



Bilgi Yönetimi Dergisi

Cilt: 2 Sayı: 2 Yıl: 2019

<https://dergipark.org.tr/tr/pub/by>



Hakemli Makaleler

Araştırma Makalesi

Makale Bilgisi

Gönderildiği tarih: 03.12. 2019

Kabul tarihi: 24.12. 2019

Yayınlanma tarihi: 31.12. 2019

Article Info

Date submitted: 03.12. 2019

Date accepted: 24.12. 2019

Date published: 31.12. 2019

Anahtar sözcükler

Makine öğrenmesi, Destek Vektör Makinesi, Metin Madenciliği, Standart Dosya Planı, Belge Sınıflama

Keywords

Machine Learning, Support Vector Machine, Text Mining, Standart File Plan, Record Classification

DOI numarası

10.33721/by.654464

ORCID

0000-0002-8071-9693

Makine Öğrenmesi Yaklaşımıyla e-Belgelere Standart Dosya Plan Numaralarının Otomatik Olarak Atanması Üzerine Bir Çalışma¹

A Study on Automatic Assignment of Standard Classification Numbers to e-Records with the Machine Learning Approach

Kasım BİNİCİ

Çankırı Karatekin Üniversitesi Bilgi ve Belge Yönetimi Bölümü Öğretim Üyesi
kbini@karatekin.edu.tr

Öz

Belge üretimindeki artış ve teknolojik olanaklar beraberinde yeni yönetim metodlarının gerekliliğini kaçınılmaz kılmıştır. Türkiye’de kamu kuruluşlarında üretilen belgeler Standart Dosya Planı’na uygun olarak düzenlenir ve yönetilirler. İlgili mevzuata koşut olarak resmi yazışmaların konusu Dosya Planından saptanmak ve konuyla bağlantılı kodları belgelere eklemek zorunluluktur. Bu kodların doğru seçilmesi, araştırma-soruşturma süreçlerinin sağlıklı işletilebilmesi ve erişim süreçlerinin başarılı şekilde sonuçlandırılabilmesi için oldukça gereklidir. Ancak kurumsal, kişisel veya yönetsel koşullara bağlı olarak belgelerin yaşam döngüsünü sekteye uğrattıkları yanlış kodlar verilebilmektedir. Bu tür yanlış uygulamaları minimize etmek ve belge sınıflandırmayı daha sağlıklı kılabilmek için yapay zekâ uygulamalarından yararlanılabilir.

Elektronik belge yönetimi sistemlerinde üretilen belgelere standart dosya plan kodlarının makine öğrenmesi yaklaşımıyla otomatik olarak atanması amaçlanan bu çalışma teorik ve analize dayalı olmak üzere iki kısımdan oluşmaktadır. İlk teorik olarak standart dosya planından yararlanarak otomatik belge sınıflandırmasının oluşturduğu güçlükler tartışılmış, ardından makine öğrenmesi ile belgelerin sınıflandırılması üzerine analiz yapılmıştır. Çeşitli yönetsel ve ön yargısal bariyerlerin aşılması ve yanı sıra kurumsal arşiv gibi otorite bir birimin olmaması belge yönetimi, eğitimi ve denetimi boşluğunu oluşturduğu ve söz konusu bu durumunun otomatik sınıflamayı sekteye uğrattığı kaygısı, dolayısıyla belgeleri yeniden sınıflandırma gerekliliği küçük bir veri kümesi ile çalışmayı zorunlu kılmıştır. Bu nedenle çalışmada analiz edilen belgeler, bu çalışmanın araştırmacısına kurum içerisinde yönlendirilen son altı aylık belgelerden oluşmaktadır. Toplamda 265 belgenin yeniden sınıflandırılması neticesinde tekil konudaki belgeler kapsam dışı bırakılmıştır. Belgelerin gövde ve konu alanları üzerinde yapılan metin madenciliği teknikleri uygulanması sonucunda, 169 belgeden oluşan bir veri seti elde edilmiştir. Bu veri setinden her konudan oransal olmak koşuluyla rastgele yöntemle belgelerin üçte biri (1/3) sınıflandırmak için seçilmiştir. Sınıflandırılmış 112 belge ve sınıflandırmak üzere oluşturulmuş 57 belgeden ibaret bu veri seti üzerinde, makine öğrenmesinde kullanılan ve son zamanlarda bilgi sektöründe popüler olan Destek Vektör Makinesi [DVM (Support Vector Machine (SVM)] algoritması çalıştırılmıştır. Çalışma sonucunda manuel olarak yapılan sınıflama ile otomatik olarak yapılan çıkarımın isabet oranı % 87.72 olarak bulunmuştur. Bir diğer ifade ile belgelerin % 87.72’si makine öğrenmesi yaklaşımıyla doğru olarak sınıflanmıştır.

¹ Bu çalışma e-BEYAS 2019 Sempozyumunda kabul edilerek sözlü sunumu yapılan bildirinin tam metnidir.

Abstract

Technological opportunities and an increase in the production of records have made inevitable the necessity of new management methods. Documents produced in public institutions in Turkey are organized and managed according to the Standard File Plan. Under the relevant law, it is mandatory to determine the subject of official correspondence from the File Plan and add the relevant codes to the records. The correct selection of these codes is essential for the healthy operating of the research- investigation processes and the successful completion of the access processes. However, incorrect codes have been given depending on institutional, personal, or managerial conditions that will interrupt the life cycle of records. Artificial intelligence applications can be utilized to minimize such misapplications and to make records classification more powerful.

This study, which is intended to automatically assign standard file plan codes with a machine learning approach to the records produced in electronic record management systems, consists of two parts as theoretical and analysis based. Firstly the difficulties of the automatic record classification were discussed in theory by using the standard file plan. Then the classification of records with machine learning was analyzed. Not to overcome the various administrative and prejudicial barriers, as well as the absence of an authority unit such as the institutional archive, were concerned document management, training, and auditing create a gap, and this will hamper automatic classification. Therefore, the necessity to reclassify records has made it necessary to work with a small data set. For this reason, the records analyzed in the study consist of records sent to the researcher of this study within the institution in the last six months. After the reclassification of 265 records in total, records on unique subjects excluded. As a result of the application of text mining techniques on the body and subject areas of the records was obtained a dataset consisting of 169 records. From this data set, provided that each subject is proportional, one-third (1/3) of the records had been randomly selected. Supported Vector Machine (SVM) algorithm used in machine learning and recently popular in the information sector was run on this dataset consisting of 112 classified records and 57 unclassified records. As a result of the study, when the manual classification and automatic classification compared, the accuracy rate was 87.72%. In other words, 87.72% of the records were classified correctly with the machine learning approach.

