{i=1}^N AP(Q_i) \quad (3.2)$$

Bilgisayarlı görüdeki nesne algılama modellerinin ve bilgi edinimindeki modellerin başarısını raporlamak için yaygın olarak kullanılmaktadır.

Recall Duyarlılık, doğru pozitifler olarak da adlandırılan başarılı bir şekilde geri getirilen benzer belgelerin sayısının, doğru pozitifler ve yanlış negatiflerin toplamı olan toplam benzer belge sayısına bölünmesiyle hesaplanır. Hesaplama aşağıdaki formülle ifade edilebilir:

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

Duyarlılık puanı, geri getirilmeyen benzer belgelere de bağlı olduğu için önemlidir.

MRR Sıralamaların terslerinin ortalaması, bir dizi sorgu için sıralamaların tersinin ortalaması alınarak hesaplanır (N). Her sorgu için sıralamanın tersi, alınan ilk ilgili belgenin sırasının ($rank_Q$) tersini temsil eder. MRR hesaplama formülü aşağıdaki gibi ifade edilebilir:

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank_{Q_i}} \quad (3.4)$$

İlgili belgeleri mümkün olan en yüksek sırada almak önemli olduğundan, MRR, Atıf Önerme görevini değerlendirmek için yararlı bir metriktir.

3.5 Sonuçlar

Bu bölümde, yukarıda bahsedilen 4 farklı deney düzeneğinden elde edilen sonuçlar sunulmakta ve yorumlanmaktadır.

Çizelge 3.1 : BM25 ve ön eğitilmiş modellerin ilk 10 makaleyi getirme performansı.

	MAP@10 ↓	Recall@10 ↓	MRR@10 ↓
BM25	0.26	0.45	0.31
SciNCL	0.18	0.33	0.23
SGPT	0.17	0.30	0.22
SPECTER	0.14	0.26	0.19
Law2Vec	0.11	0.21	0.17
LegalBERT	0.08	0.16	0.15
SciBERT	0.08	0.16	0.14

BM25 ve önceden eğitilmiş modeller: Önceden eğitilmiş çeşitli modeller ve derlenen veri kümesi eğitilmiş BM25 modelinin karşılaştırması Çizelge 3.1’de sunulmuştur. Çok sayıda bilimsel makale ile eğitilmiş olmasına rağmen SciBERT diğerlerine kıyasla en düşük performansı gösteren model olmuştur. SciBERT’ten sonra

en başarısız iki dil modeli olarak, hukuki bir derleme eğitilmiş modeller olan Law2Vec ve LegalBERT gelmektedir. Bu modellerin diğer modellere kıyasla başarısız olmasının temel nedeni, Atıf Önerme görevi için eğitilmemiş olmalarıdır. Bunu destekleyen bir argüman da, Çizelge 3.2’de görülebileceği gibi, LegalBERT’in ince ayar sonrası performansının MAP@10 için 0.08’den 0.24’e yükselmesidir. SciNCL, makalesindeki iddialara paralel olarak en iyi performans gösteren ön-eğitilmiş model olarak öne çıkmaktadır. SGPT’nin performansına bakıldığında, bilimsel makalelerle veya Atıf Önerme görevi için eğitilmemiş olmasına rağmen, SPECTER ve SciBERT’ten daha iyi performans göstererek nispeten başarılı olduğu görülmektedir. Bunun nedeni, SGPT’nin eğitildiği anlamsal arama görevinin temelde Atıf Önerme görevine benzer olmasıdır. Öte yandan BM25, önceden eğitilmiş tüm modellerden daha iyi performans göstererek en başarılı model olarak öne çıkmaktadır. Bu bulgu, literatürdeki bilimsel alan temelli değerlendirmeye odaklanan Atıf Önerme çalışmalarının [41] sonuçlarıyla da tutarlıdır.

Çizelge 3.2 : İnce ayar yapılmış modellerin ilk 10 makaleyi getirme performansı.

	MAP@10 ↓	Recall@10 ↓	MRR@10 ↓
SciNCL	0.26	0.49	0.30
SciBERT	0.26	0.49	0.29
SPECTER	0.25	0.47	0.28
SGPT	0.25	0.46	0.29
LegalBERT	0.24	0.47	0.28

İnce ayarlı modeller: Önceden eğitilmiş modellere ince ayar yapıldığında Çizelge 3.2’de verilen sonuçlar elde edilmiştir. Sonuçlar genel olarak incelendiğinde model performanslarının birbirine yakın olduğu görülmektedir. Modeller BM25’in performansına ulaşabilmiş olsalar da onu geçememişlerdir. Bu da bu modellerin veri sağlandığında test edilen alana uyum sağlayabildiğini göstermektedir. Süreç sonucunda, ince ayar yapılmış LegalBERT modelinin performansı önceden eğitilmiş modele kıyasla önemli ölçüde artmıştır. Bu noktada modelin göreve aşinalığının alan bilgisi ile birleşmesinin model performansını artırdığı düşünülmektedir. En başarılı modellere baktığımızda SciNCL ve SciBERT öne çıkmaktadır. SciNCL istikrarlı bir

gelişim gösterirken, SciBERT’in performansındaki sıçrama dikkat çekici ve araştırma gerektiren bir noktadır.

Çizelge 3.3 : BM25 tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ön eğitilmiş modellerin performansı.

	MAP@10 ↓	Recall@10 ↓	MRR@10 ↓
SciNCL	0.25	0.45	0.30
SGPT	0.25	0.45	0.30
SPECTER	0.24	0.45	0.29
LegalBERT	0.19	0.45	0.24
Law2Vec	0.19	0.45	0.24
SciBERT	0.19	0.45	0.24

BM25 ile ön-getirme ve önceden eğitilmiş modellerle yeniden sıralama:

Çizelge 3.3 BM25’in geri getirme kapasitesi ile kullanılan modellerin önceden eğitilmiş versiyonlarının sıralama yeteneklerini birleştiren deneylerin sonuçlarını göstermektedir. Sonuçlar ilk deneyle (Çizelge 3.1) uyumlu olsa da, önceden eğitilmiş modellerin BM25’in performansını artıramadığı görülmüştür. Yine de SciNCL ve SGPT en başarılı modeller olarak öne çıkmaktadırlar.

Çizelge 3.4 : BM25 tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ince ayar yapılmış modellerin performansı.

	MAP@10 ↓	Recall@10 ↓	MRR@10 ↓
SciNCL	0.30	0.45	0.34
SGPT	0.29	0.45	0.34
SciBERT	0.29	0.45	0.34
SPECTER	0.28	0.45	0.33
LegalBERT	0.28	0.45	0.33

BM25 ile ön-getirme ve ince ayarlı modellerle yeniden sıralama: Son olarak, BM25 ile getirilen makalelerin ince ayarlı modellerle yeniden sıralanmasıyla elde edilen sonuçların değerlendirilmesi Çizelge 3.4’te gösterilmiştir. Burada, tüm ince ayarlı modellerin BM25’in performansını en az 0.02 puan artırdığı görülmektedir. Diğer deneylerle uyumlu olarak SciNCL, BM25’in başarımını 0.04 puan artırarak en başarılı model olarak öne çıkmaktadır (0.30 MAP@10).



4. TARTIŞMA

Bu bölümde öncelikle BM25 algoritmasının türevleri incelenmiş, bu türevlerden literatüre paralel olarak seçilen üç tanesinin derlenen veri kümesindeki başarımı ölçülmüştür. Ek olarak önceki bölümde bahsedilen deneyler sonucu elde edilen veriler ayrıntılı bir şekilde analiz edilmiş ve bulgular giriş bölümünde öne sürülen hipotezler doğrultusunda değerlendirilmiştir.

4.1 BM25

Bu bölümde BM25 algoritması ayrıntılı olarak incelenmiş, var olan BM25 türevleri öne çıkan özellikleriyle listelenmiştir. Ardından belirlenen üç BM25 türevinin derlenen veri kümesindeki başarımı raporlanarak, önceki bölümlerde yer alan deneylerde kullanılan BM25 türevinin neden seçildiği açıklanmıştır.

