



DOI: 10.53474/ad.1277080

METİN MADENCİLİĞİ TEKNİĞİNİN OSMANLI ARŞİV BELGELERİ ÜZERİNDE UYGULANMASI: 1 NUMARALI ERZURUM AHKÂM DEFTERİ ÖRNEĞİ¹

Arş. Gör. Murat ESRİNGÜ

Marmara Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi,
Bilgi ve Belge Yönetimi Bölümü
murat.esringu@marmara.edu.tr

 <https://orcid.org/0000-0003-0260-154X>

ÖZ

Teknolojik gelişmelere bağlı olarak veriler daha hızlı ve kolay bir şekilde meydana gelmektedir. Hacmi zamanla büyük miktarlara ulaşan verilerin uygun yöntem ve araçlarla işlenmesi sonucunda ise yeni faydalı bilgiler elde edilmektedir. Bu sebeple veriler gerçekleştirilmesi muhtemel bilimsel araştırmalar için kaynak niteliği taşıyabilmektedir. Farklı yapıya sahip verilerin içerisinde gömülü hâlde bulunan bilgilerin elde edilerek kullanılmasını sağlayan çeşitli yöntemler bulunmaktadır. Metin madenciliği bu yöntemlerden biridir. Metin madenciliği, metin formatında olan veriler üzerinde gerçekleştirilen birtakım işlemler sonucunda yeni bilgiler elde edilmesini mümkün kılmaktadır. Bu çalışmada metin madenciliği tekniğinin oldukça zengin bir veri kaynağı olan Osmanlı arşiv belgeleri üzerinde uygulanabilirliğinin gösterilmesi amaçlanmıştır. Bu bağlamda, çalışma kapsamında ele alınan 1 Numaralı Erzurum Ahkâm Defteri (M.1742-1749) içerisindeki hükümlerin metin madenciliği işlemleri sonucunda konularına göre otomatik olarak sınıflandırılması için KNIME programıyla bir model oluşturulmuştur. Ayrıca çalışmada karşılaşılan kısıtlılıklar doğrultusunda Osmanlı arşiv belgeleri üzerinde gerçekleştirilebilecek metin madenciliği çalışmaları için önerilerde bulunulmuştur.

Anahtar Kelimeler: Ahkâm Defteri, Erzurum Ahkâm Defteri, Metin Madenciliği, KNIME

APPLICATION OF THE TEXT MINING TECHNIQUES ON OTTOMAN ARCHIVE DOCUMENTS: NUMBERED 1 ERZURUM AHKÂM DEFTER EXAMPLE

ABSTRACT

Due to technological advancements, data is being generated faster and more easily than ever before. Over time, the volume of this data has grown to massive amounts. New useful information is obtained by processing the large volumes of data that accumulate over time, using appropriate methods and tools. In addition, data can also serve as a resource for potential scientific research. There are various methods for obtaining and utilizing embedded information in data with different structures. Text mining is one of these methods, allowing for the acquisition of information from data in text format. This study aims to demonstrate the applicability of the text mining technique on the rich data source of Ottoman archive documents. Therefore, a model was created with the KNIME program to automatically classify the provisions in the 1st Erzurum Ahkam Defteri (1742-1749) according to their subjects, resulting from text mining operations. Additionally, recommendations were made for text mining studies that can be conducted on Ottoman archive documents based on limitations encountered during the study.

Anahtar Kelimeler: Erzurum Ahkâm Defteri, Text Mining, KNIME

¹ Bu çalışma, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Bilgi ve Belge Yönetimi Anabilim Dalında gerçekleştirilen “1 Numaralı Erzurum Ahkâm Defterinin (M. 1742-1749) Metin Madenciliği Tekniğiyle Değerlendirilmesi” adlı yüksek lisans tezi kapsamında üretilmiştir.

GİRİŞ

Hayatın normal akışı içerisinde çok farklı aktörler tarafından gerçekleştirilen iş ve işlemler sonucunda veriler meydana gelmektedir. Bu veriler teknolojinin etkisiyle birlikte hızlı ve kolay bir şekilde olduğundan zamanla birer yığın hâline gelmiş durumdadır. İhtiyaç duyulması hâlinde en doğru ve kullanılabilir veriye en hızlı şekilde erişilebilmesi için hacmi büyük miktarlara varan bu verilerin depolanması ve düzenlenmesi gerekmektedir. Veriler, potansiyel bilgi değerleri bakımından mevcut ve gelecekte popüler olması muhtemel bilim dallarında yapılabilecek çalışmalar için birer kaynak niteliği taşıyabilmektedir. Bu çalışmaların temel amacı, ulaşılmak istenilen hedef doğrultusunda uygun veriler üzerinde birtakım işlemlerin yapılması sonucunda yeni bilgilerin elde edilmesidir. Depolanan ve düzenlenen verilerden ilk bakışta bir anlam elde edilmesi mümkün olmayabilir. Veri içerisinde daha önce kimsenin fark edemediği, gömülü hâlde bulunan bilgileri elde edebilmek için bir dizi teknik ve analiz işlemlerinin yapılması gerekebilir. Bu teknik ve analiz işlemlerinin yapılabilmesini mümkün kılan program ve algoritmalar ile veri içerisindeki gömülü bilgi ortaya çıkarılabilmektedir.

Verilerin anlamlandırılmasını ve analiz edilerek daha önce bilinmeyen bilgilerin elde edilmesini mümkün kılan işlemler bulunmaktadır. Bu işlemler veri madenciliği ve metin madenciliğidir. Veri madenciliği, “verilerden belirli ancak bilinmeyen bir sınıfta yer alan nesne veya olayları ifade eden örüntülerin çıkarılması amacıyla belirli algoritmaların uygulanmasıdır” (Şentürk, 2006, s. 3). Metin madenciliği ise daha önce bilinmeyen bilgilerin farklı yazılı kaynaklardan otomatik olarak çıkarılmasıdır (Sukanya ve Biruntha, 2012, s. 269).

Verilerin içerisinde gömülü hâlde bulunan bilginin ortaya çıkarılması ve çıkarılacak bilginin sınırları veri kaynağının niteliği ve niceliğiyle ilişkilidir. Arşivler, sahip oldukları ve araştırmacıların kullanımına sundukları büyük miktardaki nitelikli veri kaynağı bakımından veri merkezi olarak kabul edilebilmektedir. Dolayısıyla Osmanlı İmparatorluğundan günümüze kadar aktarılmış olan büyük miktardaki arşiv materyalini de bünyesinde barındıran Devlet Arşivleri Başkanlığı, araştırmacıların başvurabileceği önemli bir veri kaynağı merkezi olarak görülebilir. Osmanlı Dönemi kurumlarında üretilmiş olan, resmî belge niteliği taşıyan, oluşmaları ve içerikleri bakımından farklılık gösteren arşiv materyalleri veri kaynağı olarak ele alındığında toplumun sosyal, kültürel ve ekonomik yaşamıyla ilgili analiz ve bilgi çıkarımları yapılabilir. Resmî belgelerin kurumsal kültürden, siyasal, sosyal ve tarihsel bağlamlardan bağımsız bir şekilde oluşturulamaması (Özdemirci ve Torunlar, 2015, s. 78) bu durumu destekler niteliktedir. Söz konusu analiz ve bilgi çıkarımlarının resmî belge niteliği taşıyan, yerel mahkemelerce çözüme kavuşturulamayan sorunların çözülmesi için Osmanlı Devletinde halkın Dîvân-ı Hümâyün’a başvurması (Tunç, 2022, s. 4) ve bu sorunların çözümüne dair verilen kararların kaydedilmesi sonucunda oluşan (Günay, 2013(a), s. 9) Ahkâm Defterleri üzerinde yapılması da mümkündür.

Büyük bir çoğunluğu metinsel yapıda olan Osmanlı arşiv belgeleri oluşturuldukları dönemin sosyal, ekonomik, siyasî ve idarî yapısı hakkında önemli birer bilgi kaynağı niteliği taşımaktadır. Söz konusu belgelerin metin madenciliğiyle ele alınması ilk bakışta anlaşılması kolay olmayan gömülü potansiyel bilgilerin ortaya çıkarılmasına imkân tanımaktadır. Metin madenciliğinin Osmanlı Türkçesi ile oluşturulmuş arşiv belgeleri üzerinde uygulanmamış olması sebebiyle bu çalışma “Standart bir metin madenciliği işlemi Osmanlı arşiv belgeleri üzerinde nasıl uygulanabilir?” sorusundan hareketle meydana gelmiştir. Söz konusu belgelere metin madenciliği uygulanırken temel olarak hangi işlem adımlarının kullanıldığını vurgulamak, ayrıca gerçekleştirilen metin madenciliğinin sonucunda hükümlerin otomatik olarak sınıflandırılması işleminden başarılı olarak kabul edilebilecek bir doğruluk oranının elde edilmesi hedeflenmiştir.

1 numaralı Erzurum Ahkâm Defteri bu çalışmanın veri kaynağını oluşturmaktadır. İlgili defterin kaynak olarak seçilmesinde çalışmanın süresini olumsuz anlamda etkilemesi muhtemel oldukça zorlu ve zaman alıcı bir işlem olan transkripsiyonun dört farklı yüksek lisans tezi kapsamında yapılmış olması etkili olmuştur. Ayrıca gerçekleştirilecek olan sınıflandırma işleminde hükümler arasındaki konu bütünlüğünün korunması da bir diğer etkidir. Bu çalışmanın temel amacı veri

kaynağı olarak seçilen 1 numaralı Erzurum Ahkâm Defteri içerisinde bulunan hükümlerin konularına göre başarılı bir şekilde otomatik olarak sınıflandırılabilmesi için model oluşturmaktır. Böylece manuel olarak yapılması hâlinde oldukça zaman alacak sınıflandırma işlemi otomatik olarak gerçekleştirilerek çeşitli analiz ve bilgi çıkarımları yapılabilecektir. Ayrıca Osmanlı arşiv belgelerinden oluşturulmuş veri kaynakları üzerinde yapılacak yeni metin madenciliği çalışmalarına yardımcı olması, bu çalışma süresince karşılaşılan kısıtlılıklar ile yapılan önerilerin yeni çalışmalara katkı sağlaması hedeflenmiştir. Çalışmada veri madenciliği ve metin madenciliği kavramları ele alındıktan sonra Ahkâm Defterleri ve 1 Numaralı Erzurum Ahkâm Defteri üzerinde durulacak, akabinde ise metin madenciliği için oluşturulan sınıflandırma modeli anlatılacaktır.

1. VERİ MADENCİLİĞİ VE METİN MADENCİLİĞİ

Belirli bir düzen içerisinde veri tabanlarında depolanan verilerden faydalı yeni bilgi çıktılarının elde edilebilmesi için zamanla çeşitli araç ve teknikler geliştirilmiştir. Verinin yapılandırılmış, yarı yapılandırılmış ya da yapılandırılmamış şekilde olması, amaca uygun araç ve tekniklerin belirlenmesini etkilemektedir. Verilerin yapılandırılmış olması standart bir veri tabanında satır ve sütunlara uygun olarak düzenlenmiş geleneksel bir yapıda olmalarıyla açıklanmaktadır (Babadağ, 2021, s. 47). Verilerin belirli alanlarını tanımlamak için kullanılan etiketler ve işaretçiler gibi tanımlayıcı öğeleri içeren, yapıyı tanımlayan modele ek veriler ise yarı yapılandırılmış olarak değerlendirilmektedir (Gahi, Guennoun ve Mouftah, 2016, s. 953). Verilerin yapılandırılmamış olmaları ise satır ve sütun formatına düzgün bir şekilde uymayan ve ilişkisel tablolara kolayca indekslenemeyen biçimlerinden kaynaklanmaktadır (Babadağ, 2021, s. 47).