1. Giriş

Evrendeki gizemleri öğrenip ve işlemede mahir olan insanoğlu, tarih boyunca meydana getirdiği eser, makine ve yapıtlarına kendinden, diğer canlılardan ve doğadan esinlenerek bir şeyler katmıştır. Çevresindeki canlı, cansız, soyut ve somut varlıkların özelliklerini hatta kendi özellik, yetenek ve duygularını keşfederek yapıtlarına aktarmayı başarmıştır. Doğadaki ve kendisindeki özellikleri öğrenip yapıtlarına aktarma kabiliyeti bir birikimin sonucu olarak sürekli gelişim göstermektedir. Bilgisayar teknolojilerinin ileri düzeyde gelişim gösterdiği çağımızda insanlığın kendini taklit ederek yeni yapıtlar ortaya koymanın farklı bir merhalesini yaşamaya tanıklık etmekteyiz. Şimdiye kadar meydana getirilen eserlere verilen çabalar içerisinde belki de en hayret edilecek mevzu, insanlığın kendi yapıtlarının kendisi gibi düşünmesi ve karar vermesine yönelik verdiği uğraştır. Bu uğraşın algısal ya da bilişsel süreçlerle ilgili bilgisayar modellerini geliştiren araştırma alanı bilim camiasındaki nitelendirilmesi yapay zekâ olarak bilinmektedir.

Pek çok teknoloji alanında olduğu gibi yapay zekânın da kendine özgü kavramları ve iç içe olduğu konular vardır. Yapay zekânın temel uygulama alanları, uzman sistemler, kural motorları, makine öğrenmesi, derin öğrenme, sinir ağları, doğal dil işleme, bilgisayar görüşü, bilgisayar yardımcı eğitim vb. gibi geniş bir konu listesi sıralanabilir. Çeşitli teknik ve teknolojilerin yeniden keşfi sonucunda bu teknolojilerin statüsü ya da özeliği ile ilgili karışıklık daha da artmaktadır (Rolan ve diğerleri, 2019, s. 181). Konunun daha çok bilişim alanındaki teknolojilerin uğraş alanını ilgilendirmesi nedeniyle kavramsal tartışma bu çalışmanın amacı dışında tutulmuştur. Dolayısıyla geniş konu yelpazesinden sıyrılarak bu çalışmanın konusu bilgi sınıflamada makine öğrenmesi ile sınırlandırılmıştır. Çalışmada denetimli öğrenme yaklaşımı ile belgelerinin sınıflandırılmasında Destek Vektör Makinesi [DVM (Support Vector Machine (SVM))] algoritması ve uygun veri setlerinin hazırlanmasında da metin madenciliği tekniklerinin kullanımı konu edinilmiştir.

Bilgi üretimindeki artış beraberinde yeni yönetim metotlarının gerekliliğini kaçınılmaz kılmaktadır. Oluşan büyük verilere karşı yapay zekâ, veri-metin madenciliği, makine öğrenmesi gibi tekniklerden yararlanılarak yeni bilgi veya belge yönetimi arayışları bilgi sektöründe de tartışılan ve araştırılan bir konu olmuştur. Arşivlerde yapay zekâ uygulamaları, belge yönetimi süreçlerinden başlayarak arşiv yönetimi süreçlerine kadar etkin ve uygun çözümleri üretebilecek bir yapıyı içerisinde barındırmaktadır (Cibaroğlu ve Yalçınkaya, 2019, s. 49). Kurumların yönetsel işleyişlerinde elektronik belge yönetimi

sistemlerinin kullanımı sözü edilen tekniklerin uygulanabilirliği için uygun zemin oluşturmaktadır. Bu nedenle arşiv belgelerine makine öğrenmesi yaklaşımıyla standart dosya planı numaralarının tahmin edilip, atanması yürütülen bu çalışmanın en temel amacını oluşturmaktadır. Ayrıca ülkemizde bu konudaki araştırmaların azlığı ve uygulamaya dayalı çalışmaların alan literatürüne yansımamış olması bu çalışmaya ayrı bir önem yüklemektedir.

2. Denetimli Makine Öğrenmesi

Makine öğrenmesi yapay zekanın ilişkili olduğu bir konudur. İngilizce “machine learning” olarak kullanılan bu terim “mekanik öğrenme”, “otomatik öğrenme”, “yapay öğrenme” olarak dilimize geçmiş olup “yapay zekâ problemlerinin çözümüne yönelik, gözlem ve algı verileriyle birlikte geçmişteki hataları da göz önüne alarak iyileştirmeler yapabilen algoritmaların geliştirilmesini içeren disiplin” olarak tanımlanmaktadır (TÜBA, 2019). Makine öğrenmenin denetimli, denetimsiz ve takviyeli² öğrenme gibi farklı türleri vardır. Denetimli öğrenmede DVM, Doğrusal Regresyon (Linear Regression), Lojistik Regresyon (Logistic Regression), Naïve Bayes, Doğrusal Diskriminant Analiz (Linear Discriminant Analysis), Karar Ağaçları (Decision Trees), K-En Yakın Komşu (K-Nearest Neighbor) ve yapay sinir ağlarından Çok Katmanlı Algılayıcı (Multilayer Perceptron-MLP) bilinen en yaygın algoritmalarlardır.

Denetimli makine öğrenmesi için geliştirilen algoritmaların her ne kadar formülasyon, işleyiş biçimleri ve modellendiği dış çevre farklı olsa da öğrenmede ve tahmin etmede getirdikleri yaklaşım benzerlik göstermektedir. Bu türde makinenin öğrenebilmesi için tüm algoritmalar için dış bir mekanizma tarafından gerekli giriş-çıkış verilerinin tanımlanması gerekmektedir. Sınıflama için yapılan işlemlerde bu veri kalıpları eğitim seti olarak tanımlanmaktadır. Sisteme girdi olarak tanımlanan eğitim verilerindeki parametrelere denk gelen çıktılar, etiketlerin, tanımlı olması zorunludur. Yani sınıflandırmanın yapılmış olması gerekmektedir. Burada eğitim setindeki parametreler problemin tüm özelliklerini kapsayacak şekilde tanımlanmalıdır. Makine öğrenmesinde başarı bu veri setinin kapsayıcılığına bağlıdır. Yapılan öğrenmenin ölçülmesi ise sisteme daha önce sunulmamış veri grubunu oluşturan test seti ile yapılmaktadır. Eğitim setinden farklı olarak test setindeki verilerde girdiler tanımlıdır ancak çıktılar tanımlı değildir. Bir diğer ifade ile bu veri seti sistemin daha önce karşılaşmadığı, sınıflandırılmamış kayıtları oluşturmaktadır. Bu modelde makine eğitim veri setinden girdi ve çıktılar arasındaki ilişkiden öğrendiğini, yaptığı hesaplamaları, test setine uygulayarak sınıfları üretmektedir.

3. Literatür Değerlendirmesi

Birçok alanda uygulanan makine öğrenmesinin bilgi sektöründe de kullanımı vardır. Sınıflama işlemlerinin en temel uğraş olduğu bilgi merkezleri makine öğrenmesinin odak noktası ile örtüşmektedir. Bilgi sektöründe otomatik metin sınıflama işlemlerinde makine öğrenmesinin algoritmaları kullanılmıştır. En yaygın olarak kullanılan Bayesian Model, K-En Yakın Komşu ve DVM algoritmaları kapsamlı olarak Sebastiani (2002)'nin çalışmasında ele alınmıştır. Bu tür algoritmalar daha çok spam filtreleme, haber ve dergi makalelerini kataloglama ve internet sayfalarını sınıflandırılma uygulamalarında kullanılmaktadır. Genel olarak denetimli makine öğrenmesi tabanlı bir otomatik metin sınıflama algoritması bir sınıflandırma işlemini eğitmek için el ile sınıflandırılmış belgelerin oluşturduğu bir dermeyi kullanır. Daha sonra etiketlenmemiş belgelerin sınıflarını / notasyonunu tahmin etmek için bu derme referans alınmaktadır (Joorabchi ve Mahdi, 2011, s. 500).