Bilgi getirme görevlerinde sıkça kullanılan bir sıralama algoritması olan BM25, formülü gereği dokümanlarda daha az geçen terimlere daha fazla ağırlık verdiği için akademik hukuk makalelerinin birbiriyle olan ilişkilerini ortaya çıkarmada öne çıkıyor oluşu şaşırtıcı değildir (bkz., Bölüm 3.5). Özellikle kullanıldığı alanın hukuk olması ve daha önce bahsedildiği üzere hukuki metinlerin yabancı dillerden ve eski dilde sözcükler içermesinden dolayı, bu kelimelerin BM25 tarafından kolayca tespit edilip gereken ağırlığın verildiği ve BM25'in dokümanları bu doğrultuda değerlendirdiği söylenebilir.

Girdi olarak kullanılan makale özlerinin ortalama olarak aynı uzunlukta olduğu düşünüldüğünde, formüldeki doküman uzunluğunun sonuca fazla etki etmediği de açıktır. Fakat dokümanların aynı uzunlukta olması, aynı alandaki bu dokümanların içerdiği alana özgü kelimelerin (yani daha fazla değer verilmesi gereken kelimelerin) daha doğru bir şekilde tespit edilebilmesini sağlamıştır. Bu durum BM25'in kelimelere verdiği önemin doküman uzunluğu ile ters orantılı olması yani daha

uzun dokümanlardaki sık kullanılan kelimelere daha az önem atfetmesinden ileri gelmektedir.

BM25'in yukarıda listelenen avantajlarının yanında performansına da dikkat çekmek gerekir. Derlenen veri kümesi boyut olarak her ne kadar küçük kalsa da, BM25 modeli eğitmek ve bir sorguda bulunmak birkaç saniye bile sürmezken, dil modellerinden kelime temsili elde etmek birkaç saati bulmaktadır.

Ek olarak, BM25'in sahip olduğu birtakım dezavantajlar da göz önünde bulundurulmalıdır. Bunlardan en önemlisi BM25'in bağlamı dikkate almıyor oluşudur. Terimlerin anlamsal ilişkilerinden yoksun bir değerlendirmeye sahip olması BM25'i karmaşık problemler için ikinci plana itse de deneylerde görüldüğü gibi bu durum, atıf öneri görevinde başarılı olmasını engellememektedir.

Özetlenecek olursa, BM25'in gösterdiği başarıyı aşağıdaki faktörlerle ilişkilendirilebilir:

1. Veri kümesinin kendine has dil özellikleri barındıran hukuk alanında olması.
2. Veri kümesinin benzer uzunlukta dokümanlardan oluşması.
3. Doküman havuzunun fazla geniş olmaması.

4.1.1 BM25 türevleri

BM25 algoritması öne sürüldükten sonra, zaman içinde, farklı görevlere yönelik olarak veya çeşitli problemleri çözmek amacıyla geliştirilmiş BM25 türevleri de ortaya çıkmıştır.

Bunlardan ATIRE BM25 [111] skor olarak negatif sayıların elde edilmesinin önüne geçmek için geliştirilmiştir. Geleneksel BM25 algoritması, doküman sayısının yarısından fazla doküman sıklığına sahip terimler için negatif sonuç üreten bir ters terim sıklığı formülü kullanırken, ATIRE versiyonu terimin doküman sıklığı toplam doküman sayısına (N) yaklaştıkça sıfıra yaklaşan sonuçlar üreten bir ters terim sıklığı formülü kullanmaktadır. BM25-adpt [112] versiyonunda ise, genel bir k_1 değeri kullanmaktansa, k_1 değişkeni terim bazlı bir değer olarak ele alınmıştır. Bunun sonucunda elde edilen formül, modeli, yeniden eğitime ihtiyaç duymadan başka

veri kümelerine de uyarlanabilir hale getirmiştir. Buna paralel olarak, BM25-adpt algoritmasını geliştirmek adına yine aynı yazarlar tarafından BM25T, BM25C ve BM25Q gibi BM25 türevleri de geliştirilmiştir [113].

Bu tez kapsamında, hukuki doğal dil işleme çalışmalarında, atıf önerme görevinde ve benzer görevlerde sıkça kullanılan üç BM25 türevi seçilmiştir. OkapiBM25, BM25L ve BM25+ olarak belirlenen bu türevler veri kümesi üzerinde teste tabii tutulmuş ve en başarılı BM25 türevi 1 numaralı deneyden doğrudan, iki aşamalı test içinse 3 ve 4 numaralı deneylerde 10 adet makalenin getiriminde kullanılmıştır. Aşağıda seçili BM25 türevleri detaylı bir biçimde açıklanmış ve bu algoritmalar kullanılarak elde edilen sonuçlar paylaşılmıştır.

Okapi BM25 Orijinal BM25 olarak da bilinen Okapi BM25'in formülü aşağıdaki gibidir:

$$skor_{benzerlik}(D, Q) = \sum_n^{i=1} \log\left(\frac{N - ds(q_i) + 0.5}{ds(q_i) + 0.5}\right) \cdot \left(\frac{f(q_i, D)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{|D|_{ort}})}\right) \quad (4.1)$$

Burada N derlemdeki doküman sayısı, $ds(q_i)$ ise bir sorgu teriminin (q_i) belge sıklığı ve $f(q_i, D)$ ise q_i 'nin belgedeki (D) sıklığıdır. Belge uzunluğu ($|D|$) ve ortalama belge uzunluğu ($|D|_{ort}$) da Okapi BM25'ta hesaplama etki eder.

BM25L Okapi BM25'in kısa dokümanlara uzun dokümanlara kıyasla daha yüksek skorlar vermesinin tespit edilmesi ile beraber bu durumun etkilerini azaltmak için önerilen BM25 türevidir. Öncelikle negatif skorlamayı önlemek için formül aşağıdaki hale getirilmiştir.

$$skor_{benzerlik}(D, Q) = \sum_n^{i=1} \log\left(\frac{N + 1}{ds(q_i) + 0.5}\right) \cdot \left(\frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{|D|_{ort}})}\right) \quad (4.2)$$

Ardından aşağıdaki c_{q_i} değeri için,

$$c_{q_i} = \frac{f(q_i, D)}{1 - b + b \cdot \frac{|D|}{|D|_{ort}}} \quad (4.3)$$

Formül aşağıdaki şekilde düzenlenmiş ve uzun dokümanların skorunu artıran pozitif bir δ değeri eşitliğe eklenmiştir.

$$skor_{benzerlik}(D, Q) = \sum_n^{i=1} \log\left(\frac{N+1}{ds(q_i)+0.5}\right) \cdot \left(\frac{(c_{q_i} + \delta) \cdot (k_1 + 1)}{(c_{q_i} + \delta) + k_1}\right) \quad (4.4)$$

BM25+ BM25+ algoritması Bölüm 3.1’de detaylı bir biçimde anlatılmıştır.

4.1.2 BM25 türevlerinin karşılaştırılması

Aşağıda, değerlendirmeye alınan 3 farklı BM25 türevinin daha önce bahsedilen 3 deneyde başarımları karşılaştırmalı olarak verilmiştir.

Çizelge 4.1 : BM25+, Okapi BM25 ve BM25L modellerinin derlenen veri kümesindeki başarımları.

	MAP@10 ↓	Recall@10 ↓	MRR@10 ↓
BM25+	0.267	0.452	0.317
Okapi BM25	0.260	0.431	0.312
BM25L	0.152	0.333	0.211

BM25’in doğrudan kullanımı BM25 türevlerinin derlenen veri kümesi ile eğitilip test edilmesi sonucu elde edilen sonuçlar Çizelge 4.1’de verilmiştir. Tabloya göre BM25+ ile Okapi BM25 modelleri öne çıkarken BM25+’ın daha başarılı olduğu görülmektedir. BM25L ise diğer modellere kıyasla bariz bir biçimde geride kalmıştır.