Farklı türlere sahip olan veriler üzerinde analizlerin yapılabilmesi ve gömülü bilgilerin ortaya çıkarılması işlemleri için veri madenciliği ve metin madenciliği araç ve tekniklerinden yararlanılmaktadır. Veri madenciliği, verilerden örtük, önceden bilinmeyen ve potansiyel olarak faydalı bilgilerin çeşitli yöntemlerle elde edilmesi sürecidir (Frawley, Piatetsky- Shapiro ve Matheus, 1992, s. 58) ve büyük, gürültülü ve dağınık veri kümelerinde bulunan veri öğeleri arasındaki zayıf ilişki kalıplarını bulmak için makine öğrenimi algoritmaları kullanılmaktadır (Nisbet, Elder ve Miner, 2009, s. 17).

Metin madenciliği, farklı yazılı kaynaklardan otomatik olarak bilgi ayıklayarak, daha önce bilinmeyen yeni bilgilerin bilgisayar tarafından keşfedilerek (Hearst, 2003, s. 1) metinlerin içerisindeki gizli kalıpların ve özelliklerin ayıklanması amacıyla çeşitli algoritmaların uygulanmasıdır (Kobayashi, Mol, Vrolijk ve Kismihok, 2021, s. 140). Metin madenciliği, belirgin bir formata sahip olmayan, yazı biçimindeki verilerin içerisinde gömülü olan bilginin ortaya çıkarılması ve düzensiz yapıdaki verinin formatlanması sürecidir (Oğuzlar, 2011, s. 7). Dolayısıyla metin madenciliğinin doğal dil ile oluşturulmuş metin biçimindeki veriler üzerinde uygulandığı anlaşılmaktadır. Bunun neticesinde yapılandırılmamış ya da yarı yapılandırılmış verilerden yararlı bilgilerin elde edilebilmesi için metin madenciliği araç ve tekniklerinden faydalanılabilir.

Metin Madenciliğinin Uygulandığı Alanlar

Teknolojinin gelişmesiyle birlikte, sosyal medya hesapları aracılığıyla paylaşılan mesajlar, internet ortamında meydana gelmiş olan haber yazıları ya da elektronik belge yönetim sistemi içerisinde oluşturulan belgeler gibi günlük olarak yapılandırılmış, yarı yapılandırılmış ya da yapılandırılmamış bir biçimde sıklıkla ve kolaylıkla üretilen verilerden daha önce bilinmeyen bilgilerin elde edilmesi isteği artmıştır. Yığın hâline dönüşen verilerden anlamlı sonuçların elde edilebilmesi için analizi mümkün kılan farklı teknik ve araçlar vardır. “Özellikle web ortamında elde edilen metin halindeki büyük verinin analiz edilmesinde metin madenciliği önemli bir istatistiksel tekniktir” (Doğan ve Arslantekin, 2017, s. 71).

Disiplinler arası bir alan olması, doğal dil ile oluşturulmuş verilerle ilgilenmesi ve bu verilerin miktarının hızla artması bakımından metin madenciliği uygulamaları birçok alanda kullanılmaktadır. Genel olarak kullanıldığı alanlar incelendiğinde metin madenciliği uygulamaları;

- “Yayıncılık ve medya
- Bankalar, sigorta ve finans piyasaları
- Telekomünikasyon, enerji ve diğer hizmet sektörleri
- Bilişim sektörü ve internet
- Eczacılık ve sağlık hizmetleri
- Siyasî ve adli kurumlar gibi alanlarda sıklıkla kullanılmaktadır” (Bolasco, Canzonetti, Capo, Rinaldi ve Singh, 2005, s. 36).

Yukarıda bahsedilen metin madenciliğinin sıklıkla uygulandığı alanların, bilişim teknolojileri ve araçlarının gelişmesiyle doğru orantılı bir şekilde çeşitlilik göstermesi olağandır. Miner, Delen, Elder, Fast, Hil ve Nisbet (2012, s. 31), metin madenciliğinin yedi uygulama sahasının olduğunu ve ayrıca altı alanla da ilişkili olduğunu ifade etmiştir. Söz konusu uygulama sahaları bilgi erişim, bilgi çıkarımı, doğal dil işleme, web madenciliği, kavram çıkarımı, doküman kümeleme ve doküman sınıflamadır. İlişkili olduğu alanlar ise veri madenciliği, veri tabanları kütüphane ve bilgi bilimi, hesaplamalı dilbilim, yapay zekâ ve makine öğrenimi ile istatistiktir (Miner ve diğerleri, 2012, s. 31).

Metin madenciliğinin arşivlerde uygulandığı çalışmalar da literatürde mevcuttur. Dong, Yu ve Jiang’ın (2009) acil durum karar desteği sağlamak, politika oluşturmak ve bilgi erişimi kolaylaştırarak bilgi çıkarımını sağlamak amacıyla Çin e-devlet dijital arşivlerinde metin madenciliğinden yararlandıkları çalışma örnek olarak gösterilebilir. Yzaguirre, Smith ve Warren’in (2016) dijital gazete arşivlerinden coğrafi verilerin çıkarılması ve analiz edilmesi için metin madenciliği tekniklerinden faydalandıkları çalışma ise bir diğer örnek olarak verilebilir. Literatürde veri madenciliğinin arşivlerde uygulanmasına yönelik çeşitli çalışmalar bulunmaktadır. Fakat veri madenciliği ve metin madenciliğinin temelde analiz ettiği veri türleri, bu verilere uygulanan teknik yöntemlerin farklılığı ve uygulanma amaçları bakımından farklılık göstermesi sebebiyle söz konusu çalışmalara konu bütünlüğünün bozulmasını önlemek amacıyla değinilmemiştir.

Metin Madenciliği Süreci

Metin madenciliğinde karşılaşılan problemlerin çözümü ya da hedeflenen muhtemel sonuçlara ulaşılması için gerçekleştirilen işlemlerin tamamı bir süreci meydana getirmektedir. Oğuzlar (2011, s. 9-10), metin madenciliği süreçlerinin metin koleksiyonu oluşturma, metin ön işleme, veri madenciliği, değerlendirme ve yorumlama olmak üzere beş adımdan oluştuğunu belirtmektedir. Akyüz ve Gülten (2022, s. 17), metin madenciliği süreçlerini metin ön işleme, metin sayısallaştırılması ve bilgi keşfi olarak üç adımda ele almaktadır. Oğuz, Bilge ve Saka (2007, s. 109-110) ise metin madenciliğinin metin toplama, metin ön işleme, özellik seçme, veri madenciliği, görselleştirme ve değerlendirme olmak üzere altı adımdan oluştuğunu belirtmiştir. Dolayısıyla metin madenciliği için kesin bir süreçten bahsedilemediği sonucuna ulaşılmıştır. Bu duruma, farklı veri kaynakları üzerinde çalışmaların yapılması ya da aynı veri kaynağı üzerinden farklı sonuçlara ulaşmak istenmesi, sebep olarak gösterilebilir. Aşağıda Oğuz ve diğerlerinin (2007, s.109-110) belirttiği süreç adımları açıklanacaktır;

Metin Koleksiyonu Oluşturma

Metin madenciliğinde problemin ne olduğu iyi bir şekilde anlaşılmalı, çalışma yapılacak konu ve veri kaynağı hakkında bilgi sahibi olunmalı ve çalışma sonucunda elde edilmek istenilen sonuçların neler olduğu belirlenmelidir. Bu durum yapılan metin madenciliğinden başarılı çıktılar elde

edilmesini mümkün kılacaktır. Metin koleksiyonu oluşturmak, metin madenciliği süreçlerinin ilkinin oluşturulmasıdır. Bu aşama Oğuz ve diğerleri (2007, s. 109) tarafından, üzerinde çalışma yapılması istenilen konu hakkında, bilgi erişim araçları vasıtasıyla, metin koleksiyonunun oluşturulması şeklinde ifade edilmiştir. Koleksiyon oluşturma olarak adlandırılan bu süreç yalnızca ham verilerin bir araya getirilmesi değil, aynı zamanda söz konusu verilerin anlamlandırılma ve bir bağlama oturtulma sürecidir. Bu süreçte, bir araya getirilen veriler yapılacak çalışmanın amaçları doğrultusunda analiz edilmeye ve yorumlanmaya başlar. Göker ve Tekedere'nin (2007, s.293) de belirttiği gibi, veri toplama aşaması aynı zamanda veriyi anlama ve ona yönelik bir farkındalık geliştirilme sürecidir. Dolayısıyla verilerin toplanması sadece teknik bir işlem olarak görülmemeli, araştırmacının olayları, ilişkileri ve örüntüleri daha derinlemesine kavrayabileceği bir öğrenme ve keşfetme sürecinin başlangıcı olarak değerlendirilmelidir.

Metin Ön İşleme

Metin ön işleme aşaması, üzerinde çalışma yapılacak veri kaynağına ve ulaşılmak istenilen sonuçlara göre değişiklik gösterebilmektedir. “Bu aşamada veri üzerinde yapılacak işlemlere ve bu işlemlerin hangi yöntemlerle yapılacağına karar verilir” (Şeker, 2008, s. 12). Yapılmasına karar verilen işlem ve yöntemlerin detaylı bir şekilde ve eksiksiz olarak yapılması metin madenciliğinin başarıyla uygulanmasına, dolayısıyla da elde edilecek sonucun başarı oranına etki edebilecektir. Veri ön işleme aşamaları bir projenin %80'ini ya da %90'ını kapsamaktadır (Miner ve diğerleri, 2012, s. 41) ve dolayısıyla bu aşama oldukça zaman alıcıdır (İlhan, 2001, s. 35). Literatürde metin madenciliğiyle ilgili çalışmalar incelendiğinde farklı ön işleme adımlarının olduğu gözlemlenmiştir. Genel olarak ön işleme adımlarına Kaşıkçı ve Gökçen'in (2014, s. 26) çalışması örnek gösterilebilir. Söz konusu çalışmada ön işleme adımları dört başlık altında ele alınmıştır. Bu adımlar; dönüştürme, tarama ve işaretleme, durak kelimelerin çıkarılması ve kök bulmadır.

Özellik Seçimi

“Özellik seçimi (Feature Selection), büyük boyutlu veri kümesini, daha küçük boyutta temsil edebilecek nitelikte olan bir alt küme olarak tanımlanmaktadır” (Haltaş, Alkan ve Karabulut, 2015, s. 420). Ön işlem aşamasından geçen metin verileri içerisindeki her kelime çalışmanın sonucuna anlamlı bir katkı sağlamayabilir. Bu aşama ile hedeflenen metin verileri içerisinde bulunan ve önemli olarak kabul edilebilecek kelimelerin belirlenmesidir. “Özellik seçimi, ilgilenilen problem için en faydalı ve en önemli özellikleri seçerek veri kümesindeki özellik sayısının azaltılmasını amaçlamaktadır” (Budak, 2018, s. 22).

Veri Madenciliği

Yarı yapılandırılmış ya da yapılandırılmamış formdaki veriler, üzerlerinde yukarıda bahsedilen adımların uygulanması sonucunda daha düzenli ve analiz edilebilir, diğer bir ifadeyle yapılandırılmış forma dönüştürülmektedir. Bu aşamada, karar ağaçları, yapay sinir ağları ve kümeleme gibi veri madenciliği teknikleriyle metinsel içeriklerin incelenmesi sağlanır. Oğuz ve diğerlerinin (2007, s. 110) de vurguladığı üzere bu aşamada içeriklerin daha sistematik bir şekilde analiz edilmesine ve anlamlı sonuçlar elde edilmesine katkıda bulunur.