Kütüphane ve bilgi biliminde bibliyografik sınıflandırma için makine öğrenmesi, veri ve metin madenciliği tekniklerinin kullanımı üzerine yapılmış birçok çalışma mevcuttur. Bunlar özetle dijital ortamda doküman benzerliği, bulanık (fuzzy) yaklaşım, kümeleme, kategorizasyon, TF-IDF ağırlıklandırma, bilgi çıkarımı, ontolojik yaklaşım, denetimli ve denetimsiz makine öğrenme ile otomatik metin sınıflandırma şeklinde kategorize edilebilir (Ullah, Khusro ve Ullah, 2017, ss. 51-53).

Alanyazında kütüphanelerde kullanımı yaygın olan sınıflama sistemlerini makine öğrenmesi yaklaşımı ile sınıflama numarasının tahmin edilmesi ve otomatik olarak atanması üzerine yapılan çalışmalar

² Literatürde denetimli için danışmanlı veya öğretmenli, denetimsiz için danışmansız veya öğretmensiz, takviyeli için destekleyici veya pekiştirerek gibi kelimelerin kullanıldığı görülmektedir.

vardır. Dijital kütüphaneler ve depolarda arşivlenen bilimsel literatürün Dewey Onlu Sınıflama Şemasına göre otomatik sınıflandırılmasının konu edinildiği bir çalışmada denetimsiz makine öğrenmesi yaklaşımı sergilenmiştir. Burada geleneksel kütüphanelerde kataloglanmış olarak bulunan künyelerin konu sınıflandırma meta verilerini kullanarak, ağırlık mekanizması yardımıyla dokümanın kendisi için en muhtemel notasyonu çıkarma işlemi yapılmıştır. Çalışma sonucunda makine öğrenmesi tabanlı yaklaşımla elde edilen sonuçlara göre iyi bir performans ortaya konulduğu aktarılmaktadır (Joorabchi ve Mahdi, 2011). Bir diğer çalışma ise denetimli makine öğrenmesi tekniğiyle bilim ve teknoloji alanındaki bibliyografik verilere otomatik olarak Dewey Onlu Sınıflama notasyonunun atanması üzerine yapılmıştır. On yıldan fazla künye birikimi olan veri seti üzerine Naïve Bayes (NB) ve DVM algoritmaları ile yapılan deneylerde yaklaşık %90'lık bir sınıflandırma doğruluğu elde edilmiştir (Wang, 2009). Söz konusu bu çalışmalarda da anlaşılmaktadır ki denetimli ve denetimsiz makine öğrenmesi teknikleri kütüphanelerin yoğun uğraş alanı olan sınıflandırma işlemlerinde kullanılabileceğidir.

Kütüphanelerde olduğu gibi arşiv ve belge yönetimi alanı da yapay zekâ ve makine öğrenmesi teknikleri için uygun bir ortam oluşturmaktadır. Önceki çalışmalardan yapılan çıkarımlara dayanarak, belge yönetim sürecinde yapay zekâ uygulamalarının kullanımının tartışıldığı alanlar şu şekilde özetlenmiştir. Literatürde belgelerin saklama planlarının otomatikleştirilmesi, sınıflandırmanın yapay zekâ ile yapılması, açık devlet verilerinin halka sunulması ve belge yaşam döngüsünün yapay zekâ tarafından kontrol edilmesinden bahsedilmektedir (Cibaroğlu ve Yalçınkaya, 2019, s. 50). Örneğin yapılan bir uygulama çalışmasında makine öğrenmesinin arşiv belgelerinin imha sürecinde yapılandırılmamış (sınıflandırılmamış) belgelerin sınıflandırılmasında başarılı bir sonuç ortaya çıkarttığı anlaşılmaktadır. New South Wales Devlet Arşivlerinde yapılan bir pilot projesinde imha işlemlerinde daha önce analog olarak uygulanan arşiv belgeleri üzerine Çokterimli Naif Bayes (Multinomial Naif Bayes-MNB) ve Çok Katmanlı Algılayıcı (Multi Layer Perception-MLP) makine öğrenmesi algoritmaları uygulanmıştır. Bir dizi veri işleme sürecinden sonra yapılan testler neticesinde MLP algoritmasının daha başarılı olduğu ve %84 oranında başarı sağladığı görülmüştür (Rolan ve diğerleri, 2019, ss. 190-193).

Yapay zekâ veya makine öğrenmesi konusunda Türkiye'de ise kütüphane ve arşiv alanında yapılan çalışmaların sayısı oldukça azdır. Kurumsal belgelerin saklama süreçlerinde ve arşiv değeri olan malzeme seçiminde yapay zekânın Türkiye'de kullanım koşullarını açıklama amacıyla yapılan bir çalışmada dünyada belge ve arşiv yönetimi süreçlerinde yapay zekâ uygulamaları teorik olarak ele alınmıştır (Cibaroğlu ve Yalçınkaya, 2019). Bir diğer önemli çalışma ise Ankara üniversitesinde yapılan projedir. Projede yapay zekânın bilgi-belge süreçlerine uygulanabilirliği, kurum hafızasının oluşturulması, geçmiş kararların yeni karar süreçlerine dâhil edilebilmesinin yöntemleri araştırılmaktadır. Ayrıca yapay zekâ algoritmaları tarafından yönetilen iki sanal birim arasında yazışmaların gerçekleştirilmesi, belli kategorideki belgelerin yapay zekâ tarafından, geçmiş tecrübelerle dayanarak otomatik imzalanmasının zemini, e-dosyalama sisteminin etkin kullanımı gibi konuların araştırılıp tartışılması amaçlanmıştır (Ünal ve Özdemirci, 2017, ss. 61-62).

4. Standart Dosya Planı

Dosyalama sistemi arşiv ve belge yönetiminin ayrılmaz ve önemli bir bileşenidir. Dosyalama işleminin temel amacı ihtiyaç duyulan bilgi/belgeye hızlı ve ekonomik bir şekilde erişim sağlamaktır (Özdemirci ve diğerleri, 2009, s. 174). Türkiye'de kamu kurumlarında belgelerin dosyalama işlemleri, eski adıyla Başbakanlık Devlet Arşivleri Genel Müdürlüğü, yeni adıyla Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı'nın öncülüğünde kurumların işleyişine uygun olarak hazırlanan ve bir sınıflama sistemi hüviyetini taşıyan Standart Dosya Planları referans alınarak gerçekleştirilmektedir. Standart Dosya Planı aynı konudaki belgelerin bir arada yer almasını ve alt düzeydeki konularında hiyerarşik bir yapıda tutulmasını hedeflemektedir. Her konu ve alt konuyu temsilen bir notasyon önerilmektedir. Belge üretiminin ilk safhasında belge için bir sınıflama kodu tanımlanmaktadır. Bu kodların doğru seçilmesi, araştırma-soruşturma süreçlerinin sağlıklı işletilebilmesi, belge yaşam döngüsünün verimli yönetilebilmesi ve erişim işlemlerinin başarılı sonuçlandırılabilmesi için oldukça gereklidir. Ancak kurumsal, kişisel veya yönetsel koşulların oluşturduğu belirsizlikler nedeniyle belgelerin yaşam döngüsünü sekteye uğratacak yanlış kodlar verilebilmekte ve bunun denetimi maalesef yapılamamaktadır. Sözü edilen olumsuzluğun başlıca nedenleri Türkiye'de kurumsal arşiv yapılarının

