



Yargıtay Kararlarının Suç Türlerine Göre Makine Öğrenmesi Yöntemleri İle Sınıflandırılması

Berker KILIÇ^{1*}, Prof. Dr. Yüksel ÖNER²

¹ Ondokuz Mayıs Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik, Samsun

² Ondokuz Mayıs Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik, Samsun

Özet

Mahkeme kararlarının herkese açık olması nedeni ile avukatların içtihat ve emsal karar araştırabilmesi için, pek çok internet sitesi ve ticari uygulama bulunmaktadır. Bunların birçoğu, anahtar kelime girişi ve anahtar kelimeyi içeren kararların listelenmesine dayanmaktadır. Bu durumda avukatın araştırmada kullanacağı anahtar kelimeler kendisi için uygun kararlara ulaşmasını sağlamakta ancak içtihat ve emsal kararların çok fazla olması nedeni ile pratikte zaman kazandırıcı bir yöntem olmamaktadır.

Avukatlar arasında içtihat araştırmadaki bir diğer yöntem ise birbirlerine danışmak ve birbirlerinin kullandığı içtihat ve emsal kararları kendilerinin de kullanması şeklindedir. Bir kısım avukat ise özellikle güncel içtihat kararlarını takip etmekte ve ileride işine yarayabileceğini düşündüğü güncel kararları kendince kategorik olarak kayıt altında tutmaktadır.

Bu iki yöntem de birinin güncelliği ve duruma tam olarak uygunluğu garanti edememesi değerinin ise yine oldukça zaman alıcı bir iş olması nedeni ile pratikte verimli yöntemler olmadıkları görülmektedir.

Avukatların içtihat ve emsal karar araştırabilmesi için, mahkeme kararlarının herkese açık olması nedeni ile pek çok internet sitesi ve ticari uygulama bulunmaktadır. Bunların bir çoğu, anahtar kelime girişi ve anahtar kelimeyi içeren kararların listelenmesine dayanmaktadır. İchtihat ve emsal kararlar üzerinde makine öğrenmesi yöntemlerinin kullanılarak bir içtihat/emsal karar arama motorunun alt yapısında kullanılabilir bir model geliştirilebileceği değerlendirilmiştir.

Makine öğrenmesi yöntemiyle yapılan metin sınıflandırma çalışmalarında, metinlerin, bilgisayarın anlayabileceği şekle yani sayısal bir hale dönüştürülmesi gerekmektedir. Çalışmada, Facebook tarafından geliştirilen FastText yöntemi kullanılmıştır. 2020 yılına ait Yargıtay Ceza Genel Kurulu tarafından karara bağlanan 531 adet karar metnine ulaşılmış ilgili oldukları suç türlerine göre sınıflandırılmış ve çapraz doğrulama için 3 eğitim ve test grubuna ayrılmıştır.

FastText danışanlı öğrenme yöntemi ile veri seti üzerinde tekrar sayısı (epoch), öğrenme oranı (lr: learning rate) ve kelime büyüklüğü (wordNgrams) parametrelerinin farklı değerleri için tüm kombinasyonlar ayrıca FastText otomatik optimizasyon seçenekleri ile denemeler yapılmış ve ortalama olarak %44,7 kesinlik değerine ulaşılmıştır.

Bir emsal karar arama sisteminde doğru ve yanlış etiketlenmiş kararlardan oluşacağı için, öğrenme oranı doğru etiketlenmiş sonuçların da artacağı, yanlış etiketlenmiş kararların ise azalacağı anlamına gelecektir. Ancak her halükarda, yalnızca anahtar kelimeye bağlı olarak edilen sonuçlarla kıyaslandığında, ilgililik yani istenen suç türüne ait kayıtların toplam

Makale Bilgisi

Başvuru:

18/10/2021

Kabul:

09/12/2021

sonular ierisindeki miktarı artacađı %44,7 oranı dşk bir oran olmasına rađmen uygulamada fayda sađlayabilecek bir deđer olarak dşnlmektedir.

Anahtar Kelimeler: *Hukuk, Hukukta Teknoloji, Emsal Karar, Makine đrenimi, FastText*

Classification Of Supreme Court Decisions According To The Types Of Crime By Machine Learning Methods

Abstract

Since court decisions are open to everyone, there are many websites and commercial applications for lawyers to search for case law and precedent.

Many of these are based on keyword entry and listing decisions involving the keyword.

In this case, the keywords that the lawyer will use in the research let him to reach the appropriate decisions for himself, but it is not a time-saving method in practice due to the high number of case-law and precedent decisions.

Another method of researching case law among lawyers is to consult each other and to use the case law and precedents used by each other.

On the other hand, some lawyers follow the current case-law decisions and categorically record the current decisions that they think may be useful in the future.

It is seen that these two methods are not efficient methods in practice, since one of them cannot guarantee the up-to-dateness and full compliance with the situation, and the other is a very time-consuming task.

In order for lawyers to search for legal precedent and case law, there are many websites and commercial applications because court decisions are open to everyone. Many of these are based on keyword entry and listing decisions involving the keyword. It has been evaluated that a usable model can be developed in the infrastructure of a legal precedent/case law decision search engine by using machine learning methods to case law and precedent decisions.

In the text classification studies made by machine learning method, the texts should be converted into a digital form that the computer can understand. FastText method developed by Facebook was used in the study. A total of 531 decisions concluded by the Supreme Court Penal General Assembly of the year 2020. were classified according to the types of crime they are related to and those crimes were divided between 3 training and test groups for cross-validation.

With the FastText supervised learning method, experiments were made with all combinations and FastText automatic optimization options for different values of repetitions (epoch), learning rate (lr) and word length (wordNgrams) parameters and an average of 44.7% precision was reached.

Since a case-law decision search system will consist of correct and incorrect labeled decisions, the learning rate will mean that correct labeled results will also increase, while incorrect labeled decisions will decrease. However, in any case, when compared to the results based on the keyword alone, the relevance, that is, the number of records belonging to the desired crime type, in the total results will increase, although the rate of 44.7% is low, it is considered to be a value that can be beneficial in practice.

Keywords: *Law, Technoloht in Law, Precedent Decision, Machine Learning, FastText*

* İletişim e-posta: berker.kilic@gmail.com

Giriş

Türk Yargı Sisteminde yargılama ilk derece mahkemeleri, bölge mahkemeleri (istinaf), Yargıtay (temyiz), Anayasa Mahkemesi (bireysel başvuru) ve bu noktada yargı yollarının tüketilmiş olması ile Avrupa İnsan Hakları Mahkemesi sıralamasında işlemektedir.

Yargılama sürecinin her aşamasında incelenen ve yapılan işlemler yazılı hale getirilmekte ve kararlar Türk Milleti adına herkesin erişimine açık bir şekilde verilmektedir.

