

Yargıtay Kararlarının Mükerrer Sinir Ağları ile Tahmini

Predicting Outcomes of the Court of Cassation of Turkey with Recurrent Neural Networks

Ceyhun E. ÖZTÜRK

Elektrik ve Elektronik Mühendisliği Bölümü
Bilkent Üniversitesi
ASELSAN Research Center
Ankara, Türkiye
ceyhun.ozturk@bilkent.edu.tr

Ş. Barış ÖZÇELİK

Hukuk Fakültesi
Bilkent Üniversitesi
Ankara, Türkiye
bozcelik@bilkent.edu.tr

Aykut KOÇ

Elektrik ve Elektronik Mühendisliği Bölümü
Ulusal Manyetik Rezonans Araştırma Merkezi
Bilkent Üniversitesi
Ankara, Türkiye
aykut.koc@bilkent.edu.tr

Özetçe —Doğal dil işleme (NLP) tabanlı yaklaşımlar hukuk çalışmalarında son dönemde çok popüler hâle gelmiştir. Bu çalışmada Yargıtay davalarının sonuçları derin öğrenme modelleri kullanılarak tahmin edilmiştir. Bu modeller mükerrer sinir ağı türevi olan GRU, LSTM ve BiLSTM'dir. Modeller eğitim sırasında karar metinlerinin sadece olay tanımları olan kısımlarını görmüştür. İlk olarak modeller günlük dilden metinlerle üretilen kelime temsilleriyle eğitilmiştir. Daha sonra modeller Türk mahkemelerinden indirilen davalarla üretilen kelime temsilleriyle eğitilmiştir. Modeller üzerinde yapılan deneylerin sonuçları karşılaştırmalı ve detaylı biçimde verilmiştir. Bu çalışma ve önceki çalışmalara bakılarak Yargıtay kararlarının hukukta yürütülen NLP çalışmalarında incelenen çoğu mahkemenin daha yüksek isabetle tahmin edilebildiği görülmektedir. En başarılı karar tahmini yapan model GRU olarak gözlenmiştir. GRU modeli ile karar tahmininde %96.8 doğruluk skoruna ulaşılmıştır.

Anahtar Kelimeler—Doğal dil işleme, hukuk, makine öğrenmesi, hukukta yapay zeka, hukuki metin işleme, derin öğrenme

Abstract—Natural Language Processing (NLP) based approaches have recently become very popular for studies in legal domain. In this work, the outcomes of the cases of the Court of Cassation of Turkey were predicted with the use of Deep Learning models. These models are GRU, LSTM and BiLSTM which are variants of the recurrent neural network. Models saw only fact description parts of the case decision texts during training. Firstly, the models were trained with the word embeddings that were created from the texts from daily language. Then, the models were trained with the word embeddings that were created from downloaded legal cases from Turkish courts. The results of the experiments on the models are given in a comparative and detailed manner. It is observed based on this study and the past studies that the outcomes of the Court of Cassation can be predicted with higher accuracy than most of the courts that were investigated in previous legal NLP studies. The model which is best at predicting decisions is GRU. The GRU model achieves 96.8% accuracy in the decision prediction task.

Keywords—Natural Language Processing, Law, Machine Learning, AI in Law, Legal text mining, Deep Learning.

I. GİRİŞ

Hukuk toplumların ihtiyaçlarını karşılamak için sürekli değişen bir alandır. Hukuk profesyonelleri alandaki değişimler sebebiyle sürekli artan bir yük ile karşı karşıyadırlar. Diğer taraftan, bilgisayarlar büyük miktarda hukuki metni kısa sürede işleyebilmektedir. Bu yüzden bilgisayar bilimi topluluğunda hukuki metin işleme popüler bir araştırma alanıdır.