BM25 ile ön-getirme ve önceden eğitilmiş modellerle yeniden sıralama: Okapi BM25, BM25L ve BM25+ olmak üzere belirlenen üç BM25 türevinin getirdiği 10 makalenin önceden eğitilmiş modeller yardımıyla yeniden sıralanması sonucu elde edilen başarımlar Çizelge 4.2’de gösterilmektedir. Çizelge 4.1 ile paralel olarak, en başarılı sonuçlar BM25+’ın getirdiği makaleler üzerinden elde edilmiştir. Her bir BM25 türevi için de en başarılı ilk üç model SciNCL, SGPt ve SPECTER olup değişmezken; LegalBERT, Law2Vec ve SciBERT’in sıralanmasında farklılıklar gözlemlenmiştir. Ek olarak, BM25+ ve Okapi BM25 ile yapılan deneylerde, kullanılan modellerin bu BM25 türevlerinin başarımlarını düşürdüğü, öte yandan, BM25L türevi

Çizelge 4.2 : BM25 türevleri tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ön eğitilmiş modellerin performansı.

	MAP@10 ↓	Recall@10 ↓	MRR@10 ↓
SciNCL _{BM25+}	0.25	0.45	0.30
SGPT _{BM25+}	0.25	0.45	0.30
SPECTER _{BM25+}	0.24	0.45	0.29
LegalBERT _{BM25+}	0.19	0.45	0.24
Law2Vec _{BM25+}	0.19	0.45	0.24
SciBERT _{BM25+}	0.19	0.45	0.24
SciNCL _{OkapiBM25}	0.24	0.43	0.30
SGPT _{OkapiBM25}	0.24	0.43	0.29
SPECTER _{OkapiBM25}	0.23	0.43	0.28
Law2Vec _{OkapiBM25}	0.19	0.43	0.24
LegalBERT _{OkapiBM25}	0.18	0.43	0.24
SciBERT _{OkapiBM25}	0.18	0.43	0.23
SciNCL _{BM25L}	0.21	0.33	0.27
SGPT _{BM25L}	0.20	0.33	0.26
SPECTER _{BM25L}	0.19	0.33	0.25
SciBERT _{BM25L}	0.17	0.33	0.23
LegalBERT _{BM25L}	0.17	0.33	0.23
Law2Vec _{BM25L}	0.16	0.33	0.22

içinse durumun tam tersi şekilde olduğu ve kullanılan modellerin BM25L'in başarımını artırdığı gözlemlenmiştir.

Çizelge 4.3 : BM25 türevleri tarafından getirilen ilk 10 makalenin yeniden sıralanmasında ince ayar yapılmış modellerin performansı.

	MAP@10 ↓	Recall@10 ↓	MRR@10 ↓
SciNCL _{BM25+}	0.30	0.45	0.34
SGPT _{BM25+}	0.29	0.45	0.34
SciBERT _{BM25+}	0.29	0.45	0.34
SPECTER _{BM25+}	0.28	0.45	0.33
LegalBERT _{BM25+}	0.28	0.45	0.33
SciNCL _{OkapiBM25}	0.28	0.43	0.34
SGPT _{OkapiBM25}	0.28	0.43	0.33
SciBERT _{OkapiBM25}	0.28	0.43	0.33
SPECTER _{OkapiBM25}	0.27	0.43	0.33
LegalBERT _{OkapiBM25}	0.27	0.43	0.32
SGPT _{BM25L}	0.23	0.33	0.29
SciNCL _{BM25L}	0.23	0.33	0.29
SciBERT _{BM25L}	0.23	0.33	0.29
LegalBERT _{BM25L}	0.23	0.33	0.29
SPECTER _{BM25L}	0.23	0.33	0.29

BM25 ile ön-getirme ve ince ayar yapılmış modellerle yeniden sıralama: Çizelge 4.3'de Okapi BM25, BM25L ve BM25+ olmak üzere belirlenen üç BM25 türevinin getirdiği 10 makalenin ince ayar yapılmış modeller yardımıyla yeniden sıralanması sonucu elde edilen başarımlar gösterilmektedir. Üç Bm25 türevi için de kullanılan modellerin BM25 türevlerinin başarımlarını artırdığı gözlemlenirken, Çizelge 4.1 ve Çizelge 4.2 ile paralel olarak en başarılı sonuçlar BM25+'ın getirdiği makaleler ile elde edilmiştir.

4.2 Hukuki Dokümanlarla Eğitilen Dil Modelleri

Deneylerde hukuki dokümanlarla eğitilmiş Law2Vec ve LegalBERT isimli iki dil modeli kullanılmıştır. Bu dil modelleri çok çeşitli hukuki dokümanlarla eğitilmesine rağmen, eğitim kümelerinde akademik hukuk makaleleri yer almamaktadır. Bu

doğrultuda bakıldığında hukuki atf öneri görevine alan bakımında yakın oldukları fakat görev bakımından pek fayda sağlayamayacakları tahmin edilebilir.

Nitekim, bu modellerin deneylerdeki performansları incelendiğinde, neredeyse en başarısız modeller oldukları göze çarpmaktadır (MAP@10 0.08 LegalBERT, MAP@10 0.11 Law2Vec). Fakat, ince ayar yapıldıktan sonra LegalBERT'teki performans yükselişi özellikle önceden eğitilmiş modellerin test edildiği birinci deneydeki en başarılı model olan SciNCL ile arasındaki farkı 0.10 puandan 0.02 puana düşürmüş olması gelecekteki çalışmalar için umut vaat edicidir. Öte yandan bu durum alan bilgisine sahip LegalBERT dil modelinin veri sağlandığında atf önerisi görevine de uyum sağlayabildiğini göstermektedir. Bu açıdan, sonraki çalışmalarda daha büyük veri kümeleriyle göreve özel ince ayar çalışmaları yapılabilir.

Özetlenecek olursa, hukuki dokümanlarla eğitilen Law2Vec ve LegalBERT modellerinin gösterdiği başarıyı aşağıdaki faktörlerle ilişkilendirilebilir:

1. Modellerin eğitim kümelerinden gelen alan bilgisi.
2. Modellerin bilimsel hukuk diline uzak olması.
3. Modellerin atf önerisi görevine yabancı olması.

4.3 Bilimsel Makalelerle Eğitilen Dil Modelleri

Deneylerde kullanılan diğer dil modellerinin eğitim kümeleri çeşitli alanlardan bilimsel makaleler içermektedir. Hukuk makalelerinin yer almadığı bu veri kümelerinin alan bazında yeterince dengeli olduğu da söylenemez. Kısacası bu modeller hukuk alanında yeterli bilgiye sahip değildir.

Bu tür modellerin deneylerdeki başarımına baktığımızda, SciNCL ve SGPT modellerinin tutarlı bir şekilde en tepede yer aldıklarını görüyoruz. Ayrıca başarıyı sıralamasında SciBERT < SPECTER < SciNCL şeklindeki sıralama yazarların makalelerinde savunduğu iddialarla da paralellik göstermektedir. SciNCL'in eğitim kümesinin SGPT'ye nazaran daha ağırlıklı olarak bilimsel makalelerden oluşması ve atf temelli bir örneklem seçme stratejisine sahip olması bilgileri ışığında, en başarılı modelin SciNCL olması beklenebilir. Fakat, bilimsel makaleler dışında eğitim için

birçok soru-cevap veri kümesini de kullanan SGPT dil modelinin asıl görevi anlamsal arama problemine bir çözüm üretmektir. SGPT'nin başarımının arkasında yatan sebep olarak bu görevin hukuki atf önerisi göreviyle olan paralelliği verilebilir. SGPT'nin ince ayar sonucu başarımını en az artıran model olması da bunu kanıtlar niteliktedir. Zira SGPT halihazırda hukuki atf önerisi görevine aşınadır.

Özetlenecek olursa, bilimsel makalelerle eğitilen dil modellerinin gösterdiği başarım aşağıdaki faktörlerle ilişkilendirilebilir:

1. Hukuk alan bilgisinden yoksun olmaları.
2. Bazı modellerin hukuki atf önerisi görevine aşinalığı.
3. Etkili bir örneklem seçme stratejisi.