Yargının kademeli mahkemeler hiyerarşisine sahip olması, üst mahkeme kararlarının alt mahkemeler tarafından dikkate alınmasını zorunlu hale getirmiştir.

Yargı sistemi içerisinde taraf vekili olan Avukatlar açısından ise, alt mahkemelerin, üst mahkemelerin kararlarını dikkate alma zorunluluğu müvekkillerinin haklarını korumada/savunmada bir araç olarak kullanılmaktadır.

Kısaca, avukat, müvekkilinin davasına benzer bir konuda daha önce gerçekleştirilmiş ve sonuçlandırılmış bir başka davayı emsal göstererek müvekkili lehine karar alınmasını sağlamaya çalışır. Buna da içtihat hukuku denilmektedir.

Avukatlar müvekkileri lehine bir içtihatı (emsal kararı) kullanabilmek için, üst mahkeme kararları ve emsal kararlar arasında müvekkilinin davasına en uygun geçmiş kararları bulmak, bunları incelemek, müvekkilinin durumu açısından değerlendirmek ve kendi dilekçeleri ve savunma metinleri içerisinde bunu yerleştirmek zorundadır. Müvekkilin davasının karmaşıklığı ölçüsünde, avukatın da kullanması gereken içtihat ve emsal kararların da sayısı artacaktır.

Bu süreç avukat açısından mesleki deneyimine bağlı olarak oldukça uzun mesailer harcamasına neden olabilecek niteliktedir.

Mahkeme kararlarının herkese açık olması nedeni ile avukatların içtihat ve emsal karar araştırabilmesi için, pek çok internet sitesi ve ticari uygulama bulunmaktadır. Bunların birçoğu, anahtar kelime girişi ve anahtar kelimeyi içeren kararların listelenmesine dayanmaktadır. Bu durumda avukatın araştırmada kullanacağı anahtar kelimeler kendisi için uygun kararlara ulaşmasını sağlamakta ancak içtihat ve emsal kararların çok fazla olması nedeni ile pratikte zaman kazandırıcı bir yöntem olmamaktadır.

Avukatlar arasında içtihat araştırmadaki bir diğer yöntem ise birbirlerine danışmak ve birbirlerinin kullandığı içtihat ve emsal kararları kendilerinin de kullanması şeklindedir. Bir kısım avukat ise özellikle güncel içtihat kararlarını takip etmekte ve ileride işine yarayabileceğini düşündüğü güncel kararları kendince kategorik olarak kayıt altında tutmaktadır.

Bu iki yöntem de birinin güncelliği ve duruma tam olarak uygunluğu garanti edememesi diğerinin ise yine oldukça zaman alıcı bir iş olması nedeni ile pratikte verimli yöntemler olmadıkları görülmektedir.

Çalışmanın Önemi ve Kapsamı

Bu noktada avukatların müvekkillerinin haklarını savunmada içtihat ve emsal kararlara mümkün olan en doğru ve en yüksek hızda ulaşma ihtiyaçları bu çalışmada çözmeyi amaçladığımız ana sorun olarak görülmektedir. Bu sorunun çözülebilmesi için içtihat ve emsal kararlar üzerinde makine öğrenmesi yöntemlerinin kullanılarak bir içtihat/emsal karar arama motorunun alt yapısında kullanılabilir bir model geliştirilebileceği değerlendirilmiştir.

Çalışma kapsamında avukatların bu ihtiyacının eksiksiz bir şekilde çözülmesi amaçlanmamaktadır. Amaç, bu öngörü için mantıksal bir temel oluşturulması, kısıtlı veri ile testlerin yapılması ve sonuçların değerlendirilmesi olacaktır.

Hukuk Alanında Yapay Zeka

Hukuk' ta Yapay Zeka konusu ilk defa 2018 yılında Kodex Bilişim İsimli ODTÜ Teknokent' de faaliyet gösteren bir şirketin ARYA isimli uygulaması ile somut bir şekilde üzerinde konuşulan bir konu haline gelmiştir.

ARYA isimli bu yazılıma ait teknik bilgiye açık kaynaklar aracılığı ile ulaşılammış olsa da, Kodex Bilişimin yerleşik bulunduğu ODTÜ Teknokent tarafından yayınlanan bir haberde ARYA ile ilgili olarak "Firma, bu sistemi geliştirirken Türkçe karar metinlerini yapay zekanın bir alt kategorisi olan "doğal dil işleme" teknolojisiyle sayısal verilere dönüştürüyor ve çeşitli istatistiksel yöntemlerle de sonuçları tahmin ediyor. Adalet Bakanlığının elindeki 1,5 milyon veriden çok azını analiz edebilen yazılım, şu anda sadece boşanma davası üzerinde çalışıyor. Yazılım, hakimlerin işini ellerinden almadan onlara asistanlık yapabilecek. Dava sonuçlarını tahmin eden yazılımlarının dışında firma, hukukçuların kullanacağı yapay zeka

teknolojisinin olduğu yerli arama motoru ve problemlerini tartışabildiği yerli sosyal medya platformu projeleri üzerine de yoğunlaşıyor.” ifadelerinin yer aldığı görülmüştür [1].

Bu ifadelerden geliştirilen yazılımın Türkçe karar metinleri üzerinde çalıştığı, tüm kararları kapsamadığı, yalnızca boşanma davalarına odaklanılmış olduğu anlaşılmaktadır. Ayrıca aynı firmanın hukukçuların kullanımı için yapay zeka teknolojisi kullanan bir arama motoru üzerine de çalıştığı vurgulanmaktadır.

Bu proje hakkında, Yapay Zeka Çağında Hukuk 2019 başlıklı, İstanbul, Ankara ve İzmir Baroları tarafından düzenlenen Çalıştay sonuç raporunda, ARYA isimli uygulamanın Yargıtay kararlarını Onama, Bozma ve Düzelterek Onama olmak üzere sınıflandırdığı, bunu %90 oranında doğrulukla yaptığı ve projenin TÜBİTAK Bireysel Genç Girişimcileri Destekleme Programında desteklendiği ifade edilmiştir.

Mayıs 2021 tarihi itibarıyla ilgili firmaya ait web sitesinin çalışmadığı görülmüş, ARYA isimli yazılıma ait güncel bilgiye ulaşılamamıştır.

Arama alt yapısında doğal dil işlemeyi kullanan bir diğer uygulama ise HukukWork isimli projedir.

HukukWork ile ilgili yapılan araştırmada kendi web sitesinde yer alan “Aradığımız kavramların, eş anlamlısını, kök halini, sert sessizlerin yumuşamış ve yumuşak sessizlerin sertleşmiş halini (Türkçe köke indirgeme, gövdeleme vb.) özellikler sayesinde doğal dil araması ile başarılı, kolay ve geniş kapsamlı arama yapabilirsiniz.” ifadeleri görülmüştür. Bu ifadeler ile verilen uygulama örneklerinde ise mücbir kelimesi ile zorlayıcı kelimesinin eş anlamlılıklarının göz önünde bulundurulduğu veya ipoteğin kelimesi ile ipotek kelimesinin aynı kökten gelmelerinin göz önünde bulundurulduğu ve bunun arama sonuçlarına yansıtıldığı belirlenmiştir [2].