Kelimeleri kelime temsilleri/vektörleri olarak işlemek doğal dil işlemede 1980'lerden beri popüler bir yöntemdir [1]–[3]. Günümüzde word2vec [4], GloVe [5], fastText [6] ile BERT [7] ve ELMo [8] gibi dönüştürücü (*Ing.* transformer) algoritmaları kelime temsili üretmede kullanılan popüler algoritmalar. Literatürde Türkçe hukuk çalışmaları için ön eğitimi yapılmış dönüştürücü modeli olmaması ve dönüştürücülerin eğitilmesinin yüksek kapasiteli donanımlar ve çok zaman gerektirmesi sebebiyle [9], bu çalışmada word2vec algoritmasıyla üretilmiş kelime temsilleri kullanılmıştır. Türkçe hukukta yapılacak gelecek NLP çalışmaları için alana özel dönüştürücü modeli eğitilmesi önemlidir. Kelime temsilleri 2 ayrı kümeden alınmıştır. İlk küme Wikipedia makaleleri üzerinde eğitilmiş kelime temsillerinden oluşmaktadır [10]. İkinci küme ise hukuki metinler üzerinde eğitilmiş kelime temsillerinden oluşmaktadır. İkinci küme bu bildiriye hazırlayan araştırma grubunun hukukla ilgili diğer çalışmalarında da kullanılmaktadır.

Türk hukuk sistemi üzerinde doğal dil işleme konulu bir çalışmada bölge idare ve bölge adliye mahkemeleri ile Anayasa Mahkemesinde karar tahmini yapılmıştır. Bu çalışmada çeşitli geleneksel makine öğrenmesi yöntemleri ve derin öğrenme yöntemleri kullanılmıştır [11]. Bu çalışmada kullanılan araçlardan bildiride de faydalanılmıştır. Dünya çapında mahkeme karar tahmini üzerine pek çok çalışma bulunmaktadır [12]–[19]. Ancak bu çalışmalar genellikle tek bir makine öğrenmesi yöntemi kullandıkları için sınırlı bir kapsamdadır. Bunun yanında Türkçe hukuk için doğal dil işleme çalışmaları konusunda literatürde bir boşluk bulunmaktadır. Bu bildiri kapsamında derin öğrenme modelleri bildirinin bir önceki paragrafında

bahsi geçen kelime temsilleri kullanılarak Yargıtay kararlarını tahmin etmek için eğitilmiştir. Derin öğrenme modellerinin analizi ile geniş kapsamlı bir analiz sunulmuştur. Mahkeme karar tahmini uygulamalarının kullanışlı olması için yüksek başarıya sahip olması gerekmektedir. Bu çalışmada literatürde elde edilen performans skorlarından daha iyi sonuçlar elde edilmiştir.

Bildirinin organizasyonu şu şekilde olacaktır. Literatürdeki ilgili çalışmalar Bölüm II'de bahsedilecektir. Çalışmada yapılan veri ön işleme yöntemleri ve kullanılan modeller Bölüm III'te anlatılacaktır. Modeller ile yapılan deneylerin sonuçları Bölüm IV'te analiz edilecektir. Son olarak Bölüm V'te çalışmanın sonuçları tartışılacaktır.

II. İLGİLİ ÇALIŞMALAR

Hukukun yapay zekadan faydalanabileceği fikri literatürde on yıllardır incelenmiş bir konudur. Bu fikir 1970'te yayınlanan bir makalede [20] tartışılmıştır. Makalede yapay zeka kullanımının avukatları davalar üzerinde harcanan uzun zamandan tasarruf ettirebileceği ve avukatların yaptığı hataları azaltabileceği savunulmuştur. Bununla beraber yapay zeka kullanımının ortaya çıkaracağı sosyal problemlerden çekinceler dile getirilmiştir. Örneğin, zengin kişiler davalarında yapay zeka kullanarak daha zengin olabilirlerdi [20]. Bir araştırma makalesinde yapay zeka ve hukuk alanlarının kesişimiyle ilgili araştırmalarla aktif ilgilenen topluluğun 1987'de Uluslararası Hukuk ve Yapay Zeka Konferansı (ICAIL) ilk kez toplandığında olduğu söylenmiştir [21]. [22]'de hukuktaki doğal dil işleme uygulamaları derinlemesine incelenmiştir .