yeterince olgunlaşmamış olması; belgelerin yaşam döngüsünü oluşturan iletişim kanallarının tam sağlanamamış olması; elektronik belge yönetim sistemlerinden sorumlu personellerin meslek elemanı olmamaları, yanı sıra sistem yönetimine kısıtlı erişimleri, sisteme tam hakim olamamaları ve belge yönetiminde fikir yürütme ve karar verme pozisyonunda olmamaları gösterilebilir. Dolayısıyla Standart Dosya Planı'nın amaçladığı sınıflama işlemlerinin doğru olarak yapılabilmesi için bilgisayar yeteneklerinden yararlanılarak en uygun tekniklerin belirlenmesine yönelik bulgusal araştırmalara ihtiyaç vardır.

Resmi yazışma kurallarına göre kamu kuruluşlarında belgelerin standart dosya plan numaralarını alması bir zorunluluktur. Her kurumun işleyişine bağlı olarak temelde benzerlik gösteren, fakat ayrıntılarda farklılaşan dosya planları vardır. Buna karşın yönetsel işleyişlerde benzer belgelerin üretildiği yükseköğretim üst kuruluşu ve kurumları içinse tek bir standart dosya planı hazırlanmıştır. Bununla ilgili standart çalışmaları 2005/7 sayılı Başbakanlık Genelgesi ile başlamıştır. Tüm üniversitelerin ana hizmet faaliyetlerinin aynı usul ve esaslar çerçevesinde dosyalanmasını sağlamak için Ankara Üniversitesi tarafından Devlet Arşivleri Genel Müdürlüğü'nün işbirliği ile yürütülen bir TÜBİTAK projesi kapsamında ilk çalışmalar yapılmıştır. Bu projede "üniversiteler için saklama süreli dosya planı" hazırlanmıştır. Daha sonra Yükseköğretim Kurulu Başkanlığı ana hizmet faaliyetleri için hazırlanan dosya planı ile birleştirilerek "Yükseköğretim Üst Kuruluşları ve Yükseköğretim Kurumları Saklama Süreli Standart Dosya Planı" adı altında son biçimini almıştır (Özdemirci, Torunlar ve Saraç, 2009, ss. 177-178). Tüm Yükseköğretim birimlerini ilgilendiren bu dosya planı farklı zamanlardan gözden geçirilerek yayınlanmış ve en güncel sürümü Aralık 2017 olup YÖK'ün internet sayfasında erişimi sağlanabilmektedir (T.C. Yükseköğretim Kurulu, 2017). Bu çalışmanın odak noktasında olan belgeler üniversite bünyesindeki birimler tarafından üretildiğinden söz konusu dosya planı referans alınarak sınıflandırılmıştır.

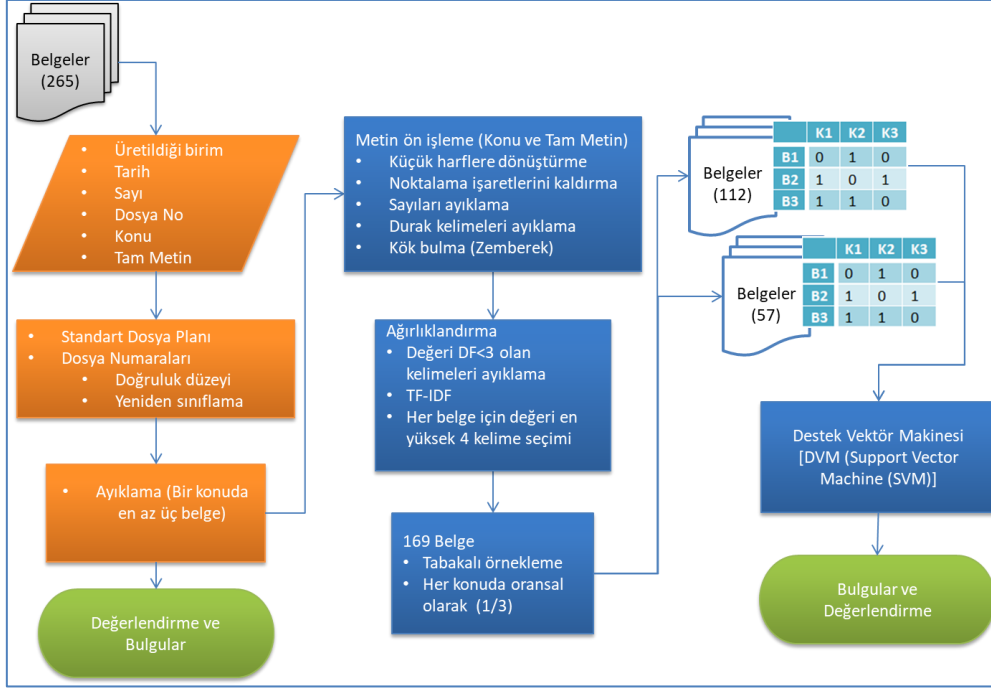
5. Yöntem

Denetimli makine öğrenmesinin olduğu araştırmalarda genellikle araştırmacılar çalışmalarını yığın veri setleri üzerinde sürdürmektedir. Yoğun veri işlem uğraşının olduğu bu tarz araştırma sonuçlarının başarısı ele alındığı veri miktarının hacmi ile yakından ilgilidir. Ancak çeşitli yönetsel ve ön yargıya dayalı sınırların aşılması araştırma için daha elverişli olan verilerin elde edilmesini güçleştirmektedir. Diğer taraftan belge üretiminde kurumsal arşiv gibi otorite bir kurumun olmaması belge yönetimi, eğitimi ve denetimi boşluğunu oluşturmaktadır. Dolayısıyla dosya numaralarının yanlış veya gelişigüzel olarak atanması olasılık dâhilindedir. Nitekim belgeler üzerinde yapılan ön incelemede dosya numaralarındaki tutarsızlıkların fazla olduğu tespit edilmiştir. Sözü edilen durum otomatik sınıflamayı sekteye uğratacağı kaygısı oluşturmuştur. Bu nedenle belgeleri yeniden sınıflandırma gerekliliği daha az bir veri kümesi ile çalışmayı zorunlu kılmıştır. Araştırmanın verileri bu çalışmanın araştırmacısına Çankırı Karatekin Üniversitesi birimleri tarafından yönlendirilen son altı aylık belgelerden oluşturulmuştur.