HukukWork projesinin genel itibarıyla yine bir anahtar kelime tabanlı arama alt yapısına sahip olmakla birlikte, doğal dil işlemeyi bir makine öğrenmesi veya yapay zeka alt yapısı olarak değil aranan kelime ve ifadelerin işlenerek arama sonuçlarının daha kapsamlı olması üzerine kurgulandığı tahmin edilmektedir.

Ayrıca, projenin kararlar ile diğer kararlar ve mevzuat arasındaki ilişkileri de kapsıyor olması diğer pek çok karar arama sistemine kıyasla daha gelişmiş bir sistem olduğunu gösterdiği, sosyal

medya hesaplarındaki takipçi sayıları itibarıyla oldukça yeni olan bu projenin ilerleyen zamanlarda bilinirliğinin artacağı değerlendirilmiştir.

Adalet Hanım isimli bir diğer projede ise, doğal dil işleme alt yapısı kullanılan, bir arama motoru formatında çalışan bir Internet sitesi ile karşılaşmıştır. Bu çalışma hakkında Yapay Zeka Çağında Hukuk 2019 Raporunda [3] her ne kadar yapay zeka tabanlı hukuk asistanı tanımı yapılmış olsa da, web sitesinde deneme amaçlı yapılan aramalarda [4], HukukWork çalışmasına olduğu gibi doğal dil işleme ile aranan ifadelerin kök temelli olarak işlendiği ve arama sonuçlarının kelimelerde yer alan alternatif eklerle birlikte listelendiği görülmüştür.

Hukuk alanında yapay zeka konusunda ulaşılabilen son gelişme ise TÜBİTAK desteği ile Ocak 2021 tarihinde başlatılmış olan “Hukuk Metinlerinde Yapay Zekâ Uygulamaları için Doğal Dil İşleme ve Makine Öğrenmesi Teknikleri” isimli proje olmuştur. Projenin oldukça yeni bir tarihe sahip olması ve bitiş tarihinin 15/07/2023 olması [5] nedeniyledir ki, bu proje hakkında başkaca bir bilgiye ulaşılamamıştır.

Bilkent Üniversitesine ait UMRAM internet sitesinde yer alan 1 Mart 2021 tarihli bilgilere göre, Bilkent Öğretim Üyesi Doç.Dr. Aykut KOÇ tarafından yürütülen projenin ilerleyen zamanlarda hukukta yapay zeka kullanımı konusunda somut ürünlerin ortaya koyulabileceği değerlendirilmiştir.

Hukuk Alanında Emsal Karar Arama

Hali hazırda yayında olan karar arama siteleri ve yazılımları araştırıldığında başta Yargıtay’ın kendisine ait bir Internet sitesi olan <https://karararama.yargitay.gov.tr/> Türkiye Barolar Birliği’ne ait yalnızca avukatlara özel ve ücretsiz olan <https://www.hukukturk.com/yargitay-kararlari> internet sitesi,

Resmi kaynaklı karar arama yapılarına ek olarak Lexperia (<https://www.lexpera.com.tr/>), Kazancı (<https://www.kazanci.com.tr/>), Full&Egal (<https://www.fullegal.com/>) gibi ticari girişimler de bulunmaktadır. Oldukça uzun bir ticari geçmişe sahip firmalar tarafından kurulan alt yapıların da anahtar kelime eşleşmesine dayalı olarak çalıştıkları bilinmektedir.

Uygun Yöntemin Belirlenmesi

Literatürde metin işlemeye dayalı makine öğrenmesi konulu makaleler incelendiğinde,

çalışmaların çoğunlukla sınıflandırma modellerinin kıyaslanmasına [6, 7, 8] ve duygusal durum belirlemeye yönelik çalışmalar [9, 10] oldukları, çalışmaların bir kısmında yazarların kendi veri setleri üzerine çalıştıkları görülmüştür [9, 10]. Çalışmaların duygu durumunun olumlu-olumsuz veya olumlu-olumsuz-nötr olarak sınıflandırılmasından farklı olarak Çelik ve Koç ile Kınık ve Güran çalışmalarında haber metinlerini içerdikleri haber türlerine sınıflandırma çalışması yaparak diğerlerinden farklılaştığı görülmüştür.

Ahmetoğlu ve Daş çalışmalarında ziyaretçi otel yorumlarının duygu analizi ile sınıflandırma yaptıkları, sınıflandırmalarında, kendi hazırladıkları veri setleri ile çalıştıkları ve Word2Vec algoritmasını kullandıkları görülmüştür. Yazarların F1 skoru ile çalışmalarının başarımlarını ölçtükleri t-SEN Python kütüphanesi ile de modellerinin görselleştirmesini yaptıkları görülmüştür.

Kınık ve Güran çalışmalarında TF-IDF ve Word2Vec modellerinin başarımlarının ardışık kelime grubu tespitleri ile artırılması üzerine çalıştıkları görülmüştür.

Aksu ve Karaman çalışmalarında Word2Vec modelinin bir uzantısı olan FastText yöntemi ile Kelime Çantası Modeli (BOW) yöntemlerinin duygu durum analizinde başarımlarının karşılaştırıldığı görülmüştür.

Onan ise çalışmasında Word2Vec, FastText, GloVe LDA2vec kelime gömme yöntemleri ile oluşturduğu kelime vektörleri üzerine K-En Yakın Komşu Algoritması, Destek Vektör Makinaları, Lojistik Regresyon ve Naive Bayes algoritmaları ile başarımlarını testleri yapmıştır.

Çelik ve Koç çalışmalarında TF-IDF, Word2Vec ve FastText modelleri ile vektörleştirdikleri veriler üzerinde Destek Vektör Makinaları, NAive Bayes, Lojistik Regresyon, Random Forest ve Yapay Sinir Ağları yöntemleri ile sınıflandırma yaptıkları görülmüştür.

Hukuk alanı ile ilgili tek akademik makale olarak karşılaşılan, Turan ve ark., yaptıkları çalışmada [11], ulusal ve uluslararası literatürde hukuk alanı ile ilgili akademik çalışmaları araştırdıkları literatür taramasına dayalı bir çalışma yaptıkları görülmüştür. Çalışmalarında, Word2Vec, Skip-Gram, n-Gram, CBOW modellerinin güncel literatürde kullanılan modeller olarak tespit ettikleri görülmüştür.