4.4 İki Aşamalı Bilgi Getirimi

Yukarıda da bahsedildiği üzere BM25 hızlı olmasına karşın anlamsal bağlantılar kurmaktan yoksun bir dil modelidir. Kullanılan diğer dil modelleri ise hantal olmalarına rağmen anlamsal ilişkiler kurma konusunda başarılıdır. Bu iki modelin birleştirilmesiyle hem hızlı hem de anlamsal ilişkilerden faydalanan bir çözüm üretilebilir.

İki aşamalı bilgi getiriminde BM25 ile hızlıca getirilen 10 makale, diğer büyük dil modelleri ile yeniden sıralanmıştır. Bu durumda önceden eğitilmiş modeller BM25'in başarımını artıramazken ince ayarlı modellerin istisnasız hepsi bunu başarabilmiştir. Bu noktada kıyas BM25'in başarımı üzerinden yapılmalıdır çünkü büyük dil modelleri için örneklem artık BM25'in getirdiği 10 adet makaleden ibarettir. Yani onların başarımı bu makaleleri ne denli doğru bir şekilde sıralayabildikleri ile ölçülmektedir. Sonuç olarak, tablolar incelendiğinde, SciNCL ile SGPT'nin, BM25'in getirdiği makaleleri sıralamada diğer modellerden daha iyi olduğu görülmüştür.

4.5 Sonuç

BM25 ile SciNCL'in kullanılması ile tüm deney düzenekleri arasında en başarılı sonucun elde edilmesi göstermektedir ki, bilimsel bir çalışma alanı olarak hukuk, elde

edilen sonuçlar bağlamında diğler alanlarla paralellik gösterse de kendine özel bir veri kümesine ihtiyaç duymaktadır. Günümüzde yapılan atıf önerisi çalışmalarında hukuk alanına yer verilmemesi büyük bir eksiklik yaratmakta ve üretilen modellerin bu alana uygulanmasını zorlaştırmaktadır.





5. SONUÇ VE ÖNERİLER

Bu çalışma kapsamında ilk İngilizce Hukuki Atf Önerisi veri kümesi derlenmiş ve bu veri kümesinde, halihazırda Atf Önerisi görevinde başarımlar gösteren modeller ile alana özel eğitilmiş modellerin başarımları karşılaştırılmıştır. Ek olarak, iki aşamalı bilgi getirme yaklaşımı uygulanmıştır.

Elde edilen sonuçlar, giriş bölümünde listelenen hipotezler ışığında şu şekilde listelenebilir:

1. Dil modellerinin Hukuki Atf Önerisi görevinde başarılı olabilmeleri için Atf Önerisi eğitim kümelerinde akademik hukuk makalelerine de yer verilmelidir. Özellikle ince ayar sonrası modellerdeki başarımların artışı bu durumun en büyük destekçisidir.
2. Paralel bir şekilde, hukuki dokümanlarla eğitilen dil modellerinin hukuki dokümanlarda daha kapsayıcı olabilmeleri için eğitim kümelerinde akademik hukuk makaleleri de olmalıdır. Her ne kadar alan aynı olsa da bu modellerin göreve uzak olması da başarımlarını etkilemiştir.
3. İki aşamalı bilgi getirme yöntemi büyük dil modelleri ile BM25'in en iyi yönlerini bir araya getirmekte ve genel başarımlarını artırmaktadır. Bu sayede hem hızlıca veri kümesinden küçük bir örneklem çekilebilmekte, hem de bu örneklem büyük dil modelleri ile daha yüksek doğruluk oranlarında yeniden sıralanabilmektedir. Bu aşamada BM25 türevleri arasında en yüksek başarımlar BM25+ ile elde edilmektedir.
4. Hukuki Atf Önerisi görevinde BM25 ile SciNCL'in beraber kullanımı en başarılı sonucu vermektedir. Bu da SciNCL'de kullanılan kontrollü en yakın komşu örnekleme stratejisinin başarımlarını ortaya koymaktadır.

Sonuçlar incelendiğinde gelecek çalışmalar adına en önemli araştırma noktasının iki aşamalı bilgi getirme yöntemi olduğu görülebilir. Öncelikle BM25'in daha başarılı bir getirme yapması adına, parametrelerin alana uyarlanarak, formülün iyileştirilmesi

üzerine çalışılmalıdır. Ek olarak, literatürde sıklıkla çalışılmadığı için potansiyel olarak düşük başarımlar göstereceği düşünülen ve tezin kapsamına alınmayan diğer bazı BM25 türevleri de araştırmaya dahil edilebilir.

Ek olarak, veri kümesinin boyutunu artırmak üzere çeşitli çalışmalar da yapılabilir, LawArXiv dışında listelen hukuk makaleleri de derlenebilir. Özellikle veri kümesinin devasa boyutlara ulaştığı durumlarda BM25'in performansı ve hız konusunda büyük dil modellerine olan üstünlüğü daha net değerlendirilebilir. Ayrıca elde edilen Hukuki Atıf Önerisi modelinin farklı Hukuki DDİ görevlerine de uygulanıp, başarımlarının test edilmesi, Atıf Öneri görevinde hızlıca işaretli veri elde edilebildiğinden, önemlidir. Özellikle emsal karar bulma, hukuki belge benzerliği ve hukuki karar tahmini gibi doküman temsilinin doğru bir şekilde elde edilmesinin önemli olduğu görevlerde bu durumun oldukça faydalı olabileceği öngörülmektedir.

KAYNAKLAR

- [1] **Governatori, G., Bench-Capon, T., Verheij, B., Araszkievicz, M., Francesconi, E. ve Grabmair, M.** (2022). Thirty years of Artificial Intelligence and Law: the first decade, *Artificial Intelligence and Law*, 30(4), 481–519, <https://doi.org/10.1007/s10506-022-09329-4>.
- [2] **Aletras, N., Tsarapatsanis, D., Preoțiu-Pietro, D. ve Lampos, V.** (2016). Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective, *PeerJ Computer Science*, 2, e93, <https://doi.org/10.7717/peerj-cs.93>.
- [3] **Ma, Y., Shao, Y., Wu, Y., Liu, Y., Zhang, R., Zhang, M. ve Ma, S.** (2021). LeCaRD: A Legal Case Retrieval Dataset for Chinese Law System, *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, <https://doi.org/10.1145/3404835.3463250>.
- [4] **Lastres, S.A.** (2013). *Rebooting legal research in a digital age*.
- [5] **Dainow, J.** (1966). The Civil Law and the Common Law: Some Points of Comparison, *The American Journal of Comparative Law*, 15(3), 419, <https://doi.org/10.2307/838275>.
- [6] **Bornmann, L., Haunschild, R. ve Mutz, R.** (2020). *Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases*.
- [7] **Wahle, J.P., Ruas, T., Mohammad, S.M. ve Gipp, B.** (2022). *D3: A Massive Dataset of Scholarly Metadata for Analyzing the State of Computer Science Research*.
- [8] **Peng, H., Liu, J. ve Lin, C.Y.** (2016). News Citation Recommendation with Implicit and Explicit Semantics, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics.
- [9] **Fetahu, B., Markert, K. ve Anand, A.** (2015). Automated News Suggestions for Populating Wikipedia Entity Pages, *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, ACM.
- [10] **Mahdabi, P. ve Crestani, F.** (2014). Query-Driven Mining of Citation Networks for Patent Citation Retrieval and Recommendation, *Proceedings of the*