word2vec algoritması 2013'te Google'daki bir araştırma takımı tarafından yayınlanmıştır. Araştırmacı takımı benzer anlamdaki kelime çiftlerinin yüksek kosinüs benzerliğine sahip olduğunu gözlemlemiştir. word2vec algoritması aslında bir nöral ağ modelidir ve aynı zamanda negatif örneklemeli skip-gram modeli (SGNS) olarak bilinir [4]. Kelime temsillerinin eğitildikleri metinlerde bulunan ön yargıları öğrendiği gözlenmiştir. Bu durum kelime temsillerinin uygulamalarda kullanılması ile ilgili çekincelere sebep olmaktadır. Bir çalışmada Türkçede İngilizceye kıyasla daha az cinsiyetçi ön yargı bulunduğu gözlenmiştir [23]. Bir başka çalışmada [24] İngilizce hukuk metinleri için eğitilen kelime temsillerindeki cinsiyetçi ön yargının giderilmesi için bir yöntem geliştirilmiştir.

Türk hukuk sisteminin incelendiği bir makalede [11] bu bildiride kullanılan ilk kelime temsili kümesi (Wikipedia makaleleri ile eğitilen temsiller) kullanılmıştır. Kümedeki kelime vektörleri 400 elemanlıdır. Bu küme bildiride Wikipedia kümesi olarak adlandırılacaktır. Makalede [11] geleneksel makine öğrenmesi yöntemi olarak karar ağacı, rassal orman ve vektör destek makinesi modelleri; derin öğrenme yöntemi olarak ise birer mükerrer sinir ağı (RNN) türevi olan mükerrer birimli sinir ağı (GRU) [25] ve uzun ve kısa vade hafızalı sinir ağı (LSTM) [26] modelleri kullanılmıştır. Öğrenme modelleri mahkemelerin karar metinlerini işlemektedir. İşlenen karar metinlerinden son kararın bulunduğu kısım çıkarılmıştır. İncelenen dava etiketleri ikilidir. Yani dava sonucu yapılan itirazın kabul edilmesi ya da reddedilmesi olarak sınıflandırılmaktadır. Davalarda başka sonuçlar da olabilmektedir ancak bu sonuçlar dikkate alınmamıştır. Her mahkeme için ayrı modeller eğitilmiştir. Her mahkeme için eğitim, geçerleme ve test kümeleri için davaların sırasıyla %70'i, %15'i ve %15'i ayrılmıştır [11].

Kanuni metin işleme alanındaki pek çok akademik çalışmada mükerrer sinir ağı (RNN) türevleri ya da RNN türevlerini içeren ağlar kullanılmıştır [16], [22], [27], [28]. Bir makalede [15] incelenen metinlerde geçen ilgili hukuk ve olguların kodlanması için çift yönlü GRU ve dikkat mekanizmalarından oluşan bir sistem geliştirilmiştir. Bu model, ikili sınıflandırma için Tayland Yüksek Mahkemesinin 1000'den fazla kararından oluşan bir veri kümesinde değerlendirilmiştir ve bulunan en iyi makro F1 skoru %63'tür [15]. Başka bir makalede GRU türevi bir model Çin Halk Cumhuriyeti Üst Mahkemesi davalarının kararlarını tahmin etmek amacıyla kullanılmıştır [16]. Dikkat mekanizmalarının (katmanlarının) uzun metinler işlemek için RNN türevlerine bağlanması oluşturulan modellerin daha iyi performans vermesini sağlamaktadır. Bu yaklaşımda kelimelelerin kendileri ya da nöral birimlerin iç durumları birer dikkat skoruyla ağırlıklandırılır ve dikkat skorlarının ortalaması ile metnin son temsili elde edilir. Dikkat katmanları parametre sayısını arttırsa da metinlerin önemli kısımlarının ayırt edilebilmesini sağlar ve öğrenme modellerinin eğitimi sırasında daha iyi gradyan akışı sağlar.