Kelime Gömme Teknikleri

Makine öğrenmesi yöntemiyle yapılan metin sınıflandırma çalışmalarında, metinlerin, bilgisayarın anlayabileceği şekle yani sayısal bir hale dönüştürülmesi gerekmektedir. Bu dönüştürme işleminde farklı yöntemler kullanılmakla birlikte literatürde model, teknik, yöntem ve benzeri şekillerde ifade edildikleri görülmüştür. Genel olarak yöntem kelimesi ile ifade edebileceğimiz metinlerin sayısallaştırılmasında kullanılan algoritmalar kısaca şunlardır:

TF-IDF

TF-IDF, bir belgedeki her bir kelimenin değerlerini, belirli bir belgedeki kelimenin sıklığı ile kelimenin görüldüğü belgelerin yüzdesinin tersiyle hesaplar. Belirli bir belgede kelimelerin göreceli sıklığını, bu kelimenin tüm veri seti üzerindeki tersine oranına göre belirleyerek çalışır. Bu hesaplama, belirli bir kelimenin belirli bir belge ile ne kadar alakalı olduğunu belirler [8].

Word2Vec

Bu yöntem kelime ile kelimenin belirli pencere boyutundaki komşuları arasında yakınlık ilişkisi kurar ve anlamca yakın kelimeleri vektör uzayında birbirine yakın olacak şekilde konumlandırır. Anlam ilişkisi kurabilmek için iki farklı öğrenme mimarisi kullanır. Bunlardan ilki Continuous Bag of Words (CBoW) mimarisidir. Bu yöntemde pencere merkezindeki kelime, kelimenin pencere boyutu kadar yakınındaki komşularına bakılarak tahmin edilmeye çalışılır. Diğer yöntem ise Skip-Gram mimarisidir. Bu yöntem de CBoW ile benzer çalışır. Ancak Skip-Gram, CBoW' un aksine pencere merkezinde konumlandırılan kelimedenden, kelimenin komşularını tahmin eder. Bu yöntemin avantajı farklı anlamlara gelebilecek kelimelerin birden fazla anlamını yakalayabilmesidir [8].

Word2vec modeli, girdi katmanı, çıktı katmanı ve gizli katmandan oluşan yapay sinir ağı tabanlı bir kelime gömme yöntemidir [7].

GloVe

Bu model, word2vec modelinin yerel bağlama dayalı öğrenmesini, küresel matris çarpanlarına ayırma ile birleştirir. Modelde, hata fonksiyonun hesaplanmasında kelimelerin olasılık oranları da dikkate alınır [7].

LDA2vec

Bu yöntemde, gizli belge düzeyi karışım vektörlerinden yoğun sözcük vektörleri, Dirichlet

dağılımına bağlı olarak birlikte elde edilmektedir. Model, metin koleksiyonlarının konularını ve konuya göre ayarlanmış sözcük vektörlerini tanımlamayı sağlar. Bu bağlamda, model her bir kelimeyi bir konuya bağlayarak konuyla zenginleştirilmiş bir kelime gömme şeması sunar. Modelde, skip gram negatif örnekleme objektif fonksiyon olarak kullanılmaktadır [7].

FastText

FastText ise 2016 yılında Facebook tarafından geliştirilen Word2Vec tabanlı bir modeldir. Bu yöntemin Word2Vec' ten farkı, kelimelerin n-gramlara ayrılmasıdır. Böylece Word2Vec ile yakalanamayan anlam yakınlığı bu yöntemle yakalanabilmektedir [8].

FastText, Word2Vec yönteminin skip-gram alt yöntemine dayanmaktadır. Word2Vec yönteminde her bir kelime en küçük birim olarak kullanılırken FastText yönteminde kelimeler n-gramlarına ayrılarak ele alınmaktadır. N-gramlarına ayrılan kelimeler, n-gramların toplamı olacak şekilde vektörel olarak temsil edilmektedir. N-gramları ele almayan diğer yöntemler, her bir kelimeyi bir vektöre dönüştürdükleri için kelimenin morfolojisini görmezden gelmektedirler. Bu durum özellikle büyük kelime dağarcığı olan ve çok nadir kelimelere sahip olan diller için bir sınırlandırma getirmektedir. FastText yöntemi, n-gram yöntemi sayesinde bu sınırlamanın üstesinden gelmektedir [10].

FastText' in n-gram modeli ile çalışması nedeni ile sağlanan bir diğer avantaj, veri setinin noktalama işaretlerinin temizlenmesi, durak kelimelerin temizlenmesi, kök belirleme gibi ön hazırlık işlemlerinin elde edilecek sonuçlar üzerindeki etkisinin göz ardı edilebilecek kadar azaltmasıdır.

Bu nedenledir ki FastText modeli, morfolojik olarak zengin diller ve nadir görülen kelimeler için daha verimli bir kelime gömme şeması sağlar [7].

1	1,"Sanık ... hakkında hizmet nedeniyle güveni kötüye kullanma suçundan açılan kamu davasının
2	2,"Sanık ... hakkında nitelikli dolandırıcılık suçundan açılan kamu davasında yapılan yargı
3	2,"Sanık ... hakkında nitelikli dolandırıcılık ve resmi belgede sahtecilik suçlarından açıl
4	3,"Silahlı terör örgütüne üye olma suçundan sanık ... hakkında ilk derece mahkemesi sıfatı
5	4,"Nitelikli yağma suçundan sanıklar ... ve ...'nın TCK'nın 149/1-c, 62 ve 53. maddeleri u
6	5,"Sanıklar ... ve ... hakkında birden fazla kişiyle birlikte tehdit ve nitelikli hırsızlı
7	6,"Sanık ...'in, mağdur ... adına sahte nüfus cüzdanı düzenleme ve kullanma eylemine ilişki
8	7,"Kasten yaralama suçundan açılan kamu davasında yapılan yargılama sonucu değişen suç vas
9	8,"Sanık ...'in, görevi yaptırmamak için direnme suçundan TCK'nın 265/1, 62, 53 ve 58. mad

Şekil 1. Kararlar listesi

Karar metni içerisinde yer alan suç bilgisi sınıflandırmada esas alınacak bir listeye dönüştürülmüştür. Toplam 78 farklı suç türü olduğu belirlenmiştir. Kararlar suç türleri ile

FastText modeli, ön işleme gereksinimlerinin az olması ve başarı oranının kelime vektörü kullanan modellere kıyasla Çelik ve Koç ile Aksu ve Karaman çalışmalarına göre daha yüksek olduğu görülmüştür. Ancak, yapay zeka mimarileri ile birlikte değerlendirildiğinde FastText'in en yüksek başarıma sahip olmadığı da Onan' ın çalışmasında görülen bir sonuç olarak dikkat çekmiştir.