- [11] **Huang, Z., Low, C., Teng, M., Zhang, H., Ho, D.E., Krass, M.S. ve Grabmair, M.** (2021). Context-aware legal citation recommendation using deep learning, *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, ACM.
- [12] **Ali, Z., Ullah, I., Khan, A., Jan, A.U. ve Muhammad, K.** (2021). An overview and evaluation of citation recommendation models, *Scientometrics*, 126(5), 4083–4119.
- [13] **Liu, H., Kong, X., Bai, X., Wang, W., Bekele, T.M. ve Xia, F.** (2015). Context-Based Collaborative Filtering for Citation Recommendation, *IEEE Access*, 3, 1695–1703, <https://doi.org/10.1109/access.2015.2481320>.
- [14] **Wang, W., Tang, T., Xia, F., Gong, Z., Chen, Z. ve Liu, H.** (2022). Collaborative Filtering With Network Representation Learning for Citation Recommendation, *IEEE Transactions on Big Data*, 8(5), 1233–1246, <https://doi.org/10.1109/tbdata.2020.3034976>.
- [15] **Sakib, N., Ahmad, R.B. ve Haruna, K.** (2020). A Collaborative Approach Toward Scientific Paper Recommendation Using Citation Context, *IEEE Access*, 8, 51246–51255, <https://doi.org/10.1109/access.2020.2980589>.
- [16] **Ali, Z., Qi, G., Kefalas, P., Abro, W.A. ve Ali, B.** (2020). A graph-based taxonomy of citation recommendation models, *Artificial Intelligence Review*, 53(7), 5217–5260.
- [17] **Pornprasit, C., Liu, X., Kiattipadungkul, P., Kertkeidkachorn, N., Kim, K.S., Noraset, T., Hassan, S.U. ve Tuarob, S.** (2022). Enhancing citation recommendation using citation network embedding, *Scientometrics*, 127(1), 233–264, <https://doi.org/10.1007/s11192-021-04196-3>.
- [18] **Beel, J., Gipp, B., Langer, S. ve Breitingner, C.** (2015). Research-paper recommender systems: a literature survey, *International Journal on Digital Libraries*, 17(4), 305–338.
- [19] **Li, C., Ye, J., Ge, J., Kong, L., Hu, H. ve Luo, B.** (2018). A Novel Convolutional Neural Network for Statutes Recommendation, *Lecture Notes in Computer Science*, Springer International Publishing, s.851–863, https://doi.org/10.1007/978-3-319-97304-3_65.
- [20] **Monroy, A., Calvo, H. ve Gelbukh, A.** (2009). NLP for Shallow Question Answering of Legal Documents Using Graphs, *Computational Linguistics and Intelligent Text Processing*, Springer Berlin Heidelberg, s.498–508, https://doi.org/10.1007/978-3-642-00382-0_40.

- [21] **Souili, A., Cavallucci, D. ve Rousselot, F.** (2015). Natural Language Processing (NLP) – A Solution for Knowledge Extraction from Patent Unstructured Data, *Procedia Engineering*, 131, 635–643, <https://doi.org/10.1016/j.proeng.2015.12.457>.
- [22] **Hendrycks, D., Burns, C., Chen, A. ve Ball, S.** (2021). *CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review*, <https://arxiv.org/abs/2103.06268>.
- [23] **Kim, M.Y., Xu, Y., Lu, Y. ve Goebel, R.**, (2017). Question Answering of Bar Exams by Paraphrasing and Legal Text Analysis, *New Frontiers in Artificial Intelligence*, Springer International Publishing, s.299–313, https://doi.org/10.1007/978-3-319-61572-1_20.
- [24] **Kim, M.Y., Rabelo, J., Okeke, K. ve Goebel, R.** (2022). Legal Information Retrieval and Entailment Based on BM25, Transformer and Semantic Thesaurus Methods, *The Review of Socionetwork Strategies*, 16(1), 157–174, <https://doi.org/10.1007/s12626-022-00103-1>.
- [25] **Bongard, L., Held, L. ve Habernal, I.** (2022). *The Legal Argument Reasoning Task in Civil Procedure*, <https://arxiv.org/abs/2211.02950>.
- [26] **Strickson, B. ve Iglesia, B.D.L.** (2020). Legal Judgement Prediction for UK Courts, *Proceedings of the 2020 The 3rd International Conference on Information Science and System*, ACM, <https://doi.org/10.1145/3388176.3388183>.
- [27] **Nguyen, C., Le, N.K., Nguyen, D.H., Nguyen, P. ve Nguyen, L.M.**, (2022). A Legal Information Retrieval System for Statute Law, *Recent Challenges in Intelligent Information and Database Systems*, Springer Nature Singapore, s.370–382, https://doi.org/10.1007/978-981-19-8234-7_29.
- [28] **Locke, D. ve Zuccon, G.** (2022). *Case law retrieval: problems, methods, challenges and evaluations in the last 20 years*, <https://arxiv.org/abs/2202.07209>.
- [29] **Bhagavatula, C., Feldman, S., Power, R. ve Ammar, W.** (2018). Content-Based Citation Recommendation, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Association for Computational Linguistics.
- [30] **Li, X., Chen, Y., Pettit, B. ve Rijke, M.D.** (2019). Personalised Reranking of Paper Recommendations Using Paper Content and User Behavior, *ACM Transactions on Information Systems*, 37(3), 1–23.
- [31] **Amami, M., Pasi, G., Stella, F. ve Faiz, R.**, (2016). An LDA-Based Approach to Scientific Paper Recommendation, *Natural Language Processing and Information Systems*, Springer International Publishing, s.200–210.

- [32] **Gu, N., Gao, Y. ve Hahnloser, R.H.R.**, (2022). Local Citation Recommendation with Hierarchical-Attention Text Encoder and SciBERT-Based Reranking, *Lecture Notes in Computer Science*, Springer International Publishing, s.274–288, https://doi.org/10.1007/978-3-030-99736-6_19.
- [33] **Nallapati, R.M., Ahmed, A., Xing, E.P. ve Cohen, W.W.** (2008). Joint latent topic models for text and citations, *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, <https://doi.org/10.1145/1401890.1401957>.
- [34] **Ma, X. ve Wang, R.** (2019). Personalized Scientific Paper Recommendation Based on Heterogeneous Graph Representation, *IEEE Access*, 7, 79887–79894.
- [35] **Khadka, A. ve Knoth, P.** (2018). Using citation-context to reduce topic drifting on pure citation-based recommendation, *Proceedings of the 12th ACM Conference on Recommender Systems*, ACM, <https://doi.org/10.1145/3240323.3240379>.
- [36] **Ma, X., Zhang, Y. ve Zeng, J.** (2018). Newly Published Scientific Papers Recommendation in Heterogeneous Information Networks, *Mobile Networks and Applications*, 24(1), 69–79, <https://doi.org/10.1007/s11036-018-1133-9>.
- [37] **Ren, X., Liu, J., Yu, X., Khandelwal, U., Gu, Q., Wang, L. ve Han, J.** (2014). ClusCite, *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, <https://doi.org/10.1145/2623330.2623630>.
- [38] **Mu, D., Guo, L., Cai, X. ve Hao, F.** (2018). Query-Focused Personalized Citation Recommendation With Mutually Reinforced Ranking, *IEEE Access*, 6, 3107–3119, <https://doi.org/10.1109/access.2017.2787179>.
- [39] **Medić, Z. ve Snajder, J.** (2020). Improved Local Citation Recommendation Based on Context Enhanced with Global Information, *Proceedings of the First Workshop on Scholarly Document Processing*, Association for Computational Linguistics, <https://doi.org/10.18653/v1/2020.sdp-1.11>.
- [40] **Lo, K., Wang, L.L., Neumann, M., Kinney, R. ve Weld, D.** (2020). S2ORC: The Semantic Scholar Open Research Corpus, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics.
- [41] **Medić, Z. ve Šnajder, J.** (2022). *Large-scale Evaluation of Transformer-based Article Encoders on the Task of Citation Recommendation*.
- [42] **Brown, P., Kulkarni, A.S., Refai, O. ve Zhou, Y.** (2019). Large expert-curated database for benchmarking document similarity detection in biomedical

literature search, *Database: The Journal of Biological Databases and Curation*, 2019.