III. METODOLOJİ

A. Veri Hazırlama ve Öğrenme Modelleri

Bu çalışmada ilk olarak Yargıtayda görülen 332.662 davanın karar metni internetten indirilmiştir. Tasnifleme programlama kodu tarafından otomatik yapılmıştır. Hazırlanan kod davalarda "Somut olayda" ve "gerektirmiştir" kelime öbeklerini aramıştır. Bu kelime öbeklerinden en az birini bulduran tüm davalar (toplamda 59.822 dava) tasniflenmiştir. Karar metinlerinde bu kelime öbekleri geçtikten sonra son hükümler yer almaktadır. Bu yüzden bu kelime öbeklerinin geçtiği satırdan sonraki satırlar öğrenme modellerine gösterilmemiştir. Sonraki satırlarda "ONANMASINA" kelimesi geçmesi hâlinde dava itirazın reddedilmesi (onanma) olarak sınıflandırılmaktadır. "BOZULMASINA" kelimesi geçmesi hâlinde ise dava itirazın kabul edilmesi (bozulma) olarak sınıflandırılmaktadır.

Oluşturulan veri kümesinin oldukça dengesiz olduğu gözlenmiştir. Davaların 55.113 tanesinin sonucu bozulma, 4.709 tanesinin sonucu ise onanmadır. Başka Türk yüksek mahkemelerinde de karar sayılarında dengesizlik bulunmaktadır [11]. Bu problemi çözmek için öğrenme modellerinde başka bir çalışmada da [29] kullanılmış olan çapraz entropi kayıp fonksiyonunun sınıf ağırlıklı versiyonu kullanılmıştır. Veri kümesindeki dengesizlik sebebiyle sınıflandırma problemlerinde az örneği bulunan sınıflardaki örneklerin doğru etiketlenmesine ağırlık veren makro yakalama ve makro F1 gibi performans metrikleri önem kazanmaktadır. Çalışmada karar metinlerinin tokenlere ayrıştırılması için literatürdeki benzer başka bir çalışmada da [11] kullanılan Zemberek [30] isimli Türkçe doğal dil işleme aracı kullanılmıştır. Her mahkeme için eğitim, geçerleme ve test kümeleri için davaların sırasıyla %70'i, %15'i ve %15'i ayrılmıştır.

Çalışmada sınıflandırma yöntemi olarak çeşitli mükerrer sinir ağı (RNN) türevi olan derin öğrenme modelleri kullanılmıştır. Bu modeller GRU, LSTM ve çift yönlü LSTM'dir (BiLSTM). Her modelin dikkat mekanizmalı ve dikkat mekanizmasız versiyonları kullanılmıştır. Mükerrer sinir ağı türevleri önemli bilgileri iç hafıza mekanizmalarında tutma becerileri sebebiyle uzun metinleri işlemek için uygun modellerdir. Dikkat mekanizmalı modellerde dikkat mekanizmasından

sonra, diğer modellerde ise sınır ağından sonra bir lineer katman kullanılmıştır. Bu lineer katmanın sonuna bağlanan softmax fonksiyonu ile karar tahmini elde edilmiştir. Modellerin değerlendirilmesi için doğruluk, makro yakalama ve makro F1 performans metrikleri kullanılmıştır. Aşağıda doğruluk metriği tanımlanmıştır.

$$\text{Doğruluk Oranı} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Yukarıdaki tanımda TP, TN, FP ve FN sırasıyla gerçek pozitif, gerçek negatif, yanlış pozitif ve yanlış negatif örnek sayılarıdır. Aşağıda yakalama, kesinlik ve F1 performans metrikleri tanımlanmıştır.

$$\text{Yakalama Oranı} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Kesinlik Oranı} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{F1 Oranı} = \frac{2 \cdot \text{Kesinlik} \cdot \text{Yakalama}}{\text{Kesinlik} + \text{Yakalama}} \quad (4)$$

Her bir sınıf sırasıyla pozitif alındığında ve bir sınıf pozitifken diğer sınıflar negatif alındığında her sınıf için yukarıdaki tanımlar kullanılarak bir yakalama skoru ve bir F1 skoru bulunabilmektedir. Yakalama skorlarının ortalaması ile makro yakalama skoru, F1 skorlarının ortalaması ile ise makro F1 skoru bulunmaktadır. Mahkeme karar tahmini için derin öğrenme yöntemleri kullanan literatürde yayınlanmış başlıca çalışmalarda elde edilen en iyi doğruluk ve makro F1 skorları aşağıdaki tabloda verilmiştir. Tabloda çalışmalarda yayınlanmamış değerler "-" olarak gösterilmiştir. vd. ise "ve diğerleri"nin kısaltmasıdır.