Yargıtay kararları içerisinde, benzer suçlar için benzer ifadelerin sıklıkla kullanılması muhtemel olması nedeni ile n-gram modeli ile FastText modelinin makul bir başarı oranı elde etmeyi sağlayacağı düşünülerek proje çalışmasında FastText modeli kelimelerin n-gramlarına ayrıştırılması ile uygulanmasına karar verilmiştir. Modelin başarı oranının artırılmasında, n-gram büyüklüğü, öğrenme hızı ve öğrenme tekrarı sayısı parametrelerinin optimizasyonundan yararlanılacaktır.

Materyal ve Metod

Verilerin Toplanması

Çalışmanın Yargıtay kararları üzerinde yapılacak olması ve hazır bir veri setinin olmaması nedeni ile kullanılacak veri seti Yargıtay Karar Arama Sitesi üzerinden edinilmiştir.

<https://karararama.yargitay.gov.tr/> adresinden erişilebilir karara arama sitesinde, 2020 yılına ait Yargıtay Ceza Genel Kurulu tarafından karar bağlanan 531 adet karar metnine ulaşılmıştır. Arama ifadesi olarak (.) ifadesi kullanılarak tüm kararlara ulaşılmıştır.

Ulaşılan karar metinleri, her bir karar tek bir satırda olmak üzere, .CSV formatında tek bir dosya içerisinde Şekil 1' de görüldüğü gibi satırlar halinde düzenlenmiştir.

etiketlenmiştir. FastText ile işlenebilmesi için her bir kaydın başlangıcına (`__label__`) ifadesi eklenmiştir. Şekil 2' de kararların suç türleri ile etiketlenmiş hali görüntülenmektedir.

1	label_görevi_kötüye_kullanma , sanık ... hakkında hizmet nedeniyle güveni kötüye kullanma
2	label_dolandırıcılık , sanık ... hakkında nitelikli dolandırıcılık suçundan açılan kamu d
3	label_dolandırıcılık , sanık ... hakkında nitelikli dolandırıcılık ve resmi belgede sahte
4	label_terör_örgütüne_üye_olma , silahlı terör örgütüne üye olma suçundan sanık ... hakkın
5	label_yağma , nitelikli yağma suçundan sanıklar ... ve ...'nın tck'nın 149/1-c, 62 ve 53.
6	label_hırsızlık , sanıklar ... ve ... hakkında birden fazla kişiyle birlikte tehdit ve ni
7	label_sahte_nüfus_cüzdanı_düzenleme , sanık ...'in, mağdur ... adına sahte nüfus cüzdanı
8	label_yaralama , kasten yaralama suçundan açılan kamu davasında yapılan yargılama sonucu
9	label_görevi_yaptırmamak_için_direnme , sanık ...'in, görevi yaptırmamak için direnme suç

Şekil 2. Suç türüne göre etiketlenmiş ve FastText ile işlenmeye uygun hale getirilmiş kararlar listesi

Kararlarda ön işleme amaçlı olarak (... , / , - , vb) bir kısım noktalama işaretleri temizlenmiştir.

FastText Metin Sınıflandırma Modeli

FastText, Word2Vec modelinin n-gram yapısını kullanan bir metin sınıflandırma modelidir. Hem denetimli hem denetimsiz öğrenme amaçlı kullanılabilir. FastText isimli bir Python kütüphanesine sahiptir ve veri setinin uygun veri formatında kullanılması gerekmektedir.

FastText'in Yüklenmesi

Python 2.6 ve üstü bir Python sürümü üzerine yüklenmesi gerekmektedir. Mevcut FastText sürümü 0.9.2 olup, pybind11, numpy ve setuptools kütüphaneleri ile birlikte çalışmaktadır. Ayrıca, C++

11 desteğine sahip bir C++ derleyicisinin de kullanılacak bilgisayarda yüklü olması gerekmektedir. C++ derleyici yüklemesi Visual Studio Installer uygulaması aracılığı ile FastText kurulumu öncesinde yüklenebilmektedir [12]

FastText İçin Veri Setinin Hazırlanması

FastText modeli, eğitim ve test verileri için kayıtların "_label_" etiketi ile başlayan satırlar halinde düzenlenmiş olmasını istemektedir. "_label_" etiketi devamında kaydın ait olduğu sınıf (tek kelime veya alt çizgi (.) vb karakterler ile birleştirilmiş ifade), ayrıca boşluk veya virgül karakteri ve devamında kayda ait metinden oluşmaktadır. Suç sınıfları (türleri) Şekil 3' de görüntülenmektedir.

1	görevi_kötüye_kullanma	52	konut_dokunulmazlığı_ihlal
2	dolandırıcılık	53	hakkı_olmayan_yere_tecavüz
3	terör_örgütüne_üye_olma	54	bozuk_gıda_ticareti
4	yağma	55	müstehcenlik
5	hırsızlık	56	avukatlık_kanununa_muhalefet
6	sahte_nüfus_cüzdanı_düzenleme	57	orman_yangınına_sebep_olma
7	yaralama	58	zimmet
8	görevi_yaptırmamak_için_direnme	59	suç_üstlenme
9	cinsel_saldırı	60	özel_belgede_sahtecilik
10	resmi_belgede_sahtecilik	61	haksız_mal edinme
11	mala_zarar_verme	62	karşılıksız_yararlanma
12	2863_sayıllı_kanuna_muhalefet	63	göçmen_kaçakçılığı
13	alacaklıyı_zora_sokma	64	veri_usul_kanununa_muhalefet
14	orman_alanını_işgal	65	mühür_bozma
15	anayasayı_ihlal	66	edimin_ifasına_fesat_karıştıрма
16	silahlı_terör_örgütüne_üye_olma	67	kaçma_imkanı_sağlama
17	öldürme	68	iftira
18	marka_hakkına_tecavüz	69	5510_sayıllı_kanuna_muhalefet
19	5846_sayıllı_kanuna_muhalefet	70	4708_sayıllı_kanuna_muhalefet
20	bankacılık_zimmeti	71	sahte_fatura_düzenleme
21	5607_sayıllı_kanuna_muhalefet	72	eziyet
22	imar_kirliliğine_neden_olma	73	haberleşmenin_gizliliğini_ihlal
23	çocuğun_cinsel_istismarı	74	ihracat_kaçakçılığı
24	uyuşturucu_madde_ticareti	75	suç_uydurma
25	banka_veya_kredi_kartlarının_kötüye_kullanı	76	tefecilik
26	trafik_güvenliğini_tehlikeye_sokma	77	genel_güvenliği_tehlikeye_sokma
27	belgelere_gizleme	78	yalan_tanıklık
28	askeri_ceza_kanununa_muhalefet	79	muhafaza_görevini_kötüye_kullanma

Şekil 3. Suç Sınıfları (Türleri)

Veri seti oluşturulurken edinilen 531 adet Yargıtay kararı, 1/3 (177/354 karar) oranlarında eğitim ve test gruplarına ayrıştırılmış, çapraz doğrulama amaçlı olarak da her bir grupta test verileri eğitim verileri ile değiştirilerek 3 adet test grubu oluşturulmuştur.