Görselleştirme

Yukarıda bahsedilen aşamaların veri kümesine uygulanması ile çok sayıda sonuç elde edilebilmektedir. Bu sonuçların etkin bir şekilde sunulmasını ve daha anlaşılabilir olmasını mümkün kılmak için yapılan görselleştirme aşamasıdır (Oğuz ve diğerleri, 2007, s. 110). Sonuçları

inceleyecek olan kullanıcıların ihtiyaçlarına, zaman metaforuna ve gerçekleştirilen çalışmanın amacına uygun olacak şekilde farklı görselleştirme teknikleri bulunmaktadır (Binici, 2016, s. 35).

Değerlendirme

Veri üzerinde gerçekleştirilen aşamaların sonucunda ortaya çıkmış olan örüntülerin (pattern) değerlendirildiği ve bilginin elde edildiği aşama olarak kabul edilebilir (Şeker, 2018, s. 16). Bu aşamada hedeflenen bilgiye ulaşılmamış ya da istenilen düzeyde bilgi çıkarımının gerçekleştirilememiş olması durumunda yukarıdaki süreç yeniden uygulanabilir. Bu aşama ile süreç adımlarının uygulanma başarısı da gözlemlenmiştir.

2. AHKÂM DEFTERLERİ VE 1 NUMARALI ERZURUM AHKÂM DEFTERİ

Ahkâm, kelime anlamı olarak “emirler, hükümler” anlamına gelmektedir ve hüküm kelimesinin çoğuludur (Devellioğlu, 2008, s. 24). Dolayısıyla öncelikle hüküm kavramının tanımlanması, ahkâm kavramının anlaşılmasını kolaylaştıracaktır. Hüküm, bir konu hakkında hâkimin vermiş olduğu karar, iki nesne ya da fikir arasında kurulan bağlantı ve bir konu hakkında verilen mantıksal karar anlamlarında kullanılabilen geniş kapsamlı bir kavramdır (Özel, 1988, s. 550). “İyileştirmek amacıyla menetmek, düzeltmek; karar vermek” manalarında kullanılan, ayrıca “ilim, derin anlayış; siyasî hâkimiyet, karar ve yargı” anlamlarına da gelen ve kelâm, mantık ve fıkıh ilimlerinde de kullanılan bir terimdir (Üzüm, 1998, s. 464). Sözlüklerde kat’î emir ve karar anlamına geldiği ifade edilerek, “her hangi bir iş ve vazife için pâdişah tarafından verilen tahrirî emre Hüküm-i hümayun denilmektedir; divan-ı hümayundan ve maliyeden, darphane ve saireden hükümdar namına yazılan emir ve tenbihâtın hepsi hükümdür” (Uzunçarşılı, 1998, s. 280). Hüküm kavramı ile ilgili yapılan tanımlar doğrultusunda ahkâm kavramı şu şekilde tanımlanabilir; Ahkâm kavramı daha geniş bir anlamda, belirli bir konuyla ilgili koyulan kuralların tamamını ifade etmek için kullanılmaktadır (Özel, 1988, s. 550). Ahkâm defteri ise Osmanlı Devleti’nde kanunnamelerle belirlenen ve genellikle malî konulara ilişkin hüküm ve nizam niteliğindeki kararların toplandığı defterdir (Sahillioğlu, 1988, s. 551; Pakalın, 2004, s. 30).

1326-1359 yılları arasını kapsayan Orhan Gazi Döneminden itibaren var olan, müessese hâline gelmesinin Fatih Sultan Mehmed (1451-1481) döneminde gerçekleştiği kabul edilen ve sahip olduğu yetkilerle Osmanlı İmparatorluğunun klasik çağını kapsayan dönemde padişahın sonraki en yetkili makamı Dîvân-ı Hümayundur (Sarıcaoğlu, 2007, s. 87). Dîvân-ı Hümayun, devletin önemli makamlarındaki kişilerin katılımıyla padişah adına karar verebildikleri, devletin idarî, siyasî ve adlî konularının görüşüldüğü, herhangi bir ayırım gözetilmeksizin toplumun her kesiminden insanın başvurabildiği ve şikâyetlerini iletebildiği en önemli karar organıdır (Mumcu, 1994, s. 430-431). Bu karar organının en temel görevi ise adaleti sağlamaktır ve buradaki adalet kavramı, halkın padişaha şikâyetlerini iletebilmesi ve padişahın verdiği emirler doğrultusunda haksızlıkların çözülmesidir (İnalçık, 1988, s. 33). Dîvân-ı Hümayunda görüşülen konular hakkında verilen kararların tarihleri esas alınarak bir düzen dâhilinde kaydedildiği defterlere Mühimme Defteri denilmektedir (Yıldırım, Atik, Cebecioğlu, Demirbaş, Karazeybek, Safi, Serin, Uslu, Yekeler ve Yıldırım, 1994, s. ix-x). Ferman kayıtlarının tamamı Hicri 1059/Miladi 1649 yılından itibaren mühimme defterleri yerine, yeni bir defter türü olarak düzenlenen şikâyet defterlerine kaydedilmiştir (Şimşir, 1994, s. 359). 1742 yılından itibaren ise şikâyet defterlerine kaydedilen, dilek ya da şikâyet niteliğindeki hükümler eyaletlere göre ayrılarak kaydedilmiştir ve dolayısıyla ahkâm-ı şikâyet veya ahkâm defterleri adı verilen yeni bir seri oluşmuştur (Temelkuran, 1975, s. 156). Eyaletlere göre kaydetme sistemi (mahalli sistem) Râgıp Efendi tarafından geliştirilmiştir ve bu sistemin geliştirilmesindeki temel sebep, meydana geldiği yerde çözülemeyen problemlerin merkeze iletilmesinde yaşanan artış ile bürokratik işlemlerde oluşan yoğunluğu azaltmaktır (Emecen, 2005, s. 125).

Ahkâm defterleri vasıtasıyla defterin temsil ettiği bölgenin sosyal ve ekonomik yapısına dair bilgilere ulaşmak mümkündür. Ayrıca yöneticilerin kendi arasında ya da halk ile yaşamış olduğu sorunların neler olduğu hakkında bilgilere de erişilebilmektedir (Emecen, 2005, s. 125). Bunların yanı sıra ahkâm defterleri söz konusu sorunlara getirilen çözümlerin neler olduğu, hukuk kurallarının pratikte ne derece uygulandığı ve uygulanabilir oldukları, hukuki açıdan bir eksikliğin var olma durumu ve varsa bu eksiklik sonucunda meydana gelen problemlerin neler olduğu gibi konularda da bilgi sunabilmektedir (Öztürk, 2001, s. 621-622). Osmanlı tarihinde ahkâm defterlerinin tutulduğu eyaletler ve söz konusu eyaletlerde oluşan defter sayıları şu şekildedir;

Tablo 1. Ahkâm defterlerinin tutulduğu eyaletler ve defter sayısı

Adana: 9	Diyarbakır: 9	Karaman: 39	Rumeli: 85
Anadolu: 185	Erzurum: 19	Maraş: 6	Sivas: 36
Bosna: 9	Halep: 9	Mora: 21	Şâm-ı Şerîf: 9
Cezair ve Rakka: 25	İstanbul: 26	Özi ve Silistre: 49	Trabzon: 8

Kaynak: Başbakanlık Osmanlı Arşivi Rehberi, 2017, s. 30-41.

1 Numaralı Erzurum Ahkâm Defteri

Çalışmanın veri kaynağı olarak ele alınan 1 Numaralı Erzurum Ahkâm Defteri, toplam 293 sayfadan oluşmaktadır ve defterin baştan bir sayfası, sondan ise iki sayfası boştur. Defterin kapağında “Erzurum Defteri min evâil-i Cemâziye’l-âhire sene 1155 ilâ evâhir-i Şevval sene 1162”² şeklinde künye ve tarih bilgisi yer almaktadır.

1 Numaralı Erzurum Ahkâm Defterini meydana getiren 1212 hüküm (3 adet iptal edilmiş) ve 4 derkenâr olmak üzere toplam 1216 kayıt bulunmaktadır. Bu hükümlerin aylara göre sayısal dağılımı ise şu şekildedir; Muharrem ayında 128 hüküm, Safer ayında 124 hüküm, Rebiülevvel ayında 126 hüküm, Rebiülâhir ayında 135 hüküm, Cemaziyelevvel ayında 130 hüküm, Cemaziyelâhir ayında 120 hüküm, Receb ayında 98 hüküm, Şaban ayında 88 hüküm, Ramazan ayında 50 hüküm, Şevval ayında 72 hüküm, Zilkade ayında 79 hüküm ve Zilhicce ayında 62 hüküm yazılmıştır (Kul, 2011, s. 23-46; Çiçek, 2011, s. 28-55; Susam, 2019, s. 19-79; Kantaroğlu, 2019, s. 26-83). Defterdeki hükümler incelendiğinde hükümlerin bazılarının tek bir yöneticiye, bazılarının ise birden çok yöneticiye yazıldığı gözlemlenmiştir. Ayrıca birden çok yöneticiye yazılan hükümlerde yöneticilerin hiyerarşik bir şekilde yazıldığı görülmektedir. Hükümlerin birden çok yöneticiye yazılmasının sebepleri olarak şu çıkarımlar yapılabilir;

- Hüküm içerisinde geçen konu farklı yöneticileri de ilgilendirebilir. Bu sebeple hüküm ilgili yöneticilere de gönderilebilir.
- Hükme konu olan problemin yaşanmaması ya da tekrarlanmaması adına yerel yönetimlerin önlem almasını ve söz konusu problemin yaşanması durumunda yerel yönetimlerin verecekleri kararlar için yol gösterici olmasını sağlamak amacıyla farklı yöneticilere iletilebilir.

² Evâil-i Cemâziyelâhir 1155-Evâhir-i Şevvâl 1162/Ağustos 1742-Eylül-Ekim 1749.

- Eyaletlerde halkın yöneticilerle ya da nüfuz sahibi kişilerle problem yaşayabilmesi muhtemeldir. Problemin çözümüne dair verilen kararın adil bir şekilde uygulanabilmesi için konunun takibinin yapılmasını mümkün kılmak ve tarafların mağduriyet yaşamamasının önüne geçilmesini sağlamak için farklı yöneticilere ulaştırılabilir.

“Divanî kırma”³ yazı biçimiyle kaleme alınan defter (Kantaroglu, 2019, s. 16), kronolojik düzen göz önünde bulundurularak kaydedilmeye çalışılmıştır. Fakat bazı sayfaların alt kısımlarında bulunan boşluklara sonradan ekleme yapılmasından dolayı bu durumun tarihsel bütünlüğü bozduğu görülmüştür. Defterde yanlışlıkla tekrar yazılan hükümlerin olduğu ve bazı hükümlerin iptal edildiği durumlar da mevcuttur. 1 Numaralı Erzurum Ahkâm Defterini oluşturan hükümler incelendiğinde üç bölümden meydana geldikleri ifade edilmektedir (Çiçek, 2011, s. 19-20):

- Makamı ya da ismi yazılan hükmün muhatabı şer’i ya da örfi yöneticisine hitap kısmı birinci bölümü oluşturmaktadır.
- Hükmün yazılmasını gerektiren olaylar/durumlar hakkında bilgi verilen kısım ikinci bölümü meydana getirmektedir.
- İkinci bölümde anlatılan olaylar/durumlar hakkında verilen kararın yer aldığı kısım ise üçüncü bölüm olarak adlandırılmaktadır.