Çalışma	Mahkeme Adı	Doğruluk (%)	Makro F1 (%)
[15]	Tayland Yüksek Mahkemesi	-	63,0
[16]	Çin Yüksek Mahkemesi	82,2	83,0
[11]	T.C. Anayasa Mahkemesi	91,8	67,0
[11]	T.C. Bölge Adliye Mahkemeleri Hukuk Mahkemesi	67,3	65,0
[11]	T.C. Bölge Adliye Mahkemeleri Ceza Mahkemesi	85,6	77,0
[11]	T.C. Bölge İdare Mahkemeleri İdari Mahkemesi	91,1	76,0
[11]	T.C. Bölge İdare Mahkemeleri Vergi Mahkemesi	93,2	85,0

TABLO I: Wikipedia Kelime Temsilleri ile Karar Tahmini Sonuçları

B. Hukuka Özel Kelime Temsilleri Kümesi

Bu bölümde bildiride kullanılan hukuka özel kullanım için üretilmiş olan kelime temsilleri kümesi tanıtılacaktır. Bu küme lawTurk2Vec olarak adlandırılmıştır. lawTurk2Vec bildiri yazarlarının yer aldığı araştırma grubu tarafından üretilmiştir. Bu kümenin vektörleri word2vec algoritmasıyla üretilmiştir. Algoritmada girdi olarak çeşitli Türk mahkemelerinden karar metinleri kullanılmıştır. Kullanılan karar metni sayıları Tablo II'de verilmiştir:

lawTurk2Vec 169.736 kelime tipinden oluşmaktadır. Kelime tiplerine sadece derlemde en az 5 kere geçen kelime tipleri dahil edilmiştir. lawTurk2Vec vektörleri 300 elemanlıdır.

Mahkeme Adı	Dava Sayısı
Yargıtay	332.662
Anayasa Mahkemesi	6.485
Bölge Adliye Mahkemeleri Hukuk Mahkemesi	47.796
Bölge Adliye Mahkemeleri Ceza Mahkemesi	9.241
Bölge İdare Mahkemeleri İdari Mahkeme	20.948
Bölge İdare Mahkemeleri Vergi Mahkemesi	8.870

TABLO II: Türk Hukuk Derlemi
IV. DENEYLER VE SONUÇLARI

Wikipedia kelime temsilleri kullanıldığında test kümesi ile elde edilen sonuçlar aşağıdaki tabloda verilmiştir. Tablo I ile III karşılaştırıldığında Yargıtay için bildiride elde edilen performansın literatürdeki çalışmaları geride bıraktığı görülmektedir.

Yöntem	Doğruluk (%)	Makro Yakalama (%)	Makro F1 (%)
GRU	96,8	90,1	86,4
GRU + Dikkat Katmanı	93,7	86,7	81,1
LSTM	95,6	88,4	85,7
LSTM + Dikkat Katmanı	96,4	89,5	87,9
BiLSTM	94,8	88,6	84,0
BiLSTM + Dikkat Katmanı	96,8	87,9	88,7

TABLO III: Wikipedia Kelime Temsilleri ile Karar Tahmini Sonuçları

lawTurk2Vec kelime temsilleri kullanıldığında test kümesi ile elde edilen sonuçlar aşağıdaki tabloda verilmiştir. Tablo III ile IV karşılaştırıldığında Wikipedia kelime temsillerinin lawTurk2Vec kelime temsillerine göre daha başarılı sonuçlar verdiği gözlemlenmiştir. Wikipedia kelime temsillerinin daha başarılı olmasının sebebinin Wikipedia derleminin çok daha büyük olması olduğu değerlendirilmektedir.