- [43] **Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L. ve Su, Z.** (2008). ArnetMiner, *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM.
- [44] **Radev, D.R., Muthukrishnan, P. ve Qazvinian, V.** (2009). The ACL Anthology Network, *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries (NLPIR4DL)*, Association for Computational Linguistics, Suntec City, Singapore, s.54–61, <https://aclanthology.org/W09-3607>.
- [45] **Bird, S., Dale, R., Dorr, B., Gibson, B., Joseph, M., Kan, M.Y., Lee, D., Powley, B., Radev, D. ve Tan, Y.F.** (2008). The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA), Marrakech, Morocco, http://www.lrec-conf.org/proceedings/lrec2008/pdf/445_paper.pdf.
- [46] **Färber, M., Thiemann, A. ve Jatowt, A.** (2018). A High-Quality Gold Standard for Citation-based Tasks, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, European Language Resources Association (ELRA), Miyazaki, Japan, <https://aclanthology.org/L18-1296>.
- [47] **Sugiyama, K. ve Kan, M.Y.** (2013). *Scholarly Paper Recommendation Datasets*.
- [48] **Saier, T. ve Färber, M.** (2019). Bibliometric-Enhanced arXiv: A Data Set for Paper-Based and Citation-Based Tasks, **G. Cabanac, I. Frommholz ve P. Mayr, (düzenleyenler)**, *Proceedings of the 8th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2019) co-located with the 41st European Conference on Information Retrieval (ECIR 2019), Cologne, Germany, April 14, 2019*, cilt2345 of *CEUR Workshop Proceedings*, CEUR-WS.org, s.14–26, <http://ceur-ws.org/Vol-2345/paper2.pdf>.
- [49] **Ozturk, S., Sankur, B., Gungor, T., Yilmaz, M.B., Koroglu, B., Agin, O., Isbilen, M., Ulas, C. ve Ahat, M.** (2014). Turkish labeled text corpus, *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, IEEE, <https://doi.org/10.1109/siu.2014.6830499>.
- [50] **Winkels, R., Boer, A., Vredereg, B. ve Someren, A.V.** (2014). Towards a Legal Recommender System, *International Conference on Legal Knowledge and Information Systems*.

- [51] **Thomas, M., Vacek, T., Shuai, X., Liao, W., Sanchez, G., Sethia, P., Teo, D., Madan, K. ve Custis, T.** (2020). Quick Check: A Legal Research Recommendation System., *NLLP@ KDD*, s.57–60.
- [52] **Ostendorff, M., Ash, E., Ruas, T., Gipp, B., Moreno-Schneider, J. ve Rehm, G.** (2021). Evaluating document representations for content-based legal literature recommendations, *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, ACM.
- [53] **Dhani, J.S., Bhatt, R., Ganesan, B., Sirohi, P. ve Bhatnagar, V.** (2021). *Similar Cases Recommendation using Legal Knowledge Graphs*.
- [54] **Rabelo, J., Kim, M.Y. ve Goebel, R.**, (2023). Semantic-Based Classification of Relevant Case Law, *New Frontiers in Artificial Intelligence*, Springer Nature Switzerland, s.84–95, https://doi.org/10.1007/978-3-031-29168-5_6.
- [55] **Locke, D. ve Zuccon, G.** (2018). A Test Collection for Evaluating Legal Case Law Search, *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, ACM, <https://doi.org/10.1145/3209978.3210161>.
- [56] **Öztürk, C.E., Ömer Köksal, Özçelik, Ş.B. ve Koç, A.** (2022). Prior Case Retrieval for the Court of Cassation of Turkey, *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, IEEE, <https://doi.org/10.1109/sitis57111.2022.00087>.
- [57] **Nigam, S.K. ve Goel, N.** (2022). *nigam@COLIEE-22: Legal Case Retrieval and Entailment using Cascading of Lexical and Semantic-based models*, <https://arxiv.org/abs/2204.07853>.
- [58] **Askari, A., Peikos, G., Pasi, G. ve Verberne, S.** (2022). *LeiBi@COLIEE 2022: Aggregating Tuned Lexical Models with a Cluster-driven BERT-based Model for Case Law Retrieval*, <https://arxiv.org/abs/2205.13351>.
- [59] **Xiao, C., Zhong, H., Guo, Z., Tu, C., Liu, Z., Sun, M., Zhang, T., Han, X., Hu, Z., Wang, H. ve Xu, J.** (2019). *CAIL2019-SCM: A Dataset of Similar Case Matching in Legal Domain*, <https://arxiv.org/abs/1911.08962>.
- [60] **Sarsa, S.** (2019). *Information Retrieval with Finnish Case Law Embeddings*.
- [61] **Harašta, J.** (2019). Case Law Retrieval: Critical Evaluation of Czech Legal Information Retrieval Systems, *Jusletter IT*, 4(21), 1–9.
- [62] **Rosa, G.M., Rodrigues, R.C., Lotufo, R. ve Nogueira, R.** (2021). *Yes, BM25 is a Strong Baseline for Legal Case Retrieval*, <https://arxiv.org/abs/2105.05686>.

- [63] **Hong, Z., Zhou, Q., Zhang, R., Li, W. ve Mo, T.** (2020). Legal Feature Enhanced Semantic Matching Network for Similar Case Matching, *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, <https://doi.org/10.1109/ijcnn48605.2020.9207528>.
- [64] **Askari, A. ve Verberne, S.** (2021). Combining Lexical and Neural Retrieval with Longformer-based Summarization for Effective Case Law Retrieval, *Biennial Conference on Design of Experimental Search & Information Retrieval Systems*.
- [65] **Tran, V., Nguyen, M.L. ve Satoh, K.** (2019). Building Legal Case Retrieval Systems with Lexical Matching and Summarization using A Pre-Trained Phrase Scoring Model, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, ACM, <https://doi.org/10.1145/3322640.3326740>.
- [66] **Kershaw, D.J., Pettit, B., Hristakeva, M. ve Jack, K.** (2020). Learning to Rank Research Articles: A Case Study of Collaborative Filtering and Learning to Rank in ScienceDirect, *BIR@ECIR*.
- [67] **Jia, H. ve Saule, E.** (2018). *Graph Embedding for Citation Recommendation*, <https://arxiv.org/abs/1812.03835>.
- [68] **Ayala-Gómez, F., Daróczy, B., Benczúr, A., Mathioudakis, M. ve Gionis, A.** (2018). Global citation recommendation using knowledge graphs, *Journal of Intelligent & Fuzzy Systems*, 34(5), 3089–3100, <https://doi.org/10.3233/jifs-169493>.
- [69] **Wang, L., Rao, Y., Bian, Q. ve Wang, S.,** (2020). Content-Based Hybrid Deep Neural Network Citation Recommendation Method, *Communications in Computer and Information Science*, Springer Singapore, s.3–20, https://doi.org/10.1007/978-981-15-7984-4_1.
- [70] **Nair, A.M., Benny, O. ve George, J.,** (2021). Content Based Scientific Article Recommendation System Using Deep Learning Technique, *Inventive Systems and Control*, Springer Singapore, s.965–977, https://doi.org/10.1007/978-981-16-1395-1_70.
- [71] **Long, K., Li, S., Wang, P., Tang, J. ve Wang, T.** (2022). Integrating Title and Citation Context Semantics of Citing Paper via Weighted Attentions for Local Citation Recommendation, *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, <https://doi.org/10.1109/ijcnn55064.2022.9892003>.
- [72] **Medić, Z. ve Šnajder, J.** (2022). An empirical study of the design choices for local citation recommendation systems, *Expert Systems with Applications*, 200, 116852, <https://doi.org/10.1016/j.eswa.2022.116852>.