Osmanlı Türkçesi ile yazılan 1 Numaralı Erzurum Ahkâm Defteri dört farklı yüksek lisans tez çalışması sonucunda transkript edilmiştir. Buna göre 1 Numaralı Erzurum Ahkâm Defterinin 1 ile 65. sayfa aralığı Muhittin Kul, 65 ile 130. sayfa aralığı Hikmet Çiçek, 131 ile 212. sayfa aralığı Feyzanur Susam ve 212 ile 293. sayfa aralığı ise Nazmiye Gül Kantaroglu tarafından tez çalışması olarak incelenmiştir. Hükümlerin transkripsiyonlarının yanı sıra farklı konu başlıklarının belirlenmesiyle hükümlerin içerikleri hakkında bilgi de verilmek istenmiştir. Fakat belirli bir standarda bağlı kalınmaksızın yapılan konu belirleme işlemi, benzer içeriklere sahip hükümlerin farklı başlıklarla ifade edilmesine sebep olmuştur. Bu sebeple 1 Numaralı Erzurum Ahkâm Defteri içerisinde bulunan hükümler bizzat incelenerek konu başlıkları belirlenmiştir. Buna göre defterde geçen hükümlerin konu dağılımları şu şekilde ifade edilebilir;

- Asayiş ile ilgili 268 hüküm,
- Sosyal, ekonomik ve idarî konularla ilgili 432 hüküm,
- Arazi meselesi ve vergi ile ilgili ise 509 hüküm ve ayrıca 4 derkenâr ile 3 adet iptal edilen hüküm bulunmaktadır.

3. 1 NUMARALI ERZURUM AHKÂM DEFTERİ ÜZERİNDE METİN MADENCİLİĞİ UYGULAMASI

1 Numaralı Erzurum Ahkâm Defteri üzerinde yapılan metin madenciliği uygulaması için Microsoft Excel ve Konstanz Information Miner (KNIME) programları kullanılmıştır. Microsoft Excel bir ve daha fazla çalışma sayfasına sahip olabilen, erişilmesi kolay ve çok sayıda hesaplama yapılabilmesini, formül yazılabilmesini ve grafik çizilmesini mümkün kılan bir programdır (Öztek, 2021; aktaran Kutluca, 2013, s. 43-44). KNIME ise hacmi büyük miktardaki ve farklı türdeki veriler ile çalışmasını sağlayan programdır. KNIME programı, açık kaynak kodlu olması, anlaşılması kolay ve tek bir ara yüze sahip olması, içerisinde hazır bir şekilde bulunan düğümler sayesinde kod yazma gereksinimini minimuma indirmesi ve her bir düğüm hakkında bilgi vermesi, ayrıca oluşturulan modellerin erişilebilir ve geliştirilebilir olmasını mümkün kılması gibi durumlar

³ Sadece divanda yazıldığı için divana mahsus anlamını taşıyan Divanî yazı çeşidinin bir türüdür. Divanî yazının hızlı bir şekilde yazılması sırasında bazı kurallara uyulmaması sonucunda ortaya çıkmıştır (Alparslan, 1994, s. 445).

dolayısıyla bu çalışmada kullanılmıştır. Microsoft Excel ile formatları farklı olan veriler bir araya getirilerek bir veri tabanı oluşturulmuştur. Sonrasında Excel programı içerisinde bulunan veriler KNIME programı vasıtasıyla bir dizi ön işlem safhalarından geçirilmiştir. Yapılandırılmış hâle dönüştürülen veriler uygun algoritmalar ile sınıflandırılmıştır ve bu sınıflandırma işleminin başarı oranları değerlendirilmiştir. Bu çalışmanın aşamaları literatürdeki metin madenciliği süreçleri ve kullanılan metinsel hüküm verileri göz önünde bulundurularak oluşturulmuştur ve aşamalar şu şekildedir;

Veri Toplama Süreci

293 sayfa, 1212 hüküm ve 4 derkenârdan oluşan 1 Numaralı Erzurum Ahkâm Defteri, Arap harfleri ile Farsça ve Türkçe sesleri karşılayan p, ç, j, g ve nazal n harfleriyle de yazılmıştır. Öncelikle bu harfler ile oluşturulmuş olan metinlerin Latin harflerine dönüştürülmesi gerekmektedir. Söz konusu defterin transkript edilmesi oldukça zaman alıcı bir süreçtir ve bu çalışmanın süresini uzatacağı öngörülmüştür.

Yukarıda bahsi geçen dört tezden elde edilen transkript edilmiş, konuları belirlenmiş ve özetleri çıkarılmış hüküm verileri Excel'e aktarılmıştır. Excel'e aktarılan hüküm verileri içerisinde bulunan 4 adet derkenâr ve 3 adet iptal edilmiş hüküm, çalışmanın sonucuna katkı sağlamayacağına öngörülmesi sebebiyle kullanılmamıştır. Dolayısıyla bu çalışma 1209 hüküm üzerinde gerçekleştirilmiştir.

Verilerin Hazırlanması

Eyalet sistemine göre kayıtların tutulmasıyla oluşan ahkâm defterleri, temsil ettikleri eyaletin, devletin ve dönemin sosyal, ekonomik, siyasî ve idarî yapısı hakkında bilgi vermektedir. Fakat bazen bu bilgiler ilk bakıldığında anlaşılır bir şekilde olmayabilir. Bu sebeple metin hâlinde olan bu hüküm verilerinin anlaşılabilir olmasını, dolayısıyla analiz edilebilir bir hâle dönüştürülmesini ve içerisinde var olan gömülü bilgilerin ortaya çıkarılmasını mümkün kılan metin madenciliğinden faydalanılabilir. Hüküm verilerinden yeni bilgilerin elde edilebilmesini muhtemel kılan işlemlerden birisi sınıflandırmadır. Yüksek hacimli veri kümelerinde manuel olarak gerçekleştirilen sınıflandırma işlemleri oldukça zaman alıcı ve hata payının yüksek olabileceği bir süreçtir. Otomatik olarak gerçekleştirilen sınıflandırma işlemi sonucunda hüküm verileri üzerinde çeşitli istatistiksel analizler yapılarak yeni bilgiler elde edilebilir. Ayrıca var olan bir bilgi için kanıt niteliği taşıyan sonuçlara ulaşılması da mümkündür.

Çalışmada kullanılan hüküm verilerinin konuları dört farklı kişi tarafından belirlenmiştir. Bu durum benzer konulardaki hükümlerin farklı konu başlıkları ile ifade edilmesine neden olmuştur. Kantaroğlu (2019, s. 21), bazı hükümlerin konularını belirlerken daha geniş bir anlam ifade eden güvenlik kelimesini başlık olarak kullanmıştır. Fakat Susam (2019, s. 19) çalışmasında Kantaroğlu'nun güvenlik başlığı altında topladığı hükümlerle oldukça benzer içeriğe sahip olan hükümleri şekavet başlığı altında bir araya getirmiştir. Bu durum doğrultusunda, konu başlıklarının belirlenmesinde belirli bir standardın olmadığı sonucuna ulaşılmaktadır. Hükümlerin konu başlıklarının belirli bir standarda göre belirlenmesi ve dolayısıyla gerçekleştirilecek olan sınıflandırma işleminden başarılı sonuçların elde edilebilmesi amacıyla hüküm konuları yeniden, tümevarım yaklaşımıyla ele alınmıştır. Hükümleri temsil eden her bir özel başlık daha genel başlıklar altında toplanmıştır ve nihai olarak tüm hükümler 3 ana konu başlığı altında toplanmışlardır. Bu durumun nedeni ise şu şekilde ifade edilebilir; sınıflandırma işleminde kullanılacak algoritmalar tarafından hüküm verileri eğitim verisi ve test verisi olmak üzere ayrılır. Eğitim verileri üzerinde algoritmalar konu başlıklarının altında hükümleri temsil edebilecek kelimeleri ve konu başlıklarını inceler. Ardından test verilerini oluşturan hükümleri temsil edebilecek kelimeleri bularak bunların eğitim verilerinden elde edilen bilgiler doğrultusunda konu atamasını gerçekleştirir. Eğitim için ayrılan veri miktarının fazla olması algoritmanın test verisi üzerinde yapacağı sınıflandırma işleminin

başarı oranını artıracaktır. Özel kapsamlı konu başlıklarının seçilmesi durumunda eğitim verisinin yeterli olmayacağı, dolayısıyla sınıflandırma işleminin başarı oranının düşebileceği öngörülmüştür. Bu sebeple daha genel konu başlıkları seçilmiştir. Oluşturulan 3 ana konu başlığı ve bu başlıkların altında bir araya getirilen alt konu başlıkları Tablo 2’de verilmektedir.

Tablo 2. 1 Numaralı Erzurum Ahkâm Defteri’ndeki Hükümlerin Konulara Göre Detaylı Dağılımı

1. ASAYİŞ	2. SOSYAL, EKONOMİK VE İDARİ KONULAR	3.ARAZİ MESELESİ VE VERGİ
<p><u>1.1. Cana ve Mala Müdahale ve Tecavüz (218)</u></p> <p>1.1.1. Eşkyalık (142)</p> <p>1.1.2. Mal Gaspsı (35)</p> <p>1.1.3. Para Gaspsı (29)</p> <p>1.1.4. Tecavüz (4)</p> <p>1.1.5. Darp (2)</p> <p>1.1.6. Hırsızlık (6)</p> <p>1.2. Fuhuş (1)</p> <p>1.3. İftira (11)</p> <p>1.4. Cinayet (24)</p> <p>1.5. Güvenliğin Sağlanması (5)</p> <p>1.5.1. Can Güvenliğinin Sağlanması (2)</p> <p>1.5.2. Resmî Evrak ve Değerli Malların Korunması (1)</p> <p>1.5.3. Şahıs Güvenliğinin Sağlanması (2)</p> <p>1.6. Diyet Talebi (8)</p> <p>1.7. Zimmete Mal Geçirme (1)</p>	<p><u>2.1. Vakıf (76)</u></p> <p>2.1.1. Vakıf Tevcihi (2)</p> <p>2.1.2. Vakıf Yönetimi Tevcihi (3)</p> <p>2.1.3. Vakfa Müdahale (71)</p> <p><u>2.2. Borç Meselesi (162)</u></p> <p><u>2.3. Anlaşmazlık (34)</u></p> <p>2.3.1. Mülk Anlaşmazlığı (4)</p> <p>2.3.2. Satış Anlaşmazlığı (10)</p> <p>2.3.3. Su Anlaşmazlığı (20)</p> <p><u>2.4. Gündelik Yardımı (1)</u></p> <p><u>2.5. Göç (14)</u></p> <p><u>2.6. Faiz (4)</u></p> <p><u>2.7. Haksız Para Talebi (6)</u></p> <p><u>2.8. Haksız Mal Talebi (2)</u></p> <p><u>2.9. İhtida (2)</u></p> <p><u>2.10. Mehir (1)</u></p> <p><u>2.11. Miras (74)</u></p> <p>2.11.1. Haksız Hisse Talebi (4)</p> <p>2.11.2. Miras İhtilafı (13)</p> <p>2.11.3. Tereke Taksimi (16)</p> <p>2.11.4. Mirasa Müdahale (41)</p> <p><u>2.12. Nikâh Akdi (9)</u></p> <p><u>2.13. Nafaka (1)</u></p> <p><u>2.14. Berat Tevcihi (2)</u></p> <p>2.14.1. Sâ’ilik Beratu (1)</p> <p>2.14.2. Müderrislik Beratu (1)</p> <p><u>2.15. Memuriyet (42)</u></p> <p>2.15.1. Sorumluluklarını Yapmama (2)</p> <p>2.15.2. Göreve Müdahale (16)</p> <p>2.15.3. Vazifeyi Kötüye Kullanma (18)</p> <p>2.15.4. Soruşturma (2)</p> <p>2.15.5. Vazife Tevcihi (3)</p> <p>2.15.6. Tayin Talebi (1)</p> <p><u>2.16. Yol Kapama (1)</u></p> <p><u>2.17. Tedarik (1)</u></p>	<p><u>3.1. Arazi ve Sınır Meselesi (112)</u></p> <p>3.1.1. Mülke Müdahale (94)</p> <p>3.1.2. Sınır İhtilafı (1)</p> <p>3.1.3. Meraya Müdahale (8)</p> <p>3.1.4. Yurtluğa ve Ocaklığa Müdahale (1)</p> <p>3.1.5. Koruya Müdahale (2)</p> <p>3.1.6. Yaylağa Müdahale (6)</p> <p><u>3.2. Zeamet (73)</u></p> <p>3.2.1. Zeamete Müdahale (66)</p> <p>3.2.2. Zeamet Tahsili (5)</p> <p>3.2.3. Zeamet Mutasarrıflığı Tevcihi (2)</p> <p><u>3.3. Malikâne (18)</u></p> <p>3.3.1. Malikâneye Müdahale (16)</p> <p>3.3.2. Malikâne Tahsili (2)</p> <p><u>3.4. Timar (126)</u></p> <p>3.4.1. Timar Tevcihi (19)</p> <p>3.4.2. Timara Müdahale (100)</p> <p>3.4.3. Timar Mutasarrıflığı (1)</p> <p>3.4.4. Timar Tahsili (4)</p> <p>3.4.5. Timar Anlaşmazlığı (2)</p> <p><u>3.5. Öşr (11)</u></p> <p>3.5.1. Öşr Tahsili (4)</p> <p>3.5.2. Öşr’e Müdahale (7)</p> <p><u>3.6. Haksız Vergi Talebi (129)</u></p> <p><u>3.7. Vergi Vermeme (32)</u></p> <p><u>3.8. Cebellü Bedeliyesi (7)</u></p> <p><u>3.9. Ulûfeye Müdahale (1)</u></p>