Yöntem	Doğruluk (%)	Makro Yakalama (%)	Makro F1 (%)
GRU	96,1	88,9	87,0
GRU + Dikkat Katmanı	95,7	87,3	85,6
LSTM	94,7	87,6	83,3
LSTM + Dikkat Katmanı	94,1	88,6	82,4
BiLSTM	92,5	86,6	79,0
BiLSTM + Dikkat Katmanı	95,9	80,0	83,8

TABLO IV: lawTurk2Vec Kelime Temsilleri ile Karar Tahmini Sonuçları

Her iki kelime kümesiyle yapılan deneyde de dikkat mekanizması bulunmayan GRU modeli en iyi sonuçları vermiştir. Deneylerde daha kompleks olan modellerin daha yüksek kapasiteye sahip olduğu için daha iyi sonuç vermesi bekleniyordu ancak en basit model olan GRU en iyi sonucu verdi. Bu durumun sebebinin modellerde herhangi bir regülarizasyon olmaması olduğu değerlendirilmektedir. Regülarizasyon uygulanarak bazı deneyler de yapıldı fakat Tablo III ve IV'te gösterilen sonuçlardan daha iyi sonuçlar elde edilemedi. Regülarizasyon için deney yapılabilecek hiperparametre uzayı çok geniş olduğu için denebilecek başka konfigürasyonlar da bulunmaktadır.

V. SONUÇ

Tablo I ile III karşılaştırıldığında Yargıtay için bildiride elde edilen performansın literatürdeki çalışmaları geride bıraktığı görülmektedir. Bu başarıda Yargıtayda çok sayıda dava bulunmasının önemli payı olduğu düşünülmektedir. Deneylerde daha basit modellerin daha başarılı olduğu gözlemlenmektedir.

Beklenti ise daha kompleks modellerin daha başarılı olmasıydı. Bu beklenti modellere regülarizasyon uygulanmasıyla karşılanabilir. Seyreltme (*İng.* dropout) ve L2 regülarizasyonu gibi regülarizasyon yöntemleri ile daha iyi skorlar da elde edilebilir.

lawTurk2Vec ile alınan performans skorları genel olarak Wikipedia skorlarına göre düşük kalmıştır. Bu yüzden lawTurk2Vec'in tek başına kullanıldığında hukuk uygulamalarında yetersiz kalabileceği gözlenmiştir. Bu sebeple lawTurk2Vec'in gelecek çalışmalarda Wikipedia kelime temsilleriyle beraber kullanılarak modellerde denemesi fikri ortaya çıkmıştır. Hukuki metinleri işlerken Wikipedia kelime temsillerini ve lawTurk2Vec temsillerini ayrı ayrı işleyen RNN türevi yapılar gelecek çalışmalarda denenerek daha iyi sonuçlar elde edilebileceği tahmin edilmektedir.

BİLGİLENDİRME

Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) (1001-120E346) fonuyla desteklenmiştir.