Denetimli makine öğrenmesinden yararlanarak belge sınıflamak için uygun veri matrislerinin olması gerekmektedir. Veri hazırlamak için izlenen yöntem aşamalı olarak Şekil-1'de görselleştirilmeye çalışılmıştır. Araştırma verilerini elde etmek için öncelikle kurumun elektronik belge yönetimi sisteminde toplamda 265 belgenin üst veri ve tam metinleri manuel olarak indirilmiştir. Resmi yazışma usul ve esaslarında belgelere ekler eklenebilmektedir. Yanı sıra kurum yazışma geleneğinde üst birimden gelen yazılar, üst yazı ile alt birimlere iletilmektedir. Bu tür ekler bir belgenin konusu ve kapsamı hakkında zengin içerikler barındırmaktadır. Bu araştırmanın odağındaki belgelerin ek miktarı da oldukça yüksektir. Belge ekleri araştırma sürecine yoğun bir iş yükü getireceği düşüncesiyle kapsama dâhil edilmemiştir. İndirilen belgelerden uygun öz nitelikleri elde etmek için üst veri ve tam metin dosyalarından belgenin üretildiği birim, tarih, sayı, dosya numarası, konu ve tam metin değerleri elde edilerek bir veri seti oluşturulmuştur.

Araştırma kapsamındaki veri kümesi üzerindeki incelemeler iki aşamada ele alınmıştır. Birinci aşamada teorik bir yaklaşım sergilenerek atanmış dosya numaralarının doğruluğu ve yeniden numaralandırması üzerinde çalışılmıştır. Bu aşamada belge üretiminin başlangıç evresinde atanan kodlamaların tutarlılığı ile belgelerin içerik ve yapısal boyutu incelenmiş, makine öğrenmesi ile metin sınıflandırmasını sekteye uğratan olgular saptanmıştır. Farklı birim ve personelin yaklaşımlarının bir doğal sonucu olarak aynı

konudaki belgelerin farklı konu kategorisinde konumlandırılması ve bazı konulardaki veri (belge) yetersizliği bilgisayar ile metin sınıflandırmasını güçleştirmektedir. Bu güçlüklerin üstesinden gelmek için dosya plan numaraları, bir arşivci yaklaşımı ile yeniden gözden geçirilerek sınıflandırılmıştır. Makine öğrenmesinde sınıflandırmanın yapılabilmesi için eğitilmiş veri setinde yeteri miktarda kaydın olması gereklidir. Buradaki kayıtların zenginliği ve tutarlılığı çıkarım yapılacak kayıtların doğruluk oranını artırmaktadır. Bu nedenle veri kümesindeki belgelerin yeniden sınıflandırılması sonucunda bir konuda en az üç belgenin bulundurulması koşulu ortaya konulmuştur. Konusal olarak temsiliyeti üçten aşağı olan gruplardaki belgeler kapsam dışı bırakılmıştır (Şekil-1).



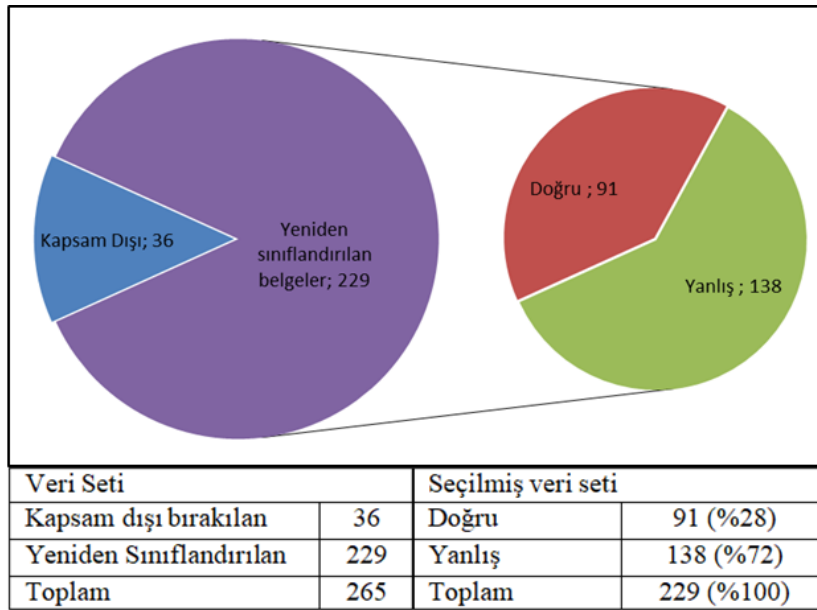
Şekil 1: Belgelerin denetimli makine öğrenmesi ile sınıflandırılması amacıyla uygun verinin hazırlanmasında izlenen yöntem

Çalışmanın ikinci aşamasında ise belgelerin makine öğrenmesi ile sınıflandırılması amacına yönelik olarak veriler işlenerek uygun matrisler elde edilmiş ve bu veriler üzerinde işlemler yürütülmüştür. Belgelerin konu ve gövde kısmı tam metin indeksleme için zengin verileri içermektedir. Temsil etme kabiliyeti en yüksek olan indeks terimlerinin çıkarımı için bu alanlara metin madenciliği teknikleri uygulanmıştır. İlk metin küçük harflere dönüştürülmüş daha sonra noktalama işaretleri, sayılar ve analiz için anlam ifade etmeyen bağlaç ve edat gibi durak kelimeler (stop words) ayıklanmıştır. Farklı kalıplarda yer alan kelimelerin tekil sayılmasına yönelik olarak da kök bulma işlemi uygulanmıştır. Bunun için Türkçe metinler için geliştirilen zemberek kök bulma algoritmasının yeteneğinden yararlanılmıştır. Sözü edilen metin ön işleme tekniklerinin uygulanmasından sonra her kelime için doküman sıklığı (DF) hesaplanarak değeri 3'den aşağı olan kelimeler kapsam dışı bırakılmıştır. Belgelerde geçen kelimeleri ağırlıklandırmak için TF-IDF algoritması çalıştırılmış ve her belgeye değeri en yüksek olan dört anahtar kelime atanmıştır. Tüm bu metin ön işleme ve ağırlıklandırma işlemleri sonucunda 169 belgenin yer aldığı doküman ve kelimelerin oluşturduğu bir veri matrisi elde edilmiştir. Bu veri setinden her konudan oransal olmak koşuluyla tabakalı örnekleme yöntemi kullanılarak belgelerin üçte biri (1/3) bilgisayar aracılığıyla sınıflandırılmak için seçilmiştir. Sınıflandırılmış 112 belge ve sınıflandırmak üzere oluşturulmuş 57 belgeden ibaret bu veri seti üzerinde, makine öğrenmesinde kullanılan ve son zamanlarda bilgi sektöründe popüler olan DVM algoritması çalıştırılmıştır (Şekil-1). Tüm bu teknik işlemler KNIME veri analitiği yazılımı aracılığıyla yapılmıştır.

6. Bulgular

Standart dosya numaralarının belgelere otomatik olarak atanması üzerine yürütülen bu araştırmanın bulguları iki aşamalı olarak değerlendirilmiştir. İlk veri yapısı ve düzeni hakkında bulgular sunulmuş ardından DVM algoritması ile yapılan sınıflamanın sonuçları açıklanmıştır.