Veri Ön İşleme

Microsoft Excel programında bir araya getirilerek oluşturulan veri kümesi incelendiğinde yarı yapılandırılmış metinsel veri oldukları belirlenmiştir. Bu veriler üzerinde analizlerin yapılabilmesi ve bilgi çıkarımı işlemlerinin gerçekleştirilebilmesi için yarı yapılandırılmış veri kümesinin yapılandırılmış veri hâline dönüştürülmesi gerekmektedir. Bu sebeple çalışmanın verileri üzerinde KNIME programı vasıtasıyla bir dizi ön işlem uygulanmıştır. Veriler üzerinde gerçekleştirilen ön işlemler, verilerin analizler yapılabilir bir hâle dönüştürülmesini ve bu analizlerin daha başarılı

sonuçlar verebilmesini sağlamak için oldukça önemlidir (Emre ve Selçukcan Erol, 2007, s. 165). Hüküm verileri üzerinde gerçekleştirilen ön işlem adımları ile bu adımların uygulanması mümkün kılan KNIME düğümleri şu şekildedir;

- a) *Excel Reader (Excel Okuyucu)*: Farklı dosya türleri ile çalışma imkânı sağlayan KNIME programı içerisinde bulunan bu düğüm vasıtasıyla Excel programında bir araya getirilmiş olan veriler KNIME programına aktarılmıştır.
- b) *Strings To Document (Doküman Dizinleri)*: Excel programının içerisinde bulunan hüküm numarası, hükmün tarihi, konusu, özeti ve hüküm metni gibi satır ve sütunlardaki her bir veri, Strings To Document düğümü ile dokümana dönüştürülmüştür. Böylece her bir veri, üzerinde işlem yapılabilir hâle getirilmiştir.
- c) *Case Converter (Büyük-Küçük Harf Dönüştürücü)*: Üzerinde işlem yapılacak olan hüküm verilerini oluşturan tüm harfler bu düğüm ile küçük harfe dönüştürülmüştür. Bu durumun temel sebebi ise algoritmaların aynı kelime olmasına rağmen içerisinde büyük harf bulunan ve bulunmayan kelimeleri farklı birer karakter olarak kabul etmesidir. Örneğin “Vakıf” ile “vakıf” kelimeleri aynı olmalarına rağmen baş harfi büyük olan kelime farklı bir karakter olarak, baş harfi küçük olan kelime ise farklı bir karakter olarak kabul edilecek ve bu kelimeler için ayrı ayrı frekans hesaplaması yapılacaktır. Bahsedilen bu durumun engellenmesi için tüm büyük harflerin küçük harflere dönüştürülmesi işlemi gerçekleştirilmiştir.
- d) *Punctuation Erasure (Noktalama İşaretlerini Silme)*: Bu düğüm vasıtasıyla metin verileri içerisinde kullanılan noktalama işaretleri kaldırılmıştır. Noktalama işaretleri ve bu işaretlerle kullanılan kelimeler için ayrı frekans değeri hesaplanmasının engellenmesi adına bu işlem gerçekleştirilmiştir.
- e) *Zemberek Stemmer (Zemberek Kök Bulucu)*: Zemberek, Türkçe için geliştirilen ve açık kaynak kodlu olan doğal dil işleme kütüphanesidir. Zemberek Kütüphanesi, Türkçe bir metni meydana getiren kelimeleri eklerinden ayırarak bu kelimelerin kökünü bulmaya yarayan bir algoritma şeklinde ifade edilebilir ve yapılan bu işleme ise gövdeleme (kök bulma) denilmektedir. Bu çalışmada kullanılan veriler Osmanlı Türkçesi ile yazılan hükümlerdir ve bu sebeple Arapça ve Farsça kelimelerle tamlamalar bulunmaktadır. Bu kelimeleri köküne ayırabilecek Osmanlı Türkçesi için bir kütüphane mevcut değildir. Zemberek Kütüphanesinin bu çalışmada kullanılmasının sebebi hükümlerin içerisinde var olan Türkçe kelimelerin köklerine ayrılmasını sağlamaktır. Zemberek Kütüphanesinin köklerine ayıramadığı diğer kelimeler için kök bulma işlemi manuel olarak yapılmıştır.
- f) *Number Filter (Sayı Filtresi)*: Veri kümesini oluşturan hüküm metinleri içerisinde bulunan sayısal ifadelerin, sınıflandırma işleminin sonucuna etki etmeyeceği öngörülmüştür. Bu sebeple sayısal ifadeler Number Filter düğümüyle çıkarılmıştır. Gerçekleştirilen bu işlemle birlikte algoritmalar tarafından sayısal ifadelerin birer karakter olarak frekans hesaplarının yapılmaması ve dolayısıyla veri boyutunun azaltılması sağlanmıştır.
- g) *Stop Word Filter (Durak Kelime Filtresi)*: Bir metin verisi içerisinde bağlaçların ve edatların bulunması olağandır. Hüküm metinlerinin tamamında sıklıkla geçen bu bağlaç ve edatlar, hükümlerin konularına göre sınıflandırılmasında belirleyici değildir ve gerçekleştirilebilecek analiz işlemlerine anlamlı bir katkı sağlamayacaktır. Osmanlı Türkçesi ile oluşturulmuş metinlerin içerisindeki tüm durak kelimeleri kapsayan bir sözlüğün olmaması sebebiyle, yalnızca hüküm metinlerinin içerisinde bulunan (ve, veya gibi) Türkçe durak kelimeler çıkarılmıştır.
- h) *Dictionary Replacer (Sözlük Değiştirici)*: Dört farklı kişi tarafından transkript edilen hüküm metinlerinde bazı kelimelerin yazılışlarında farklılık gözlemlenmiştir. Örneğin bir araya getirilen verilerin içerisinde geçen İsmail ismi İsmâ'il ve İsmâ'il şeklinde de transkript edilmiştir. Bu durum ismin her bir yazımı için ayrı bir frekans değerinin hesaplanmasına

neden olacaktır. Frekans değerinin doğru bir biçimde belirlenebilmesi için buna benzer farklılıkların tek bir yazım biçimine dönüştürülmesi, diğer bir ifade ile standartlaştırılması gerekmektedir. Hüküm metnini oluşturan kelimelerin yazım hatası veya farklılığı sonucunda algoritmalarca birbirinden bağımsız kelimeler olarak algılanabilmesi ya da Türkçe yazım kurallarına uymadığı için algoritmalar tarafından anlaşılabilmesi gibi problemlerin giderilmesi için bir liste oluşturulmuştur. Bu liste içerisinde kelimelerin farklı ve yanlış yazımlarının düzeltilmiş ve algoritmalarca anlaşılabilir olacak biçimleri bulunmaktadır. Böylelikle kelimelerin kök bulma işlemleri ve durak kelimelerin çıkarılması süreçlerinde bu liste kullanılarak frekans hesaplamalarında daha başarılı sonuçlar elde edilmiştir.

- i) *N Chars Filter (N Karakter Filtresi)*: Metin verileri içerisinde bulunan kelimelerin bazıları tek başına bir anlam ifade etmemektedir. Bu sebeple N Chars Filter düğümü kullanılmıştır. Bu düğüm ile hedeflenen n sayısına verilecek değer kadar harften oluşan kelimelerin çalışmanın dışında bırakılmasıdır. Çalışmada n değeri 2 (iki) olarak belirlenmiştir ve iki harften oluşan kelimeler filtrelenmiştir.
- j) *Dictionary Tagger (Sözlük Etiketleyici)*: Yukarıda ifade edilen işlemler sonucunda 1559 anahtar kelime elde edilmiştir ve bu kelimeler incelenerek kişi adları ile yer adları belirlenmiştir. Table Creator (Tablo Oluşturucu) düğümü ile bir tablo oluşturulmuştur. Table Creator düğümü ile Dictionary Tagger düğümü birleştirilmiştir. Böylece yer adları ve kişi adlarının etiketlenmesi işlemi gerçekleştirilmiştir. Bu işlem yer ve kişi adları ile ilgili gerçekleştirilmesi muhtemel analizler için oldukça önemlidir.
- k) *Bag of Words Creator (Kelime Çantası Oluşturma)*: Kelimelerin frekans değerlerinin hesaplanması için gerçekleştirilen bu adımda, her bir hüküm oluşturan kelimeler ayrı ayrı satırlarla ifade edilmiştir. Ayrıca kelimelerine ayrılan her bir hükümde geçen yer ve kişi adları Dictionary Tagger düğümüyle gerçekleştirilen etiketleme işlemi sayesinde otomatik olarak belirlenmektedir.
- l) *DF (Document Frequency, Doküman Frekansı)*: Bag of Words Creator işlemi sonucunda elde edilen her bir kelimenin toplamda kaç hükümde geçtiğinin hesaplanması işlemidir. Bu işlem sonucunda frekans değerleri hesaplanan kelimelerin hükümlerde hangi sıklıkla kullanıldıkları belirlenmiştir.
- m) *Frequency Filter (Frekans Filtresi)*: Frekans değerleri hesaplanan kelimeler incelendiğinde bazı kelimelerin, hüküm verilerinin yarısından fazlasında sıklıkla kullanıldığı görülmüştür. Dolayısıyla bu kelimeler, hükümlerin konularına göre sınıflandırılması işleminde belirleyici olmayacaktır. Bu sebeple çalışmada bir kelimenin maksimum frekans değeri 450 olarak belirlenmiştir. 450 ve üzeri hükümde sıklıkla kullanılan kelimeler filtrelenmiştir. Bu işlem ile hükümlerin konularının belirlenmesinde, konuları temsil edebilecek kelimelerin kullanılması amaçlanmıştır.

Gerçekleştirilen ön işlem adımları sonucunda metinsel veriler, analiz yapılabilecek bir yapıya dönüştürülmüştür. Ayrıca yukarıda gerçekleştirilen işlemler sonucunda hükümlerde sıklıkla kullanılan kelimelerin neler olduğu ve kullanım sıklıkları, hükümlerde en çok kullanılmış olan kişi ve yer adları ile bunların kullanım sıklıkları gibi üzerinde çeşitli analiz ve çıkarımların yapılmasını mümkün kılacak tablolar elde edilebilmiştir.