KAYNAKLAR

- [1] G. E. Hinton, J. L. McClelland, and D. E. Rumelhart, *Distributed Representations*. Cambridge, MA, USA: MIT Press, 1986, p. 77–109.
- [2] D. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, pp. 533–536, 1986.
- [3] J. L. Elman, “Finding structure in time,” *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.
- [4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2013.
- [5] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [6] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching Word Vectors with Subword Information,” *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.
- [8] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237.
- [9] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, “LEGAL-BERT: The Muppets straight out of Law School,” in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, 2020, pp. 2898–2904.
- [10] A. Köksal, “Turkish pre-trained Word2vec model,” <https://github.com/akoksal/Turkish-Word2Vec>, 2018.
- [11] E. Mumcuoğlu, C. E. Öztürk, H. M. Ozaktas, and A. Koç, “Natural language processing in law: Prediction of outcomes in the higher courts of Turkey,” *Information Processing & Management*, vol. 58, no. 5, p. 102684, 2021.
- [12] N. Aletras, D. Tsarapatsanis, D. Preoțiu-Pietro, and V. Lampos, “Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective,” *PeerJ Computer Science*, vol. 2, 2016.
- [13] A. Y. Ikram and L. Chakir, “Arabic text classification in the legal domain,” in *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*. IEEE, 2019, pp. 1–6.
- [14] D. M. Katz, M. J. Bommarito, and J. Blackman, “A general approach for predicting the behavior of the supreme court of the United States,” *PLoS one*, vol. 12, no. 4, p. e0174698, 2017.
- [15] K. Kowsrihawatt, P. Vateekul, and P. Boonkwan, “Predicting judicial decisions of criminal cases from Thai Supreme Court using bi-directional GRU with attention mechanism,” *2018 5th Asian Conference on Defense Technology (ACDT)*, pp. 50–55, 2018.
- [16] S. Long, C. Tu, Z. Liu, and M. Sun, “Automatic judgment prediction via legal reading comprehension,” in *China National Conference on Chinese Computational Linguistics*. Springer, 2019, pp. 558–572.
- [17] O.-M. Şulea, M. Zampieri, M. Vela, and J. van Genabith, “Predicting the law area and decisions of French Supreme Court cases,” in *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, Varna, Bulgaria, Sep. 2017*, pp. 716–722.
- [18] O. Şulea, M. Zampieri, S. Malmasi, M. Vela, L. P. Dinu, and J. van Genabith, “Exploring the use of text classification in the legal domain,” in *Proceedings of the Second Workshop on Automated Semantic Analysis of Information in Legal Texts co-located with the 16th International Conference on Artificial Intelligence and Law (ICAIL 2017)*, London, UK, June 16, 2017, ser. CEUR Workshop Proceedings, K. D. Ashley, K. Atkinson, L. K. Branting, E. Francesconi, M. Grabmair, M. Lauritsen, V. R. Walker, and A. Z. Wyner, Eds., vol. 2143. CEUR-WS.org, 2017.
- [19] M. B. Virtucio, J. Aborot, J. K. Abonita, R. Aviñante, R. J. Copino, M. Neverida, V. Osiana, E. Peramo, J. Syjuco, and G. B. Tan, “Predicting decisions of the Philippine supreme court using natural language processing and machine learning,” in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, 2018, pp. 130–135.
- [20] B. G. Buchanan and T. E. Headrick, “Some speculation about artificial intelligence and legal reasoning,” *Stanford Law Review*, vol. 23, pp. 40–62, 05 1970.
- [21] T. Bench-Capon, A. M. Araszkievicz, A. K. Ashley, K. Atkinson, F. Bex, F. Borges, D. Bourcier, P. Bourguin, J. G. Conrad, E. Francesconi, T. F. Gordon, G. Governatori, J. L. Leidner, D. D. Lewis, R. P. Loui, L. T. McCarty, H. Prakken, F. Schilder, E. Schweighofer, P. Thompson, A. Tyrrell, B. Verheij, D. N. Walton, and A. Z. Wyner, “A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law,” *Artificial Intelligence and Law*, vol. 20, pp. 215–319, 2012.
- [22] I. Chalkidis and D. Kampas, “Deep learning in law: early adaptation and legal word embeddings trained on large corpora,” *Artificial Intelligence and Law*, vol. 27, no. 2, pp. 171–198, 2019.
- [23] N. Sevim and A. Koç, “Investigation of Gender Bias in Turkish Word Embeddings,” in *29th Signal Processing and Communications Applications Conference, SIU 2021, Istanbul, Turkey, June 9-11, 2021*. IEEE, 2021, pp. 1–4.
- [24] N. Sevim, F. Şahinuç, and A. Koç, “Gender Bias in Legal Corpora and Debiasing It,” *Natural Language Engineering*, pp. 1–34, 2022.
- [25] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.
- [26] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [27] G. Tang, H. Guo, Z. Guo, and S. Xu, “Matching law cases and reference law provision with a neural attention model,” in *IBM China Research, Beijing*, 2016.
- [28] C. Çetindağ, B. Yazıcıoğlu, and A. Koç, “Named Entity Recognition in Turkish Legal Texts,” *Natural Language Engineering*, pp. 1–29, 2022.
- [29] E. Mumcuoğlu, C. E. Öztürk, H. M. Ozaktas, and A. Koç, “law-turk,” <https://github.com/koc-lab/law-turk>, 2021.
- [30] A. A. Akın and M. D. Akın, “Zemberek, an open source NLP framework for Turkic languages,” <https://github.com/ahmetaa/zemberek-nlp>, 2013.