Makine öğrenmesinde sınıflandırma işleminde çıkarımların isabet oranının yüksek olması eğitim veri setinin gerçek ve kesin bir sınıflandırma yapısında bulunması ile doğru orantılıdır. Dolayısıyla bu araştırmada belgelerin üretildiği kurumda kurumsal bir belge yönetimi biriminin olmaması nedeniyle, standart dosya plan numaralarının verilmesinde yanlışlıkların olabileceği kaygısı ile belgelerin yeniden sınıflandırılması yoluna gidilmiştir. Nitekim “Yükseköğretim Üst Kuruluşları ve Yükseköğretim Kurumları Saklama Süreli Standart Dosya Planı” kaynağından hareketle yaptığımız yeniden sınıflandırma neticesinde 265 belgeden 36’sı veri setindeki diğer belgeler ile biçimsel olarak benzeşmediğinden (üst yazısız dış yazı, izin formları gibi) kapsam dışı bırakılmış; geriye kalan 229 belgenin 91’ine (%28) doğru, 138 (%72) tanesine ise yanlış dosya numarasının verildiği görülmüştür (Şekil -2).



Şekil 2: Belgelerin standart dosya planı numaralarının doğruluk durumuna göre dağılımı

Sınıflama işlemlerinde yanlış notasyonun atanmasının birçok nedeni olabilmektedir. Yanlış atanan dosya numaraları üzerinde yaptığımız detaylı incelemede 49 belgenin hiyerarşi düzeyi detaylandırılmadığı, geriye kalan 89 belgenin ise farklı dosya numarası aldığı belirlenmiştir. Bu yanlış atamaların muhtemel nedenleri ve yeniden sınıflandırmada tercihimiz maddeler halinde aşağıda verilmiştir.

- Standart dosya planının hiyerarşik ve içerik yapısı yanlış notasyonun verilmesine neden olabilmektedir. Örneğin Şemadaki 105 notasyonu ve alt hiyerarşisi ders programları için ayrılmıştır. Dosya planında bu konuda derin bir hiyerarşi yapıldığı görülmektedir. Ancak kurumdaki yazışmaların genellikle alt hiyerarşideki birkaç konuyu birlikte ele alması karışıklığa neden olmuştur. Bu tip belgelere sadece bir konu numarası verilmiştir.
- Kurumun belge sınıflamada ilke kararlarının olmaması farklı tercihlerin ortaya çıkmasına neden olmuştur. Mesela topluca alınan kurul kararları için 050 notasyonu kullanılırken, kişiye özgü kurul kararları için öğrenci veya personel özlük işlerini temsil eden notasyon kullanılmıştır. Bu durumda kurul kararları yazışmalarında bütünlük bozulmuştur. Yeniden sınıflandırmada tüm kurul kararları için 050 notasyonu kullanılmıştır.
- Dosyalama sisteminde konu başlıklarının sadeliği ve ek açıklamalarının olmaması bir belirsizlik oluşturmaktadır. Örneğin personel hareketliliği Bologna sürecinde mi (109.01) yoksa Öğretim Elemanı Değişim Programları (203.01.01) altında mı yer alacaktır? Benzer bir durum mezun bilgi sistemini içeren yazışmalar için geçerlidir. Bu belgeler mezuniyet işlemleri (302.15) altında mı,

yoksa bilgi sistemleri (700) altında mı dosyalanacaktır? Sınıflandırma yapılırken tercihimiz belgelerin yoğun olarak biriktiği konu numarası üzerinde olmuştur.

- Standart dosya planı şemasındaki alt konu başlıklarının diğer hiyerarşideki alt konu başlıkları ile benzerlik taşıması belgelerin yanlış numara almasına neden olmuştur. Bu durum elektronik belge yönetim sistemlerinde personel davranışlarını yönlendirmektedir. Örneğin “izin” ve “sınavlar” alt başlığı birçok alt hiyerarşide tekrarlanmıştır.
- Standart dosya planında bir konunun birden fazla hiyerarşi altında verilmesi karışıklığa neden olmaktadır. Örneğin öğrenci işleri için burslara 304 notasyonunda, eğitim işleri için burslara ise 772 notasyonunda yer verilmiştir. Burada da tercihimiz belgenin yoğun olduğu notasyon olmuştur.
- Belge içeriklerinin birden fazla konuyu içermesi hangi notasyonun verileceği hususunda bir belirsizlik oluşturmaktadır. Örneğin yüksek lisans ve doktora sınav programları dosya planında ayrıca verilmektedir. Belge içeriğinin her iki konuyu kapsadığı durumlarda hangi dosya numarasının verileceği yaklaşımı resmi yazışma kurallarında açıklanmadığından bu tip belgeler için sadece bir konu numarası tercih edilmiştir.
- Yeni konuların standart dosya planlarında belirlenmemiş olması farklı numaraların atanmasına neden olmuştur. Örneğin akademik teşvik konusu.
- Dosya planında detaylı açıklamaların olmaması konu saptamasını zorlaştırdığı tahmin edilmektedir. Örneğin BAP ile ilgili alınan kararlar, “kurullar ve toplantılar” (050) altın da mı yoksa “proje işleri” (604) altında mı sınıflandırılacağı bir belirsizliktir.
- Kolaycılığa kaçma ve özensizlik belgelerin yanlış numara almasına neden olacak bir diğer sebeptir. Örneğin yüksek lisans mülakat ve jüri üyelikleri (301.10.04) ile ilgili belgelerin eğitim ve öğretimle ilgili diğer işlerde (199) ele alındığı görülmüştür.

Denetimli makine öğrenmesinde sınıfların tahmin edilebilmesi için eğitim verisinde, önceden sınıflanan kayıtlarda, yeteri miktarda verinin bulunması gerekmektedir. Aksi takdirde ilişkilendirme başarısızlıkla sonuçlanır. Bu nedenle veri setindeki belgelerin konusal dağılımına dikkat edilmiştir. Hem sınıflama işleminden önce hem de sınıflama işleminden sonra belgelerin konu ve sayısal dağılımı Tablo 1’de verilmiştir. İşlenmek üzere seçilen 229 belgenin konusal dağılımına bakıldığında sınıflandırma işleminden önce toplamda 96 konuda dağılım gösterdiği anlaşılmaktadır. Sınıflandırma işleminden sonra konusal dağılım 76’ya inmiştir. Veri kümesindeki 46 konudan oluşan 60 belge, denetimli öğrenme için gerekli olan yeterli veriyi taşınamaması nedeniyle kapsam dışı bırakılmıştır. İşlenmek üzere seçilen belgelerin konu ve sayısal dağılımı ise toplamda 30 konuyu kapsayan 169 belge şeklinde olmuştur.

| | Önce | | Sonra | |
|-----------------------|-------------|--------------|-------------|--------------|
| | Konu Sayısı | Belge Sayısı | Konu Sayısı | Belge Sayısı |
| Belge sayısı ≥ 3 | 25 | 140 | 30 | 169 |
| Belge sayısı < 3 | 71 | 89 | 46 | 60 |
| Toplam | 96 | 229 | 76 | 229 |