Sınıflandırma İşlemi

Metinsel hüküm verilerinin algoritmalarca anlaşılabilir ve dolayısıyla üzerlerinde işlem yapılabilir olması diğer bir ifadeyle kelimelerin vektör olarak sayısal ifadelerle dönüştürülmesi işlemi Document Vector düğümü ile gerçekleştirilmiştir. Bu düğüm ile her bir hükümde geçen kelimelerin birleştirmeyen kelimelerin ise sıfırlarla ifade edildiği vektörler meydana gelmiştir. Ardından hükümlerin konularına göre sınıflandırılması işlemine geçilmiştir. Partitioning (Bölme) düğümü ile sınıflandırma

işlemi için kullanılan 1.209 adet hükmün %70'i eğitim verisi olarak, %30'u ise test verisi olarak ayrılmıştır⁴. Dolayısıyla eğitim verisi olarak toplamda 847 adet hüküm, test verisi olarak da 362 adet hüküm kullanılmıştır. Hükümlerin sınıflandırılması için çeşitli algoritmalar kullanılmıştır. Bu algoritmalar KNIME programı içerisinde bulunmaktadır ve Learner (Öğrenici) ile Predictor (Tahminci) olmak üzere iki yapıya sahiptir. Learner düğümü ile algoritma, eğitim için ayrılan 847 adet hüküm öğrenmektedir ve ardından Predictor düğümü ile test için ayrılan 362 hükmün konularını tahmin etmektedir. Çalışmada kullanılan sınıflandırma algoritmaları, literatürde sınıflandırma işlemlerinde sıklıkla kullanılan algoritmalar olmaları sebebiyle tercih edilmişlerdir. Bu algoritmalar ise şunlardır;

- Decision Tree (Karar Ağacı)*: “Karar ağacı, kararları ve muhtemel sonuçlarını göstermek için ağaç benzeri bir grafik ya da model kullanılan bir karar destek aracıdır” (Uysal ve Güyer, 2014, s. 34).
- Support Vector Machine (Destek Vektör Makinesi)*: Yüksek hacimli verilerin sınıflandırılması problemlerinin çözülmesinde kullanılabilen klasik makine öğrenmesi yöntemidir (Suthaharan, 2016, s. 207).
- Gradient Boosted Trees (Gradyan Artırma Ağaçları)*: Birden fazla ağacın oluşturulduğu ve gerçekleştirilen tahminlerdeki hataların göz önünde bulundurulmasıyla sürecin devam ettirilemeyeceği düşünülene kadar ağaç oluşturan Decision Tree benzeri bir algoritmadır.
- Random Forest Learner (Rastgele Orman Öğrenicisi)*: Sınıflandırma işlemini birden çok sayıdan karar ağacı vasıtasıyla gerçekleştiren algoritmadır. Verinin ait olduğu sınıf, karar ağaçlarının çoğunluğunun belirlediği sınıftır (Breiman, 2001, s. 6).
- K Nearest Neighbor (K En Yakın Komşu)*: Sınıflandırılacak verinin, belirlenen “k” değeri kadar, veri kümesindeki diğer verilere olan uzaklığının ölçülmesi, dolayısıyla en yakınında bulunan verilerin çoğunun sınıfının nihai sınıf olarak atandığı algoritmadır. Söz konusu uzaklıklar genellikle Öklid uzaklığı kullanılarak hesaplanmaktadır. Manhattan uzaklığı ise kullanılabilir başka bir alternatiftir (Nisbet vd., 2009, s. 239). Çalışmada 3, 5 ve 7 olmak üzere üç farklı şekilde k değeri belirlenmiştir. Bu durumun temel sebebi ise en doğru konu sınıflandırmasına ulaşılma hedefidir.

Kullanılan algoritmaların gerçekleştirdiği sınıflandırma işleminin başarı oranları ise her bir algoritma için çeşitli ağırlıklandırma yöntemlerine göre değerlendirilmiştir. Ağırlıklandırma yöntemlerine göre algoritmaların başarı oranları Scorer (Performans Hesaplayıcı) düğümü vasıtasıyla hesaplanmıştır. Kullanılan ağırlıklandırma yöntemleri ve elde edilen başarı oranları şu şekildedir;

- Binary (Bitsel) Ağırlıklandırma*: Bir kelimenin metin veri kümesi içerisinde kullanılmış ya da kullanılmamış olmasına bağlı olarak yapılan ağırlıklandırma yöntemidir (Canhası, 2007, s. 53). Bu ağırlıklandırma yöntemine göre en başarılı sonuç Gradient Boosted algoritması ile elde edilmiştir.

Tablo 3. Binary Ağırlıklandırma Yöntemiyle Sınıflandırma Algoritmalarının Başarısının İncelenmesi

Sınıflandırma algoritmaları	F-ölçüsü	Doğruluk oranı
Karar ağaçları	0,711	0,718

⁴ Metin ve veri madenciliği çalışmalarında eğitim ve test verileri genellikle bu oranda bölünmüştür. Bu duruma Toplu'nun (2019 s. 57) ve Budak ile Organ'ın (2022, s. 1009) çalışmaları örnek olarak verilebilir. Eğitim için ayrılan veri miktarının %70-80'den fazla olması algoritmanın verileri ezberlemesine yol açacaktır. Test verileri üzerinde başarı oranı yüksek çıktılar elde edilmesine rağmen yeni veriler üzerinde gerçekleştirilecek çalışmalarda başarı oranı düşecektir. %70-80 oranından daha az miktarda verinin eğitim için ayrılması durumu da algoritmanın başarılı sonuç verme oranını düşürecektir (Nisbet vd., 2009, s. 287). Dolayısıyla literatürdeki çalışmalarda veriler genellikle %70 eğitim ve %30 test verisi olarak bölünmektedir.

SVM (Destek Vektör Makineleri)	0,748	0,757
Rastgele Orman (Random Forest)	0,760	0,765
K-En yakın komşu (k=3)	0,579	0,605
K-En yakın komşu (k=5)	0,575	0,608
K-En yakın komşu (k=7)	0,614	0,644
Gradient Boosted ağaçları	0,765	0,773

- *Term Frequency (Terim Frekansı-TF)*: Bir kelimenin, bir belgede görünme sayısının söz konusu belgede geçen tüm terim sayısına bölünmesi sonucu elde edilen değerdir (Kotu ve Deshpande, 2015, s. 278). TF ağırlıklandırma yöntemi sonucunda en başarılı sonuç Random Forest algoritması ile elde edilmiştir.

Tablo 4. TF Ağırlıklandırma Yönteminde Sınıflandırma Algoritmalarının Başarısının İncelenmesi

Sınıflandırma algoritmaları	F-ölçüsü	Doğruluk oranı
Karar ağaçları	0,736	0,749
SVM (Destek Vektör Makineleri)	0,726	0,751
Rastgele Orman (Random Forest)	0,793	0,804
K-En yakın komşu (k=3)	0,627	0,660
K-En yakın komşu (k=5)	0,646	0,680
K-En yakın komşu (k=7)	0,662	0,696
Gradient Boosted ağaçları	0,767	0,782

- *Document Frequency (Doküman Frekansı-DF)*: Bir terimin tüm dokümanlarda kullanılma sayısıdır (Binici, 2016, s. 25). Bu ağırlıklandırma yöntemi sonucunda ise TF ağırlıklandırma yönteminde olduğu gibi, en yüksek başarı oranı Random Forest algoritması vasıtasıyla elde edilmiştir.

Tablo 5. DF Ağırlıklandırma Yöntemiyle Sınıflandırma Algoritmalarının Başarısının İncelenmesi

Sınıflandırma algoritmaları	F-ölçüsü	Doğruluk oranı
Karar ağaçları	0,715	0,724
SVM (Destek Vektör Makineleri)	0,724	0,735
Rastgele Orman (Random Forest)	0,805	0,812
K-En yakın komşu (k=3)	0,523	0,552
K-En yakın komşu (k=5)	0,512	0,541
K-En yakın komşu (k=7)	0,521	0,555
Gradient Boosted ağaçları	0,799	0,812

- *Inverse Document Frequency (Ters Doküman Frekansı-IDF)*: Bir terimin kullanıldığı doküman sayısının belirlenmesinin ardından terimin ağırlığının tersine çevrilmesi ile elde edilen ağırlıklandırma yöntemidir (Binici, 2016, s. 25-26). IDF ağırlıklandırma yöntemine göre en başarılı sonuç Gradient Boosted Trees algoritması ile elde edilmiştir.

Tablo 6. IDF Ağırlıklandırma Yönteminde Sınıflandırma Algoritmalarının Başarısının İncelenmesi

Sınıflandırma algoritmaları	F-ölçüsü	Doğruluk oranı
Karar ağaçları	0,710	0,718
SVM (Destek Vektör Makineleri)	0,744	0,754
Rastgele Orman (Random Forest)	0,789	0,793
K-En yakın komşu (k=3)	0,575	0,627
K-En yakın komşu (k=5)	0,541	0,624
K-En yakın komşu (k=7)	0,542	0,624
Gradient Boosted ağaçları	0,809	0,812

- *Inverse Class Frequency (Ters Sınıf Frekansı-ICF)*: Bir sınıf için belirleyici olan kelime ya da terimlerin diğer sınıflara da ait olabilmesi sonucunda bu kelime ya da terimlerin önemine vurgu yapabilmek için geçiş sıklıklarının hesaplandığı yöntemdir (Lernattee ve Theeramunkong, 2008, s. 1146). Bu ağırlıklandırma yöntemi sonucunda elde edilen en yüksek başarı oranı ise Support Vector Machine algoritması sonucunda elde edilmiştir.

Tablo 7. ICF Ağırlıklandırma Yönteminde Sınıflandırma Algoritmalarının Başarısının İncelenmesi

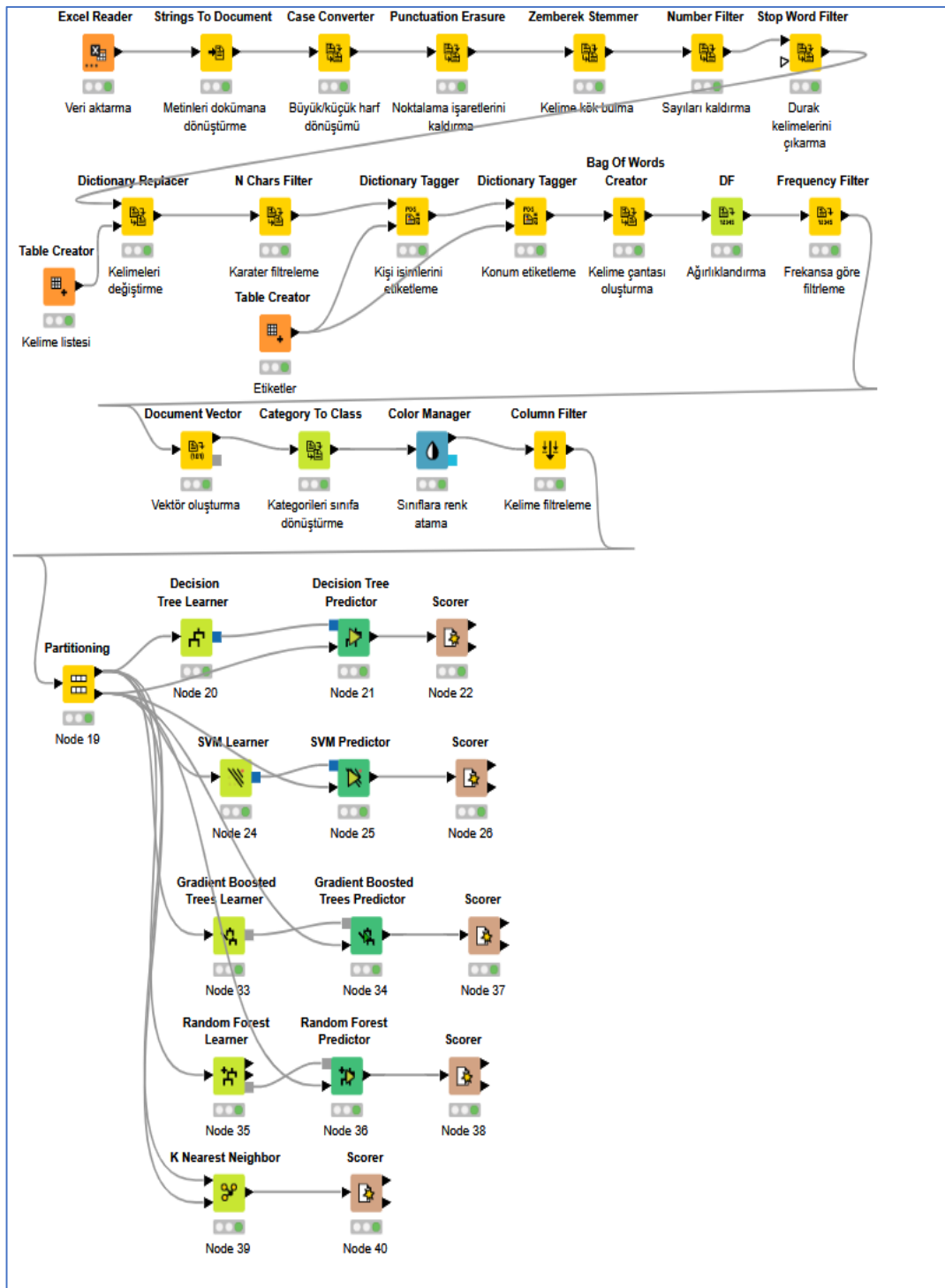
Sınıflandırma algoritmaları	F-ölçüsü	Doğruluk oranı
Karar ağaçları	0,753	0,751
SVM (Destek Vektör Makineleri)	0,795	0,801
Rastgele Orman (Random Forest)	0,760	0,765
K-En yakın komşu (k=3)	0,537	0,577
K-En yakın komşu (k=5)	0,513	0,572
K-En yakın komşu (k=7)	0,505	0,572
Gradient Boosted ağaçları	0,783	0,787