FastText Test Prosedürü

Modelin eğitilmesinde, danışanlı öğrenme modeli kullanılmıştır. FastText, herhangi bir optimizasyon parametresi olmadan da çalışabilmektedir. Çalışmada parametresiz ve sonrasında tekrar sayısı (epoch), öğrenme oranı (lr:learning rate) ve kelime

büyükülüğü (wordNgram) parametreleri için 3' er değer her bir parametre için sırasıyla; tekrar sayısı için (5;25;50), öğrenme oranı için (0.1;0.5;1.0) ve kelime büyükülüğü için (2;3;5) şeklinde alınmıştır.

Parametre değerlerine ek olarak, otomatik optimizasyon seçeneği de uygulanarak tüm elde edilen sonuçlar arasında elde edilen en iyi değerler belirlenmiştir.

Fasttext, herhangi bir karar için birinci, ikinci, üçüncü tahminleri de belirleyebilmeyi sağlamaktadır. Yani tahmin oranlarına bağlı olarak bir kararın hangi suç türüne ait olabileceğine dair birden fazla tahminde bulunabilmektedir.

Elde edilen en iyi değerler için, modelin ikinci, üçüncü, dördüncü ve beşinci tahminleri de hesaplanmıştır.

Son olarak modelin farklı tahmin sayıları için kesinlik ve duyarlılık değerleri kullanılarak F değerleri ve bu değerlerin ortalamaları hesaplanmıştır.

FastText Parametresiz Kullanım

Danışanlı öğrenme yönteminde, FastText, tekrar sayısı (epoch), öğrenme oranı (lr:learning rate), kelime büyükülüğü (wordNgram) temel parametreleri ile kullanılmaktadır. Ancak, herhangi bir parametre olmaksızın da varsayılan değerler ile kullanılması mümkündür.

FastText Parametreleri

FastText Tekrar Sayısı (epoch) Parametresi

Bu parametre, veri setinin kaç tekrar ile eğitileceğini belirlemektedir. Veri setinin tekrarlı olarak eğitilmesi, öğrenme oranını artıran bir parametredir. Tekrar sayısı parametresi 5-50 aralığında değer alabilmektedir.

FastText Öğrenme Oranı (lr) Parametresi

Bu parametre, veri setinin öğrenme oranının ne olacağını yani, her bir örnek işlendikten sonra ne kadar değişiklik yapılacağını belirlemektedir. Öğrenme oranı parametresi Öğrenme oranı parametresi 0.1-1.0 aralığında değer alabilmektedir.

FastText Kelime Büyükülüğü (wordNgrams) Parametresi

Bu parametre, veri seti içerisinde yazılan kelime büyükülüğünden bağımsız olarak işlenecek kelime büyükülüğünü belirtmektedir. Kelime büyükülüğü parametresi 1-5 aralığında değer alabilmektedir.

Bulgular ve Tartışma

Veri Setlerinin Eğitilmesi ve Test Edilmesi

Yargıtay Ceza Genel Kurulu tarafından 2020 yılına ait 531 adet karardan oluşan veri seti hazırlanmıştır. Veri seti 1/3 oranında eğitim ve test verisi olarak ayrıştırılmış, çapraz doğrulama amaçlı olarak test ve eğitim verileri karşılıklı değiştirilerek 3 adet test grubu elde edilmiştir.

Veri setinden elde edilen test gruplarının her biri danışanlı öğrenme yöntemi kullanılarak, parametresiz kullanıma ek olarak tekrar sayısı (epoch), öğrenme oranı (lr:learning rate) ve kelime büyükülüğü (wordNgrams) parametrelerinin 3 farklı değeri ve tüm kombinasyonlar için denenmiştir.

Parametrelili denemelerde tekrar sayısı parametresi için 5, 25 ve 50 değerleri, öğrenme oranı parametresi için 0.1, 0.5 ve 1.0 değerleri ve kelime büyükülüğü parametresi için 2, 3 ve 5 değerleri kullanılmıştır.

Karareğitim1-Karartest1 Grubu Test Sonucu

Test grubu için yapılan denemeler sonrasında en iyi sonucun epoch=50 ve lr=1.0 parametreleri ile elde edildiği, kesinlik değerinin %36,9, duyarlılık değerinin ise %36,9 olduğu görülmüştür.

Karareğitim2-Karartest2 Grubu Test Sonucu

Test grubu için yapılan denemeler sonrasında en iyi sonucun epoch=50 ve lr=1.0 olduğu değerlerde ortaya çıktığı, kesinlik değerinin %41,7 duyarlılık değerinin ise %41,7 olduğu görülmüştür.

Karareğitim3-Karartest3 Grubu Test Sonucu

Test grubu için yapılan denemeler sonrasında en iyi sonucun epoch=50 ve lr=1.0 olduğu değerlerde ortaya çıktığı, kesinlik değerinin %55,6 duyarlılık değerinin ise %55,6 olduğu görülmüştür.

Test grupları arasında elde edilen kesinlik ve duyarlılık değerleri itibariyle büyük bir fark olmadığı değerlendirilmiştir.

Otomatik Optimizasyon Testleri

Otomatik optimizasyon seçeneğinde, en iyi öğrenme oranına ulaşmak için eğitim ve test bölümleri üzerinde, FastText, verilen süre içerisinde kendi denemelerini yapmakta ve ulaştığı en iyi öğrenme oranı ile modeli eğitmektedir.

Bu çalışmada ele alınan her bir test grubu üzerinde, 5 dakikalık optimizasyon süresi için ulaşılan en iyi öğrenme oranları Tablo 1' de sunulmuştur. Farklı deneme sayıları ile elde edilen öğrenme oranlarının,

parametre kombinasyonları ile elde edilen öğrenme oranlarından daha düşük olduğu görülmektedir.

Otomatik optimizasyon işlemlerinde tekrar sayısı, öğrenme oranı ve kelime büyüklüğü parametreleri için herhangi bir değer belirlenmemiştir.

Tablo 1. FastText otomatik optimizasyon test sonuçları

Test Veri Seti	Otomatik Optimizasyon Deneme Sayısı	Kesinlik (precision)	Duyarlılık (recall)
Kararegitim1-Karartest1	12 deneme	0.119	0.119
Kararegitim2-Karartest2	20 deneme	0.115	0.115
Kararegitim3-Karartest3	10 deneme	0.154	0.154

Sonraki Tahminlerin İncelenmesi

Her 3 test grubu için elde edilen en iyi öğrenme oranlarına ait parametreler ile ikinci, üçüncü, dördüncü ve beşinci suç türü sınıflandırma tahminler hesaplanmıştır.