- [73] **Yang, L., Zhang, Z., Cai, X. ve Dai, T.** (2019). Attention-Based Personalized Encoder-Decoder Model for Local Citation Recommendation, *Computational Intelligence and Neuroscience*, 2019, 1–7, <https://doi.org/10.1155/2019/1232581>.
- [74] **Dai, T., Zhu, L., Wang, Y., Zhang, H., Cai, X. ve Zheng, Y.** (2019). Joint Model Feature Regression and Topic Learning for Global Citation Recommendation, *IEEE Access*, 7, 1706–1720, <https://doi.org/10.1109/access.2018.2884981>.
- [75] **Ali, Z., Qi, G., Muhammad, K., Khalil, A., Ullah, I. ve Khan, A.** (2021). Global Citation Recommendation employing Multi-view Heterogeneous Network Embedding, *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, IEEE, <https://doi.org/10.1109/ciss50987.2021.9400311>.
- [76] **Dai, T., Yan, W., Zhang, K., Qiu, C., Zhao, X. ve Pan, S.** (2021). Gated relational stacked denoising autoencoder with localized author embedding for global citation recommendation, *Expert Systems with Applications*, 184, 115359, <https://doi.org/10.1016/j.eswa.2021.115359>.
- [77] **Cai, X., Wang, N., Yang, L. ve Mei, X.** (2022). Global-local neighborhood based network representation for citation recommendation, *Applied Intelligence*, 52(9), 10098–10115, <https://doi.org/10.1007/s10489-021-02964-5>.
- [78] **Choi, J., Honda, U., Watanabe, T., Ouchi, H. ve Inui, K.** (2022). Law Retrieval with Supervised Contrastive Learning Using the Hierarchical Structure of Law, *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, De La Salle University, Manila, Philippines, s.590–599, <https://aclanthology.org/2022.pacllic-1.65>.
- [79] **Li, H., Ai, Q., Chen, J., Dong, Q., Wu, Y., Liu, Y., Chen, C. ve Tian, Q.** (2023). *SAILER: Structure-aware Pre-trained Language Model for Legal Case Retrieval*, <https://arxiv.org/abs/2304.11370>.
- [80] **Fink, T., Recski, G., Kusa, W. ve Hanbury, A.** (2023). *Statute-enhanced lexical retrieval of court cases for COLIEE 2022*, <https://arxiv.org/abs/2304.08188>.
- [81] **Bellandi, V., Castano, S., Ceravolo, P., Damiani, E., Ferrara, A., Montanelli, S., Picascia, S., Polimeno, A. ve Riva, D.** (2022). Knowledge-Based Legal Document Retrieval: A Case Study on Italian Civil Court Decisions.
- [82] **Kim, M.Y., Rabelo, J., Goebel, R., Yoshioka, M., Kano, Y. ve Satoh, K.** (2023). COLIEE 2022 Summary: Methods for Legal Document Retrieval and Entailment, *New Frontiers in Artificial Intelligence*, Springer Nature Switzerland, s.51–67, https://doi.org/10.1007/978-3-031-29168-5_4.

- [83] **Bhattacharya, P., Ghosh, K., Ghosh, S., Pal, A., Mehta, P., Bhattacharya, A. ve Majumder, P.** (2019). FIRE 2019 AILA Track, *Proceedings of the 11th Forum for Information Retrieval Evaluation*, ACM, <https://doi.org/10.1145/3368567.3368587>.
- [84] **Robertson, S., Walker, S., Jones, S., Hancock-Beaulieu, M.M. ve Gattford, M.** (1995). Okapi at TREC-3, *Overview of the Third Text REtrieval Conference (TREC-3)*, Gaithersburg, MD: NIST, s.109–126.
- [85] **Mikolov, T., Chen, K., Corrado, G. ve Dean, J.** (2013). *Efficient Estimation of Word Representations in Vector Space*.
- [86] **Chalkidis, I.** (2018). *Law2Vec: Legal Word Embeddings*, <https://archive.org/details/Law2Vec>.
- [87] **Mandal, A., Ghosh, K., Ghosh, S. ve Mandal, S.** (2021). Unsupervised approaches for measuring textual similarity between legal court case reports, *Artificial Intelligence and Law*.
- [88] **Devlin, J., Chang, M.W., Lee, K. ve Toutanova, K.** (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.
- [89] **Beltagy, I., Lo, K. ve Cohan, A.** (2019). SciBERT: Pretrained Language Model for Scientific Text, *EMNLP*, arXiv:1903.10676.
- [90] **Cohan, A., Feldman, S., Beltagy, I., Downey, D. ve Weld, D.** (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics.
- [91] **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. ve Polosukhin, I.** (2017). *Attention Is All You Need*, <https://arxiv.org/abs/1706.03762>.
- [92] **Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N. ve Androutsopoulos, I.** (2020). LEGAL-BERT: The Muppets straight out of Law School, *Findings of the Association for Computational Linguistics: EMNLP 2020*, Association for Computational Linguistics.
- [93] **Ostendorff, M., Rethmeier, N., Augenstein, I., Gipp, B. ve Rehm, G.** (2022). *Neighborhood Contrastive Learning for Scientific Document Representations with Citation Embeddings*.
- [94] **Muennighoff, N.** (2022). *SGPT: GPT Sentence Embeddings for Semantic Search*.
- [95] **Radford, A., Narasimhan, K., Salimans, T. ve Sutskever, I.** (2018). Improving language understanding by generative pre-training.
- [96] **Schroff, F., Kalenichenko, D. ve Philbin, J.** (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering, <https://arxiv.org/abs/1503.03832>.

- [97] **Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q. ve Rush, A.M.** (2019). *HuggingFace's Transformers: State-of-the-art Natural Language Processing*.
- [98] **Reimers, N. ve Gurevych, I.** (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, <https://arxiv.org/abs/1908.10084>.
- [99] **Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., Abrego, G.H., Yuan, S., Tar, C., Sung, Y.H., Strope, B. ve Kurzweil, R.** (2019). *Multilingual Universal Sentence Encoder for Semantic Retrieval*, <https://arxiv.org/abs/1907.04307>.
- [100] **Reimers, N. ve Gurevych, I.** (2020). *Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation*, <https://arxiv.org/abs/2004.09813>.
- [101] **Schweter, S.** (2020). *BERTurk - BERT models for Turkish*, <https://zenodo.org/record/3770924>.
- [102] **Clark, K., Luong, M.T., Le, Q.V. ve Manning, C.D.** (2020). *ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators*, <https://arxiv.org/abs/2003.11055>.
- [103] **Jiang, Z., Yu, W., Zhou, D., Chen, Y., Feng, J. ve Yan, S.** (2020). *ConvBERT: Improving BERT with Span-based Dynamic Convolution*, <https://arxiv.org/abs/2008.02496>.
- [104] **Dadgostari, F., Guim, M., Beling, P.A., Livermore, M.A. ve Rockmore, D.N.** (2020). Modeling law search as prediction, *Artificial Intelligence and Law*, 29(1), 3–34.
- [105] **Locke, D. ve Zuccon, G.** (2022). *Case law retrieval: problems, methods, challenges and evaluations in the last 20 years*.
- [106] **Bhattacharya, P., Ghosh, K., Pal, A. ve Ghosh, S.** (2020). Methods for Computing Legal Document Similarity: A Comparative Study, *ArXiv*, [abs/2004.12307](https://arxiv.org/abs/2004.12307).
- [107] **Cui, J., Shen, X., Nie, F., Wang, Z., Wang, J. ve Chen, Y.** (2022). *A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges*.
- [108] **Aydemir, İ.A.** (2010). TÜRK HUKUK DİLİ ÜZERİNE ARAŞTIRMALAR I, *Türk Dili Araştırmaları Yıllığı-Belleten*, 58(2), 19–36.

- [109] **Chalkidis, I.** (2021). *Deep neural networks for information mining from legal texts*, Athens University Economics and Business (AUEB); Οικονομικό Πανεπιστήμιο Αθηνών.
- [110] **Altay, A.A.** (2002). İngilizce ve Türkçe hukuk dillerinin özellikleri: karşılaştırmalı bir çalışma, *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 19(2).
- [111] **Trotman, A., Jia, X. ve Crane, M.** (2012). Towards an Efficient and Effective Search Engine, *OSIR@SIGIR*.
- [112] **Lv, Y. ve Zhai, C.** (2011). Adaptive term frequency normalization for BM25, *Proceedings of the 20th ACM international conference on Information and knowledge management*, ACM, <https://doi.org/10.1145/2063576.2063871>.
- [113] **Lv, Y. ve Zhai, C.,** (2012). A Log-Logistic Model-Based Interpretation of TF Normalization of BM25, *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, s.244–255, https://doi.org/10.1007/978-3-642-28997-2_21.
- [114] *CORE Dataset*, <https://core.ac.uk/documentation/dataset>.
- [115] *CiteSeerX Data*, <https://csxstatic.ist.psu.edu/downloads/data>.
- [116] *MEDLINE/PubMed Data*, https://www.nlm.nih.gov/databases/download/pubmed_medline.html.
- [117] *LawArXiv Papers*, <https://osf.io/preprints/lawarxiv>.
- [118] *Google Scholar*, <https://scholar.google.com>.
- [119] *Google Scholar API*, <https://serpapi.com/google-scholar-api>.
- [120] *pdfplumber*, <https://github.com/jsvine/pdfplumber>.