Tablo 1: Manuel sınıflama işleminden önce ve sonra belgelerin konusal dağılımı

Bu 169 belgeye yöntem bölümünde de belirtildiği gibi uygun parametrelerin tanımlanması için metin madenciliğinde kullanılan bir dizi işlem uygulanmıştır. Metin ön işleme ve ağırlıklandırma işlemleri neticesinde belgeleri temsil eden en uygun anahtar kelimeler seçilmiştir. Denetimli öğrenmede önceden verilen sınıflamadan hareketle yeni kayıtların sonuçlarının tahmin edilmesi için iki veri matrisine ihtiyaç vardır. Bu nedenle belgelerin kayıt, kelimelerin ise öznitelikleri oluşturduğu bit vektörleri oluşturulmuştur. Veri seti yaklaşık 1/3 oranında olmak üzere iki kısma ayrılmıştır. Tablo- 2’de standart dosya planına göre konusal dağılımı verilen bu veri setlerinde bulunan toplamda 112 belgenin oluşturduğu eğitim seti, 57 belgenin oluşturduğu kısım ise öngörü (Prediction), sınıflandırılacak belgeler, olarak ayarlanmıştır.

| Dosya No | 045 | 050.99 | 050.01.04 | 051.01 | 051.02 | 051.04 | 060 | 101.03 | 103 | 104.04 | 105.02.04 | 105.04 | 203.01.01 | 204.01 | 299 | 301.01.02 | 301.10.04 | 302.02.03.02 | 302.04.16 | 302.14 | 310.01.01.01 | 310.02.01.01 | 604 | 700 | 880 | 821.05 | 824.02 | 903.02 | 903.02.02 | 903.05 | toplam |
|------------------|-----|--------|-----------|--------|--------|--------|-----|--------|-----|--------|-----------|--------|-----------|--------|-----|-----------|-----------|--------------|-----------|--------|--------------|--------------|-----|-----|-----|--------|--------|--------|-----------|--------|--------|
| Tüm veri seti | 3 | 3 | 18 | 5 | 6 | 6 | 5 | 3 | 5 | 4 | 3 | 9 | 5 | 3 | 6 | 3 | 7 | 4 | 9 | 6 | 5 | 6 | 7 | 3 | 5 | 6 | 5 | 7 | 3 | 9 | 169 |
| Eğitim veri seti | 2 | 2 | 12 | 3 | 4 | 4 | 3 | 2 | 3 | 3 | 2 | 6 | 3 | 2 | 4 | 2 | 5 | 3 | 6 | 4 | 3 | 4 | 5 | 2 | 3 | 4 | 3 | 5 | 2 | 6 | 112 |
| Öngörü seti | 1 | 1 | 6 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 3 | 2 | 1 | 2 | 1 | 2 | 1 | 3 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 3 | 57 |

Tablo 2: Veri setlerindeki belgelerin standart dosya planındaki konulara göre dağılımı

Bu aşamaya kadar yürütülen tüm bu işlemler makine öğrenmesi için uygun veri setlerini hazırlamak için yapılmıştır. Bundan sonra elde edilen eğitim ve öngörü veri matrisleri ilişkilendirilerek, veri setleri üzerine DVM algoritması çalıştırılmıştır. Bu şekilde eğitim seti matrisinde öğrenilenler ile sınıflandırılmak istenilen kayıtların dosya numaraları tahmin edilmiştir. İşlem sonucunda daha önce manuel olarak yapılan sınıflama ve DVM algoritması ile öngörülen çıkarımları kıyasladığımızda doğruluk isabet oranının % 87,72 (50) olduğu görülmüştür. Diğer bir deyişle belgelerin %12,28'inin (7) dosya numaraları yanlış olarak atanmıştır. Makine tarafından yanlış atanan dosya numaralarının önceki numaralar ile kıyaslandığında ve belge içeriklerinde yaptığımız incelemede yanlış öngörülmesinin nedeni büyük oranda indeks için seçilen kelimelerin benzerlik göstermesidir. Bunun en temel nedeni ise bazı belgelerin tam metin içeriğinin çok kısa olmasıdır. Metinlerin kısa olması, indekisleme için uygun öznelik seçimini engellemektedir. Ayrıca eğitim veri setindeki benzeri belgelerin yeterli miktarda olmaması da yanlış sınıflamaya neden olmuştur. Ancak söz konusu bu olumsuzluklara rağmen %87,72 oranında doğru tahmin, makine öğrenimi ile belgelere standart dosya plan numarası atama işi için yüksek bir başarı olduğu söylenebilir.

7. Sonuç ve Öneriler

Yapay zekâ uygulamalarının gündelik hayatımızda yer edinimi her geçen gün giderek artmaktadır. Güvenlikte, sağlıkta, eğitimde, bankacılıkta, alışverişte, mobil uygulamalarda ve daha birçok yerde uygulama alanına tanıklık etmekteyiz. Etki alanının büyük olduğu yapay zekânın arşiv ve kütüphaneleri kapsamı alanına almaması düşünülemez bir gerçekliktir. Özellikle dijital dermelerin bilgi merkezlerinde giderek artış göstermesi yapay zekâ uygulamaları için zengin bir veri hazinesini oluşturmaktadır. Kütüphane yönetim sistemleri ve elektronik belge yönetim sistemlerindeki hemen hemen tüm aktivitelere yapay zekâ uygulamaları adapte edilebilir. Nitekim belge yönetimi ve arşiv biliminde yeni bir alan olan Hesaplama Arşiv Bilimi (Computational Archival Science) hem yapay zekâ uygulamalarını hem de diğer bilinen yaklaşımları içerisinde barındırması uygulanabilirliğin en somut göstergesidir. Dolayısıyla insan hatasını minimuma indirmek bilgisayar teknolojilerden maksimum seviyede yararlanmak için bilgi profesyonellerinin bu alanda uğraş göstermesi gerekmektedir.

Kütüphane ve bilgi bilimi literatüründe yapay zekâ, makine öğrenmesi, veri ve metin madenciliği uygulamalarının tartışıldığı görülmektedir. Yakın bir gelecekte bu konuların çok daha ayrıntılı olarak araştırılmaya devam edeceği ve konu hakkında tartışmaların yaygınlaşacağı öngörüsünde bulunmak yanlış olmayacaktır. Ancak bu alanda Türkçe literatürde ciddi bir boşluk olduğunu aktarmakta fayda vardır. Bilgi profesyonellerinin, konunun daha çok bilişim alanındaki teknologların alanı olduğu yanlış düşüncesine kapılmaması gerekmektedir. Çünkü olayın modelleme ve algoritmik tasarım olmak üzere iki boyutu vardır. Kanımca konunun tüm boyutlarının ortaya konulması makine öğrenmesi veya yapay zeka için işi tasarlayıp en uygun modeli ortaya koyma işlemi bilgi profesyoneli, bu modeli kodlayıp bilgi yönetim sistemlerine entegre etme işlemi de bilişim camiasının uğraş alanına girmektedir. Bu nedenle bilgi sektöründeki işlemlerin yapay zekâyâ geçiş sürecinde optimizasyonun sağlanması için daha çok uygulama araştırmalarına ihtiyaç vardır ve bilgi profesyonellerinin burada aktif rol almaları gerekmektedir.