Daha sonra en iyi suç türü tahmin sayılarını belirleyebilmek için F1-Skor değerleri aşağıdaki Denklem (1) ile hesaplanmıştır.

$$\frac{2 \times \text{kesinlik} \times \text{duyarlılık}}{\text{kesinlik} + \text{duyarlılık}} \quad (1)$$

Tablo 2' den görüleceği üzere, Kararegitim1-Karartest1 test grubu için en yüksek F1 skorlarına ilk iki suç türü tahmininde ulaşıldığı görülmüştür.

Tablo 2. Kararegitim1-Karartest1 Test Grubu F1 Skorları

test k	Tekrar Sayısı (epoch)	Öğrenme Oranı (lr: learnin g rate)	Kelime Büyüklüğü (wordN grams)	precisi on (Kesinlik)	recall (Duyarlılık)	F1-Skoru (F1-Score)
1	50	1.0	-	0.369	0.369	0.369
2	50	1.0	-	0.295	0.589	0.393
3	50	1.0	-	0.215	0.643	0.322
4	50	1.0	-	0.177	0.708	0.283
5	50	1.0	-	0.151	0.756	0.252

Tablo 3' den görüleceği üzere, Kararegitim2-Karartest2 test grubu için en yüksek F1 skorlarına ilk iki suç türünün tahmininde ile ulaşıldığı görülmektedir.

Tablo 3. Kararegitim2-Karartest2 Test Grubu F1 Skorları

test k	Tekrar Sayısı (epoch)	Öğrenme Oranı (lr: learnin g rate)	Kelime Büyüklüğü (wordN grams)	precisi on (Kesinlik)	recall (Duyarlılık)	F1-Skoru (F1-Score)
1	50	1.0	-	0.417	0.417	0.417
2	50	1.0	-	0.266	0.532	0.355
3	50	1.0	-	0.199	0.596	0.298
4	50	1.0	-	0.172	0.686	0.275
5	50	1.0	-	0.144	0.718	0.240

Tablo 4' den görüleceği üzere, Kararegitim3-Karartest3 test grubu için en yüksek F1 skorlarına ilk iki suç türünün tahmininde ulaşıldığı görülmektedir.

Tablo 4. Kararegitim3-Karartest3 Test Grubu F1 Skorları

test k	Tekrar Sayısı (epoch)	Öğrenme Oranı (lr: learnin g rate)	Kelime Büyüklüğü (wordN grams)	precisi on (Kesinlik)	recall (Duyarlılık)	F1-Skoru (F1-Score)
1	50	1.0	-	0.556	0.556	0.556
2	50	1.0	-	0.340	0.679	0.453
3	50	1.0	-	0.239	0.716	0.358
4	50	1.0	-	0.182	0.728	0.291
5	50	1.0	-	0.149	0.747	0.248

Tablo 2,3,4' de görüleceği üzere, suç türü sınıflandırmaları için en yüksek tahmin oranlarının ilk iki suç türünün tahmininde elde edilmiştir. Test gruplarının ilk iki tahmin ortalamaları ise Tablo 5' de verilmektedir. Birinci tahminlerinde ortalama F1 skorunun 0.447, ikinci tahminlerinde ortalama 0.400 olduğu hesaplanmıştır.

Bu suretle modelin mevcut öğrenme oranları için ilk iki sınıflandırma tahminin kullanılmasının uygun olacağı görülmüştür.

Tablo 5. Ortalama Kesinlik, Duyarlılık ve F1-Skoru değerleri

	Ortalama Kesinlik	Ortalama Duyarlılık	Ortalama F1-skoru)
1. suç türü tahmini	0.447	0.447	0.447
2. suç türü tahmini	0.300	0.600	0.400

Kesinlik ve Duyarlılık Değerleri

Kesinlik (precision) ve duyarlılık (recall) değerleri, tahminlerin gerçek pozitif (GP), gerçek negatif (GN), yalancı pozitif (YP) ve yalancı negatif (YN) değerleri üzerinden hesaplanmaktadır. Bu iki değer sonuçların ilgililiğine bağlıdır.

Denklem(2)' de görülen, kesinlik değeri, doğru tahminlerin tüm tahminler içerisindeki oranıdır. Yani, modelin öğrenme oranıdır.

$$\text{kesinlik} = \frac{GP}{GP + YP} \quad (2)$$

Denklem(3)' de görülen, duyarlılık değeri ise doğru tahminlerin ilgili tahminler içerisindeki oranıdır.

$$\text{duyarlılık} = \frac{GP}{GP + YN} \quad (3)$$

Bu durumda, modelin birincil tahminleri için kesinlik ve duyarlılık değerleri aynı olurken, modelin ikincil tahminleri de göz önünde

bulundurulduğunda, modelin kesinlik değeri düşerken duyarlılık değeri yükselmekte, yani iki sınıf tahmini içerisinde doğru tahmine ulaşma ihtimali artmaktadır.

Sonuç

Yargıtay Ceza Genel Kurulu tarafından 2020 yılına ait 531 adet karar suç türlerine göre etiketlenmiş, 1/3 (177/354 karar) oranlarında eğitim ve test gruplarına ayrıştırılmış, çapraz doğrulama amaçlı olarak da her bir grupta test verileri eğitim verileri ile değiştirilerek 3 adet test grubu oluşturulmuştur.

Her üç test grubu FastText metin sınıflandırma modelinin danışanlı öğrenme yöntemi ile eğitilmiştir.

Eğitimde FastText modeli parametresiz ve tekrar sayısı (epoch), öğrenme oranı (lr:learning rate) ve kelime büyüklüğü (wordNgrams) parametrelerinin her birisi için 3' er farklı değer ile tüm kombinasyonlar için 3 test grubu üzerinde çalışılmıştır.

Ayrıca, FastText otomatik optimizasyon seçeneği ile parametrelere bağlı denemelerden daha iyi bir öğrenme oranına ulaşılmaya çalışılmıştır. Ancak; her üç test grubu için de parametrelere bağlı denemelerde otomatik optimizasyon seçeneğine göre daha yüksek oranlar elde edilmiştir.

Karareğitim1 - Karartest1 test grubu için parametrelere bağlı denemelerde kesinlik ve duyarlılık değerleri %36,9 iken otomatik optimizasyon seçeneğinde %11,9 olarak elde edilmiştir.

Karareğitim2 - Karartest2 test grubu için parametrelere bağlı denemelerde kesinlik ve duyarlılık değerleri %41,7 iken otomatik optimizasyon seçeneğinde %11,5 olarak elde edilmiştir.

Karareğitim3 - Karartest3 test grubu parametrelere bağlı denemelerde kesinlik ve duyarlılık değerleri %55,6 iken otomatik optimizasyon seçeneğinde %15,4 olarak elde edilmiştir.

Test grupları arasında elde edilen kesinlik ve duyarlılık değerleri itibariyle büyük bir fark olmadığı değerlendirilmiştir.