EKLER

EK A1: Kanada Yüksek Mahkemesi'nin Örnek Kararı

EK A2: Türkiye Cumhuriyeti Yargıtay'ının Örnek Kararı

EK B1: Terimler Sözlüğü





EK A1



SUPREME COURT OF CANADA

CITATION: *R. v. Haevischer*, 2023 SCC 11

APPEAL HEARD: October 4, 2022
JUDGMENT RENDERED: April 28, 2023
DOCKET: 39635

BETWEEN:

His Majesty The King
Appellant

and

Cody Rae Haevischer and Matthew James Johnston
Respondents

- and -

Director of Public Prosecutions, Attorney General of Ontario, Criminal Lawyers' Association of Ontario, Independent Criminal Defence Advocacy Society, Criminal Trial Lawyers' Association, Trial Lawyers Association of British Columbia and Canadian Civil Liberties Association
Interveners

...

[45] Since 1996, *Vukelich* hearings have been extended well beyond search warrant cases and are frequently used in criminal trials in British Columbia. While it is more common for the Crown to request a *Vukelich* hearing to summarily dismiss a defence application, the trial judge's screening function applies equally to Crown applications, and defence counsel do sometimes request *Vukelich* hearings (see *R. v. Cody*, 2017 SCC 31, [2017] 1 S.C.R. 659, at para. 38; *R. v. Biring*, 2021 BCSC 2678, at para. 5 (CanLII); *R. v. Kuntz-Angel*, 2020 BCSC 1777, at para. 71 (CanLII)).

(2) Underlying Values: Trial Efficiency and Trial Fairness

[46] The standard selected for summary dismissal on a *Vukelich*-type hearing will be based on the two sets of underlying values at play in such proceedings: trial efficiency and trial fairness. These values coexist and "both must be pursued in order for each to be realised: they are, in practice, interdependent" (*R. v. Jordan*, 2016 SCC 27, [2016] 1 S.C.R. 631, at para. 27, quoting B.C. Justice Reform Initiative, *A Criminal Justice System for the 21st Century* (2012), at p. 75).

...

Şekil A.1 : Kanada Yüksek Mahkemesi'nin Örnek Kararı.

EK A2

Ceza Genel Kurulu 2022/403 E. , 2022/584 K.

"İçtihat Metni"

Mahkemesi :Asliye Ceza

Hükümlü : ...

Hükümlü ... hakkında hırsızlık suçundan açılan kamu davasında yapılan yargılama sonucunda hükümlünün atılı suçtan 5237 sayılı TCK'nın 142/1-b, 53 ve 63. maddeleri gereğince 4 yıl hapis cezası ile cezalandırılmasına, hak yoksunluğuna ve mahsuba ilişkin ... 2. Asliye Ceza Mahkemesince verilen 05.10.2006 tarihli ve 79-643 sayılı hükmün, hükümlü ve katılanlar tarafından temyiz edilmesi üzerine dosyayı inceleyen Yargıtay 6. Ceza Dairesince 03.12.2013 tarih ve 2238-25647 sayı ile; "... Diğer temyiz itirazları yerinde görülmemiştir.

...

Adres kayıt sistemindeki adresine (MERNİS) tebliğatin yapılma şekli ise anılan Kanun'un "Tebliğ imkansızlığı ve tebellüğüden imtina" başlıklı 21. maddesinin 2. fıkrasında; "Gösterilen adres muhatapın adres kayıt sistemindeki adresi olup, muhatap o adreste hiç oturmamış veya o adresten sürekli olarak ayrılmış olsa dahi, tebliğ memuru tebliğ olunacak evraki, o yerin muhtar veya ihtiyar heyeti azasından birine veyahut zabıta amir veya memurlarına imza karşılığında teslim eder ve teslim edenin adresini ihtiva eden ihbarnamayı gösterilen adresteki binanın kapısına yapıştırır. İhbarnamenin kapağa yapıştırıldığı tarih, tebliğ tarihi sayılır." biçiminde açıklanmıştır. Diğer taraftan, 2797 sayılı Yargıtay Kanunu'nun "Yargıtay Cumhuriyet Başsavcısı Başyardımcısı ile Yardımcılarının Görevleri" başlıklı 28. maddesi; "Yargıtay Cumhuriyet Başsavcısı Başyardımcısı, Yargıtay Cumhuriyet Başsavcısının veya Yargıtay Cumhuriyet Başsavcivekilinin vereceği görevleri yapar. Özel hükümler saklı kalmak kaydıyla Yargıtay Cumhuriyet Başsavcısı ve Vekilinin yokluğunda Yargıtay Cumhuriyet Başsavcısına vekillik eder. Başyardımcının da bulunmadığı zamanlarda bu görevi en kıdemli Yargıtay Cumhuriyet Başsavcıyardımcısı yapar. Yargıtay Cumhuriyet Başsavcıyardımcıları, kendilerine verilen dosyaların tebliğnamelerini, karar düzeltme ve itiraz yoluna başvurma işlemlerini Yargıtay Cumhuriyet Başsavcısı adına düzenler ve onun yerine imza ederler. Yargıtay Cumhuriyet Başsavcısının vereceği diğer işleri görürler.

Zamanaşımı Türk Ceza Kanununun 102 nci maddesinin birinci fıkrasının (5) ve (6) numaralı bentleri kapsamında olan suçlara ilişkin dava dosyaları temyiz üzerine Yargıtay Cumhuriyet Başsavcılığınca tebliğname düzenlenmeksizin ilgili daireye gönderilir. Daire kararına karşı Yargıtay Cumhuriyet Başsavcılığının karar düzeltme ve itiraz kanun yollarına başvurma yetkisi vardır." şeklinde hüküm altına alınmıştır.

...

Şekil A.2 : Türkiye Cumhuriyeti Yargıtay'ının Örnek Kararı

EK B1

Çizelge B.1 : Tez kapsamında kullanılan kavramların İngilizce literatürdeki karşılıkları.

Türkçe	İngilizce
Atıf önerisi	Citation recommendation
Çizge temelli filtreleme	Graph-based filtering
Doğal dil işleme	Natural language processing
Hukuki atıf önerisi	Legal citation recommendation
Hukuki gerekçelendirme	Legal reasoning
Hukuki metin ilişkilendirme	Legal textual entailment
Hukuki soru cevaplama	Legal question answering
İçerik temelli filtreleme	Content-based filtering
İçtihat bulma	Case retrieval
İlk oylayan problemi	First-rater problem
İşbirlikçi filtreleme	Collaborative filtering
Küresel atıf önerme	Global citation recommendation
Mean average precision	Ortalama kesinliklerin ortalaması
Mean reciprocal rank	Sıralamaların terslerinin ortalaması
Otomatik kodlayıcı	Autoencoder
Öz	Abstract
Soğuk başlangıç problemi	Cold-start problem
Şilin saldırısı	Shilling attack
Tüzük bulma	Statute retrieval
Üçlü kayıp fonksiyonu	Triplet loss function
Yerel atıf önerme	Local citation recommendation



ÖZGEÇMİŞ

Adı Soyadı: Dođukan ARSLAN

ÖĐRENİM DURUMU:

- **Lisans:** 2020, İstanbul Teknik Üniversitesi, Bilgisayar ve Bilişim Fakültesi, Bilgisayar Mühendisliği Bölümü

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE PATENTLER:

- **Arslan D.**, Erdoğan S. S., Eryiđit G. (2023). Citation Recommendation on Scholarly Legal Articles. *Seventeenth International Workshop on Juris-informatics on International Symposia on AI*, Haziran 5-6, 2023 Kyoto, Japan.

DİĐER YAYINLAR, SUNUMLAR VE PATENTLER:

- **Arslan D.** and Eryiđit G. (2021). Evaluation of Wizard-of-Oz and Self-Play Data Collection Techniques for Turkish Goal-Oriented Dialogue Agents. *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2021 Kocaeli, Turkey.