Yukarıda ifade edilen algoritmalar ve ağırlıklandırma yöntemlerine göre gerçekleştirilen sınıflandırma işlemlerinin sonuçları genel olarak incelendiğinde en yüksek başarılı sonucun IDF ağırlıklandırma yöntemine göre Gradient Boosted Trees olduğu belirlenmiştir. Dolayısıyla hükümlerin sınıflandırılması işleminin IDF ağırlıklandırma yöntemine göre Gradient Boosted Trees algoritmasıyla gerçekleştirilebileceği sonucuna varılmaktadır. Söz konusu algoritmanın hüküm verilerine uygulanması sonucunda elde edilen karışıklık matrisi şu şekildedir;

Tablo 8. Gradient Boosted Ağaçları Algoritması Sonucunda Elde Edilen Karışıklık Matrisi (n=362)

		Tahmin Edilen Sınıf			
		Asayiş	Sosyal, Ekonomik ve İdarî Konular	Arazi Meselesi ve Vergi	Başarı Oranı %
Doğru Sınıf	Asayiş	60	14	8	73,2
	Sosyal, Ekonomik ve İdarî Konular	8	108	14	83,1
	Arazi Meselesi ve Vergi	4	17	129	86

Yukarıdaki tablo incelendiğinde, konularına göre sınıflandırılmak için ayrılan 362 test hüküm verisinin konu başlıklarına göre sınıflandırma başarı oranları görülmektedir. Asayiş ile ilgili 82 adet test hüküm verisinden 60 hükmün konusu doğru tahmin edilmiştir. Geriye kalan 22 hükmün konuları ise yanlış belirlenmiştir. Asayiş ile ilgili hükümlerin %73,2'lik oranla başarılı bir şekilde sınıflandırıldığı görülmektedir. Sosyal, ekonomik ve idarî konularla ilgili 130 test hüküm verisi incelendiğinde, 108 hükmün konusu doğru belirlenmiştir. %83,1 oranında konuları doğru belirlenen hükümlere rağmen 22 hükmün konusu yanlış belirlenmiştir. Son olarak arazi meselesi ve vergi ile ilgili 150 hükümden 129'u doğru sınıflandırılmıştır. Bu hükümlerin konu başlığı %86'lık bir başarı oranıyla belirlenmiştir. 21 hükmün ise konuları yanlış belirlenmiştir. Dolayısıyla bu bilgiler ışığında gerçekleştirilen sınıflandırma modelinin başarılı olduğu sonucuna ulaşılabilir. KNIME programı ile gerçekleştirilen ve **başarılı** olduğu kabul edilen sınıflandırma modeli ise şu şekildedir.



Şekil 1. KNIME programında oluşturulan sınıflandırma modeli

SONUÇ VE ÖNERİLER

Osmanlı arşiv belgeleri, coğrafya, etnoloji, kronoloji, filoloji, felsefe, hukuk, tarih, iktisat, edebiyat, sosyoloji ve tıp bilimlerinde gerçekleştirilen çalışmalar için veri kaynağı niteliği taşımaktadır. Osmanlı arşiv belgeleri üzerinde gerçekleştirilen çalışmaların ise derin öğrenme, makine öğrenmesi ve yapay zekâ gibi gelişen bilişim teknolojilerinin sağladığı imkânlar dâhilinde yapılması önem arz etmektedir. Bu duruma, söz konusu teknolojilerin araştırmacılar için zamandan tasarruf etme ve iş yükünü en aza indirme gibi kolaylıkları sağlayabilmesi ve daha önce keşfedilmemiş potansiyel

bilgiye erişmeyi mümkün kılabilmesi sebep olarak gösterilebilir. Genellikle metin formatında olan Osmanlı arşiv belgeleri üzerinde gerçekleştirilecek çalışmaların manuel olarak yapılması oldukça zorlu ve zaman alıcı bir süreç olabilmektedir. Defteri meydana getiren hükümlerin otomatik şekilde, konularına göre sınıflandırılmasını mümkün kılan bir model oluşturmak amacıyla gerçekleştirilen bu çalışma ile şu sonuçlar ifade edilebilir;

- Yarı yapılandırılmış hâlde olan hüküm metin verisi, metin madenciliği vasıtasıyla üzerinde işlem yapılabilir ve bilgisayarlarca anlaşılabilir hâle, diğer bir ifade ile yapılandırılmış veriye dönüştürülmüştür.
- Hüküm konularının belirlenebilmesi için KNIME programı vasıtasıyla bir model oluşturulabilmiştir. Söz konusu model ile birlikte Osmanlı arşiv belgelerinden farklı ve sınıflandırmaya uygun fonlara ait veri kümeleri üzerinde sınıflandırma işleminin gerçekleştirilebilmesi mümkün olabilmektedir.
- Hüküm konularının manuel olarak belirlenmesi işlemi oldukça zahmetli ve zaman alıcı bir işlemdir. Gerçekleştirilen metin madenciliği işlemi ile birlikte konu başlıklarının otomatik olarak belirlenmesi, zaman kaybını minimize etmiştir.
- Manuel olarak konu başlıklarının belirlenmesi işleminde hatalı konu başlığı atanmasının yapılması muhtemeldir. Bir hükmün birden fazla konuyu kapsayabilmesi benzer hükümlerin farklı konu başlıkları altında değerlendirilmesine dolayısıyla hatalı konu başlığının seçilmesine neden olabilmektedir. Otomatik olarak gerçekleştirilen sınıflandırma işleminde kelimelerin ağırlıklandırma yöntemlerine göre değerlendirilmesi, yüksek oranda başarılı konu başlığının atanmasını mümkün kılmaktadır.
- Oluşturulan model ile birlikte çalışmanın veri kümesine eklenen yeni verilerin sınıflandırılması da hızlı ve daha yüksek başarı oranıyla gerçekleştirilebilmektedir.
- Kelimelerin frekans değerlerine göre çeşitli istatistiksel veriler elde edilebilmektedir. Bu veriler kelime bulutu gibi çeşitli görselleştirme yöntemleriyle kolay anlaşılabilir ve analiz edilebilir bir hâle getirilebilmektedir. Dolayısıyla veri içerisinde gömülü olan çeşitli potansiyel bilgiye erişilebilmesi mümkün olmuştur.

Osmanlı arşiv belgeleri üzerinde gerçekleştirilecek metin madenciliği çalışmalarına katkı sağlaması ve muhtemel çalışmalara yardımcı olabilmesi amacıyla şu önerilerde bulunulabilir;

- Genellikle Osmanlı Türkçesi ve Arap harfleriyle oluşturulmuş veri kaynakları üzerinde çalışmalar gerçekleştirilecektir. Bu veri kaynaklarının Latin harflerine dönüştürülmesi gerekmektedir. Bunun yanı sıra kelimelerin Türkçe yazım kurallarına göre düzenlenmesi, ön işlem aşamalarının başarısına doğrudan etki etmesi sebebiyle oldukça önemlidir. Bu işlem zaman alıcı ve özel beceri gerektiren bir süreçtir. Dolayısıyla transkript edilmiş verilerin kaynak olarak ele alınması ya da Arap harflerini Latin harflerine çevirmek için geliştirilen uygulamaların kullanılması gerekmektedir.
- Transkript işleminde aynı kelimelerin yazım biçimlerinde bir standart oluşturulmalıdır.
- Transkript işlemi gerçekleştirilirken verilerin içerisinde bulunan kişi, yer ve mekân isimleri, verilerin konuları ve kaynakları kapsamlı bir üstveri standardına göre belirlenmelidir.
- Konularına göre sınıflandırma işleminin gerçekleştirileceği çalışmalarda eğitim verisinin konuları belirlenirken bir standarda bağlı kalınmalıdır. Üzerinde işlem gerçekleştirilecek veri sayısı az ise konu başlıkları daha genel kelimelerle, veri sayısı çok ise daha özel kelimelerle ifade edilmelidir. Her iki duruma göre önceden hazırlanmış bir konu başlığı şablonu geliştirilmelidir ve konu başlığı belirlenirken söz konusu şablona bağlı kalınmalıdır.

- Zemberek Kütüphanesi ile aynı işleve sahip, Osmanlı Türkçesi için kelimelerin köklerini bulacak, diğer bir ifade ile gövdeleme işlemini gerçekleştirecek bir kütüphane oluşturulmalıdır.
- Veriler üzerinde gerçekleştirilen işlemlerin sonunda amaca uygun algoritmalar tercih edilmelidir. Çalışmanın çıktıları disiplinler arası bir yaklaşımla ele alınmalıdır.
- Çalışmadan elde edilen çıktıların yapısına uygun görselleştirme yöntemleri tercih edilmeli ve çıktılar görselleştirilmelidir. Dolayısıyla gerçekleştirilmesi muhtemel yeni çalışmalara uygun zemin hazırlanmalıdır.