Makine öğrenmesi yaklaşımıyla kurumsal belgelere standart dosya plan numaralarının otomatik olarak atanması üzerine yaptığımız bu incelemede ilkin belgeler "Yükseköğretim Üst Kuruluşları ve Yükseköğretim Kurumları Saklama Süreli Standart Dosya Planı" kaynağı referans alınarak sınıflandırılmış, ardından denetimli makine öğrenmesi ile dosya numaraları tahmin etme uygulaması gerçekleştirilmiştir. Bu yöntemle öngörü veri seti üzerinde yaptığımız kıyaslama çalışmasında belgelerin % 87,72 oranında DVM algoritması ile doğru tahmin edilmiştir. Diğer taraftan yanlış notasyonun öngörülmesi kısıtlı içerik (metin), eğitim verisinin azlığı ve farklı konularda benzeşen belge içeriği etki ettiği anlaşılmıştır.

Bu bulgular ışığında DVM'nin belgeler için standart dosya numaralarının tahmininde uygulanabilir bir teknik olduğu sonucuna ulaşılmıştır. Çalışmada elde edilen bulgunun doğruluk oranının yüksek olması aslında büyük oranda yapay zekâ ve algoritmik yaklaşımların arşivlerin derin sorunu olan tasnif ve araştırma süreçlerine katacağı yararlılıkları açıklar niteliktedir. Dolayısıyla bu tür çözümler ile zenginleştirilmiş elektronik belge yönetim sistemlerinin kullanımının faydaları kısaca özetlemek gerekirse şunları söylemek mümkündür.

- Sistemlerin otomatikleşmesi ile daha az işgücüne ihtiyaç duyulacaktır.
- Manuel olarak sürdürülen işlemlerdeki insani hatalar en aza indirgenebilir.
- Kurumlarda iş ve yönetim süreçleri daha isabetli, doğru ve az maliyetli şekilde tamamlanabilir.
- Belgeleri arşivleme süreçlerinde karmaşa önlenir ve bunun sonucunda vatandaşın kamu kurumlarına olan saygınlığı artacaktır.
- Erişilemeyen belgelerin maddi ve manevi yaptırımlarına maruz kalma durumu minimize edilebilir.
- Doğru sınıflandırmalar sonucunda elde edilen farklı verilerle çapraz analizler yapılarak, kurumsal üretkenlik değerlendirmeleri yapılabilir.

Makine öğrenmesinde öngörünün doğru olması sisteme girdi olarak sağlanan verinin kalitesi ile doğru orantılıdır. Belgeleri temsil eden değişken değerlerin heterojen dağılım göstermesi, bir diğer ifade ile konusal dağılımların keskin çizgilerle belirgin olması sınıflamada ayrıştırmayı kolaylaştırır. Bu bakış açısıyla çalışmada eğitim veri seti oluşturan belgelerin dosya numaraları üzerinde ayrı bir inceleme yapılmıştır. İnceleme sonucunda belgelerin büyük bir oranının kurumun elektronik belge yönetim sisteminde personele dayalı nedenlerle yanlış sınıflandırıldığı görülmüştür. Sınıflandırma işleminde olası hataların kurumun sınıflama işleminde ilke kararlarının olmaması, bir belgenin birden çok konuyu içermesi, yeni konuların tanımlanmamış olması, standart dosya planının hiyerarşik ve içerik yapısı gibi nedenlerin yanı sıra personel kaynaklı olduğu değerlendirilmiştir.

Bu çalışmanın az bir veri kümesi üzerinde yürütmüş olması ve belgelerin standart dosya numaralarının sağlıklı olmaması yürüttüğümüz bu inceleme için bir dezavantaj olarak görülebilir. Ancak bu tür veri üzerinde araştırma yürütme ciddi zaman ve emek gerektirdiğinden araştırmanın maliyetini artırdığı durumu göz ardı edilmemelidir. Bu nedenle ileride Türkçe belgeler/dokümanlar üzerinde makine öğrenmesi yaklaşımıyla sınıflama işlemlerinin yürütüleceği benzeri çalışmalarda belge yönetiminde yetkili arşiv birimine sahip kurum belgelerinin tercih edilmesi ve daha büyük veri setlerinin kullanımı, dahası belge eklerini de hesaba katarak araştırmaların yürütülmesi önerilmektedir.

Makine öğrenmesinde denetimli, denetimsiz veya hibrit olmak üzere benzer amaç için geliştirilen birçok algoritma bulunmaktadır. Bu nedenle kütüphane ve arşivlerde yürütülen sınıflama işlemlerinde optimizasyonunun sağlanması için DVM algoritması dışındaki algoritmaların da belge/doküman sınıflandırma kabiliyetlerinin ölçülmesi gerektiği düşünülmektedir.

Kaynakça

- Cıbaroğlu, M. O. ve Yalçınkaya, B. (2019). Belge ve Arşiv Yönetimi Süreçlerinde Büyük Veri Analitiği ve Yapay Zeka Uygulamaları. *Bilgi Yönetimi*. doi:10.33721/by.570634
- Joorabchi, A. ve Mahdi, A. E. (2011). An Unsupervised Approach to Automatic Classification of Scientific Literature Utilizing Bibliographic Metadata. *Journal of Information Science*, 37(5), 499-514. doi:10.1177/0165551511417785
- Özdemirci, F., Torunlar, M. ve Saraç, S. (2009). *Üniversiteler İçin Belge Yönetimi ve Arşiv Sistemi / İşlemleri (BEYAS) El Kitabı*. Ankara: Bayut Tanıtım Matbaacılık.
- Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoukova, T. ve Stuart, K. (2019). More Human Than Human? Artificial Intelligence In The Archive. *Archives and Manuscripts*, 47(2), 179-203. doi:10.1080/01576895.2018.1502088
- Sebastiani, F. (2002). Machine Learning In Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1-47. doi:10.1145/505282.505283
- T.C. Yükseköğretim Kurulu. (2017). Yükseköğretim Üst Kuruluşları ve Yükseköğretim Kurumları Saklama Süreli Standart Dosya Planı. https://www.yok.gov.tr/Documents/Universiteler/Standart_Dosya_Planı.pdf adresinden erişildi.
- TÜBA. (2019). Yapay öğrenme. *Türkçe Bilim Terimleri Sözlüğü*. <http://www.tubaterim.gov.tr/> adresinden erişildi.

- Ullah, A., Khusro, S. ve Ullah, I. (2017). Bibliographic Classification in the Digital Age: Current Trends & Future Directions. *Information Technology and Libraries*, 36(3), 48-77. doi:10.6017/ital.v36i3.8930
- Ünal, M. A. ve Özdemirci, F. (2017). EBYS (e-BEYAS) ve e-Arşiv Sistemlerinde/ Uygulamalarında Yapay Zeka Yaklaşımı. F. Özdemirci ve Z. Akdoğan (Ed.), *Bilgi Sistemleri ve Bilişim Yönetimi Beklentiler ve Yeni Yaklaşımlar* içinde (ss. 57-63). Ankara: BİL-BEM.
- Wang, J. (2009). An extensive study on automated Dewey Decimal Classification. *Journal of the American Society for Information Science and Technology*, 60(11), 2269-2286. doi:10.1002/asi.21147