FastText epoch=50 ve lr=1.0 parametre değerleri için birinci suç türü tahmininde en yüksek öğrenme oranlarına ulaşılmıştır. Ulaşılan öğrenme oranlarının ortalamasının %44,7 olduğu görülmüştür.

Karareğitim1-Karartest1 test grubu için F1 skorunun ikincil tahminlerde daha yüksek olduğu görülmüştür. Bu nedenle tüm gruplar için ikincil tahmin değerleri de hesaplanmıştır.

FastText epoch=50 ve lr=1.0 parametreleri için ikincil suç türü tahmininde ortalama öğrenme oranının %30.0 olduğu görülmüştür.

Çalışmanın amacı olan, Yargıtay Kararları ile eğitilen bir modelin, avukatlık hizmetlerinde, emsal karar arama sistemlerinin alt yapısı olarak kullanılarak avukatların anahtar kelime tabanlı emsal karar arama sistemlerine göre daha kısa sürede daha ilgili emsal kararlara ulaşması açısından elde edilen %44,7 öğrenme oranı değerlendirilmiştir.

Bu öğrenme oranı oldukça düşük bir seviyedir, ancak; anahtar kelime tabanlı elde edilen sonuçlarla kıyaslandığında, elde edilen sonuçların hangi suç türüne ait olduğu bilinmiyorken, FastText ile eğitilmiş bir arama altyapısı kullanıldığında elde edilen sonucun yaklaşık yarısının istenen suç türüne ait olduğu bilinecektir.

Uygulamada 531 Yargıtay kararının 78 farklı suç türüne ait olduğu düşünülecek olduğunda, herhangi bir suç türü için yarı yarıya doğru tahminde bulunmak arama sonuçları içerisinde suç türüne bağlı filtreleme yapıldığında doğru suç türlerinin yarısının elde edilmesi diğer yarısının ise yanlış etiketlendiği için gözden kaçırılmasına neden olacaktır.

Bu durumda, eğer doğru etiketlenen kararlar ulaşılmak istenen emsal karara ulaşmayı sağlamış ise istenen duruma yani en kısa süre içerisinde doğru suç türüne ait bir emsal karara ulaşma amacına hizmet etmiş olacaktır.

Bu halde, öğrenme oranına bağlı olarak elde edilen sonuçlar doğru ve yanlış etiketlenmiş kararlardan oluşacağı için, öğrenme oranı doğru etiketlenmiş sonuçların da artacağı, yanlış etiketlenmiş kararların ise azalacağı anlamına gelecektir. Ancak her halükarda, yalnızca anahtar kelimeye bağlı olarak edilen sonuçlarla kıyaslandığında, ilgililik yani istenen suç türüne ait kayıtların toplam sonuçlar içerisindeki miktarı artacaktır. Her ne kadar %44,7 oranı yetersiz olsa da uygulamada fayda sağlayabilecek bir sonuç olarak değerlendirilmiştir.

Teşekkür

Bu çalışma, Berker KILIÇ tarafından Ondokuz Mayıs Üniversitesi, Lisantüstü Eğitim Enstitüsü, Veri Bilimi Anabilim Dalında, Prof. Dr. Yüksel ÖNER danışmanlığında yürütülen “Yargıtay Kararlarının Suç Türlerine Göre Veri Bilimi Yöntemleri İle Sınıflandırılması” başlıklı Yüksek Lisans projesinden türetilmiştir. Makale özeti 5. Uluslararası Adli Bilimler ve Hukuk Kongresi’ nde 17/08/2021 tarihinde sözlü olarak sunulmuştur.

Kıymetli Hocam Prof. Dr. Yüksel ÖNER’ e katkıları için teşekkürlerimi sunarım.

Çalışma içerisinde yer alan kaynak kodlara ve tüm test sonuçlarına

<https://github.com/berkerkilic/yargitaykarar>

adresinden erişilebilir.

Kaynaklar

- [1] Ortadoğu Teknik Üniversitesi, “Kodex Bilişim’ den Yargıtay Davalarına Yapay Zekalı Destek”, <https://odtuteknokent.com.tr/tr/haber/kodex-bilisimden-yargitay-davalarina-yapay-zekali-destek> (12.05.2021).
- [2] Hukukwork, “HukukWork Özellikler”, <https://www.hukukwork.com/ozellikler/>, (12.05.2021).
- [3] İstanbul, Ankara ve İzmir Baroları, “Yapay Zeka Çağında Hukuk Çalıştay Raporu”. *İstanbul, Ankara ve İzmir Baroları Çalıştay Raporu*, 2019.
- [4] AdaletHanım, “AdaletHanım Hakkımızda”, <http://adalethanım.com/#hakkımızda>, (12.05.2021).
- [5] Bilkent Üniversitesi, “Hukuk Metinlerinde Yapay Zekâ Uygulamaları için Doğal Dil İşleme ve Makine Öğrenmesi Teknikleri”, <http://umram.bilkent.edu.tr/index.php/2021/03/01/42-hukuk-metinlerinde-yapay-zeka-uygulamaları-icin-dogal-dil-isleme-ve-makine-ogrenmesi-teknikleri/>, (12.05.2021).
- [6] Kınık D, Güran A. “TF-IDF ve Doc2Vec Tabanlı Türkçe Metin Sınıflandırma Sisteminin Başarım Değerinin Ardışık Kelime Grubu Tespiti ile Arttırılması”, *Avrupa Bilim ve Teknoloji Dergisi*, (21), 323-332, 2021.
- [7] Onan A. “Evrışimli Sinir Ağı Mimarilerine Dayalı Türkçe Duygu Analizi”, *Avrupa Bilim ve Teknoloji Dergisi*, Özel Sayı, 374-380, 2020.
- [8] Çelik Ö, Koç BC. “TF-IDF, Word2vec ve Fasttext Vektör Model Yöntemleri ile Türkçe Haber Metinlerinin Sınıflandırılması”, *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 23(67), 121-127, 2020.
- [9] Ahmetoğlu H, Daş R. “Türkçe Otel Yorumlarıyla Eğitilen Kelime Vektörü Modellerinin Duygu Analizi

İle İncelenmesi”, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 24(2), 455-463, 2020.

- [10] Aksu MÇ, Karaman E. “FastText ve Kelime Çantası Kelime Temsil Yöntemlerinin Turistik Mekanlar İçin Yapılan Türkçe İncelemeler Kullanılarak Karşılaştırılması”, *Avrupa Bilim ve Teknoloji Dergisi*, (20), 311-320, 2020.
- [11] Turan T, Kemaloğlu N, Küçüksille EU. “Hukuk’ ta Yapay Zeka: Çalışmalar ve Gelecek Öngörülleri”, *Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 11(2), 246-255, 2020.
- [12] Facebook, “Get Started”, <https://fasttext.cc/docs/en/support.html>, (12.05.2021).