KAYNAKÇA

- Akyüz, F. ve Gülten, S. (2022). Metin Madenciliği Kelime Muhasebesi ve Denetimi. Ankara: Gazi Kitabevi Yayınları.
- Alparslan, A. (1994). Divanî. Türkiye Diyanet Vakfı İslam Araştırmaları Ansiklopedisi içinde (Cilt 9, 445-446). İstanbul: TDV İslâm Araştırmaları Merkezi.
- Babadağ, M. (2021). İstihbaratta Büyük Veri ve Veri/Metin Madenciliği: Bir Uygulama. (Yayımlanmamış yüksek lisans tezi). Polis Akademisi Güvenlik Bilimleri Enstitüsü, Ankara.
- Binici, K. (2016). Kütüphane ve Bilgi Bilimi Çalışmalarında Dönemsel Konu Analizi. (Yayımlanmamış doktora tezi). İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Bolasco, S., Canzonetti, A., Capo, F.M., della Ratta-Rinaldi, F. ve Singh, B.K. (2005). Understanding Text Mining: A Pragmatic Approach. Spiros Sirmakessis (Ed.). Knowledge Mining: Studies in Fuzziness and Soft Computing içinde. (31-50 ss.). Heidelberg: Springer.
- Breiman, L. (2001). Random Forests, Machine Learning, 45, ss. 5-32.
- Budak, H. (2018). Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 22, ss. 21-31.
- Budak, İ. ve Organ, A. (2022). Veri ve Metin Madenciliği ile Hava Yolu İşletmelerinin Covid-19 Öncesi ve Sonrası Sosyal Medya Yorum ve Skorlarının Değerlendirilmesi. Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15(4), ss. 998-1022.
- Canhası, E. (2007). Edebiyat Eserlerinin Web Verilerine Dayanarak Sınıflandırılması. (Yayımlanmamış yüksek lisans tezi), Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Çiçek, H. (2011). Erzurum Vilayeti 1 Numaralı ve 1155-1162/1742-1749 Tarihli Ahkâm Defteri'nin Transkripsiyonu ve Değerlendirilmesi (s. 65-130). (Yayımlanmamış yüksek lisans tezi). Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum.
- Devellioğlu, F. (2008). Osmanlıca-Türkçe Ansiklopedik Lügat. Ankara: Aydın Kitapevi.
- Dong, H., Yu, S. ve Jiang, Y. (2009). Text Mining on Semi-structured E-Government Digital Archives of China. Xianzhi Tian, Luo Qi (Ed.). 2009 Second Pacific-Asia Conference on Web Mining and Web-based Application içinde, (11-14 ss.). Wuhan: Computer Society Press.
- Doğan, K. ve Arslantekin, S. (2018). Elektronik Belge Yönetimi, Dijital Arşivleme Sistemleri ve Büyük Veri. Fahrettin Özdemirci, Zeynep Akdoğan (Ed.). Bilgi Sistemleri ve Bilişim Yönetimi: Beklentiler ve Yeni Yaklaşımlar içinde, (65-80 ss.). Ankara: Ankara Üniversitesi.
- Emecen, F. M. (2005). Osmanlı Divanının Ana Defter Serileri: Ahkâm-ı Mirî, Ahkâm-ı Kuyûd-ı Mühimme ve Ahkâm-ı Şikâyet. Türkiye Araştırmaları Literatür Dergisi, 3(5), ss. 107-139.
- Emre, İ. E. ve Selçukcan, E. Ç. (2017). Veri Analizinde İstatistik mi Veri Madenciliği mi?. Bilişim Teknolojileri Dergisi, 10(2), ss. 161-167.

- Frawley, W. J., Piatetsky-Shapiro, G. ve Matheus, C. J. (1992). Knowledge Discovery in Databases an Overview. *AI Magazine*, 13(3), ss. 57-60.
- Gahi, Y., Guennoun, M. ve Mouftah, H. T. (2016). Big Data Analytics: Security and privacy challenges. 2016 IEEE Symposium on Computers and Communication (ISCC) içinde, (952-957 ss.). Messina: IEEE.
- Göker, H. ve Tekedere, H. (2017). Fatih Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(3), ss. 291-299.
- Günay, R. (2013). Diyarbakir Ahkâm Defterleri ve Muhtevaları. *Elektronik Sosyal Bilimler Dergisi*, 12(45), ss. 290-307.
- Haltaş, A., Alkan, A. ve Karabulut, M. (2015). Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 30(3), ss. 417-427.
- Hearst, M. (2003, October 17). What is Text Mining?. 07.01.2023 tarihinde <http://www.sims.berkeley.edu/~hearst/text-mining.html> adresinden erişildi.
- İlhan, U. (2001). Application of K-NN and FPTC Based Text Categorization Algorithms to Turkish News Reports. (Yayımlanmamış yüksek lisans tezi). Bilkent Üniversitesi Mühendislik ve Fen Bilimleri Enstitüsü, Ankara.
- İnalcık, H. (1988). Şikâyet Hakkı: ‘Arz-ı Hal ve ‘Arz-ı Mazhar’lar. *Osmanlı Araştırmaları Dergisi*, 7-8(7-8), ss. 33-54.
- Kantaroğlu, N. G. (2019). 1 Numaralı ve 1155-1162/1742-1749 Tarihli Erzurum Ahkâm Defterinin Transkripsiyonu ve Değerlendirilmesi (s. 212-293). (Yayımlanmamış yüksek lisans tezi). Amasya Üniversitesi Sosyal Bilimler Enstitüsü, Amasya.
- Kaşıkcı, T. ve Gökçen, H. (2014). Metin Madenciliği ile E-Ticaret Sistemlerinin Belirlenmesi. *Bilişim Teknolojileri Dergisi*, 7(1), ss. 25-32.
- Kobayashi, V. B., Mol, S. T., Vrolijk, J. ve Kismihok, G. (2021). Text Mining in Career Studies: Generating Insights from Unstructured Textual Data, Ed. Wendy Murphy v.d., *Handbook of Research Methods in Careers*, içinde (139-163 ss.), Cheltenham, UK: Edward Elgar Publishing
- Kotu, V. ve Deshpande, B. (2015). Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner. Waltham, Morgan Kaufmann..
- Kul, M. (2011). 1 Numaralı ve 1155-1162/1742-1749 Tarihli Erzurum Ahkâm Defterinin Transkripsiyonu ve Değerlendirilmesi (s. 1-65). (Yayımlanmamış yüksek lisans tezi). Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum.
- Kutluca, T. (2013). Excel Yazılımı İle Geliştirilen Bilgisayar Destekli Bir Öğretim Materyalinin Tasarlanması. *Elektronik Eğitim Bilimleri Dergisi*, 2(4), ss. 40-49.
- Lernattee, V. ve Theeramunkon G. T. (2008). Improving Thai Academic Web Page Classification Using Inverse Class Frequency and Web Link Information. 22nd International Conference on Advanced Information Networking and Applications-Workshops (aina workshops 2008) içinde, (1144-1149 ss.). Okinawa: IEEE
- Miner, G., Delen, D., Elder, J., Fast, A., Hill, T. ve Nisbet R. A. (2012). Practical Text Mining and Statistical Analysis for Nonstructured Text Data Applications. Waltham: Academic Press.
- Mumcu, A. (1994). Dîvân-ı Hümayun. Türkiye Diyanet Vakfı İslam Araştırmaları Ansiklopedisi içinde (Cilt 9, 430-432). İstanbul: TDV İslâm Araştırmaları Merkezi.
- Nisbet, R., Elder, J. ve Miner, G. (2009). Handbook of Statistical Analysis and Data Mining Applications. MA, Academic Press.

- Oğuz, B., Bilge, U. ve Saka, O. (2008). Tıpta Metin Madenciliği. Hakan Gülkesen (Ed.). Tıp Bilişimi'07: 4. Ulusal Tıp Bilişimi Kongresi içinde (108-112 ss.). Antalya.
- Oğuzlar, A. (2011). Temel Metin Madenciliği. Bursa: Dora.
- Özdemirci, F. ve Torunlar, M. (2015). Bilgi Çağında Arşivsel Bilgi Analizi: Bilgi-İktidar-İdeoloji-Devlet. Ankara: Ankara Üniversitesi Basımevi.
- Özel, A. (1988). Ahkâm. Türkiye Diyanet Vakfı İslam Araştırmaları Ansiklopedisi içinde (Cilt 1, 550-551). İstanbul: TDV İslâm Araştırmaları Merkezi.
- Öztürk, S. (2001). Sosyo-Ekonomik Tarih Kaynağı Olarak Ahkam Defterleri. Kemal Çiçek (Ed.). Pax Ottomana Studies in Memoriam Prof. Dr. Nejat Göyünç içinde, (611-639 ss.). Haarlem-Ankara: Sota-Yeni Türkiye.
- Pakalın, M. Z. (2004). Osmanlı Tarih Deyimleri ve Terimleri Sözlüğü. İstanbul: Millî Eğitim Bakanlığı Yayınları.
- Sahillioğlu, H. (1988). Ahkâm Defteri. Türkiye Diyanet Vakfı İslam Araştırmaları Ansiklopedisi içinde (Cilt 1, 551). İstanbul: TDV İslâm Araştırmaları Merkezi.
- Sarıcaoğlu, F. (2007). Dîvân-ı Hümayûn'un Kronolojik Toplanma ve Merasim Günleri (1153-1210, 1740-1795). Osmanlı Araştırmaları Dergisi, 30(30), ss. 87-143.
- Sukanya, M. ve Biruntha. S. (2012). Techniques on Text Mining. 2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT) içinde, (269-271. ss.). Ramanathapuram: IEEE.
- Susam, F. (2019). 1 Numaralı ve 1155-1162/1742-1749 Tarihli Erzurum Ahkâm Defterinin Transkripsiyonu ve Değerlendirilmesi (s. 131-212). (Yayımlanmamış yüksek lisans tezi). Amasya Üniversitesi Sosyal Bilimler Enstitüsü, Amasya.
- Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. Greensboro: Springer.
- Şeker, Ş. E. (2018). CRISP-DM: Endüstriler Arası Standart İşleme-Veri Madenciliği İçin (Cross Industry Standard Processing-Data Mining). YBS Ansiklopedi, 5(2), ss. 10-16.
- Şentürk, A. (2006). Veri Madenciliği: Kavram ve Teknikler. Bursa: Ekin Yayınevi.
- Şimşir, N. (1994). Ahkâm Defterlerinin Tarihi Kıymeti ve 107 No'lu Ahkâm Defterindeki İzmir ile İlgili Hükümler. Tarih İncelemeleri Dergisi, 9(1), ss. 357-390.
- T.C. Başbakanlık Devlet Arşivleri Genel Müdürlüğü, Osmanlı Arşivi Daire Başkanlığı. (1994). 5 Numaralı Mühimme Defteri (973/1565-1566): Özet ve İndeks (Yayın No. 21). Erişim adresi: <https://www.devletarsivleri.gov.tr/cdn/file/download?fileId=47>.
- T.C. Başbakanlık Devlet Arşivleri Genel Müdürlüğü, Osmanlı Arşivi Daire Başkanlığı. (2017). Başbakanlık Osmanlı Arşivi Rehberi (Yayın No. 147). Erişim adresi: <https://www.devletarsivleri.gov.tr/cdn/file/download?fileId=173>.
- Temelkuran, T. (1975). Divân-ı Hümayun: Mühimme Kalemî. İstanbul Üniversitesi Edebiyat Fakültesi Tarih Enstitüsü Dergisi, (6), ss. 128-175.
- Toplu, S. G. (2019). Metin Madenciliği ve Sağlık Alanında Bir Uygulama. (Yayımlanmamış yüksek lisans tezi). Sağlık Bilimleri Enstitüsü, Düzce.
- Tunç, M. (2022). Ahkâm Defterleri'nin Vakıf Araştırmalarındaki Yeri (1 Numaralı Erzurum, Karaman ve Maraş Ahkâm Defterleri). Çankırı Karatekin Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 13(1), ss. 1-34.
- Uysal, M. ve Güyer, T. (2014). İstatistiksel Veri Analizine İlişkin Genişleyebilir Bir Karar Ağacı Tasarımı. Bilişim Teknolojileri Dergisi, 7(3), ss. 33-43.

- Uzunçarşılı, İ. H. (1998). Osmanlı Devleti'nin Saray Teşkilatı. Ankara: Türk Tarih Kurumu
- Üzüm, İ. (1998). Hüküm. Türkiye Diyanet Vakfı İslam Araştırmaları Ansiklopedisi içinde (Cilt 18, 464-466). İstanbul: TDV İslâm Araştırmaları Merkezi.
- Yzaguirre, A., Smit, M. ve Warren, R. (2016). Newspaper Archives + Text Mining = Rich Sources of Historical Geo-spatial Data. Hugh Millward, Dirk Werle, Danika van Proosdij (Ed.). The 9th Symposium of the International Society for Digital Earth (ISDE) içinde, (1-8 ss.). Halifax: IOP Publishing.