



Erişime Açık Terörizm Veri Kümeleri Kullanarak Makine Öğrenmesi ve Büyük Veri Mimarileri ile Terörle Mücadeleye Yönelik Tahminleme Yaklaşımları

Caner DERELİ* - Fatih ADIGÜZEL** - Pınar KARAGÖZ***

Öz

Terör yüzyıllardır insanlığın mücadele içinde bulunduğu bir güvenlik problemidir. Son yıllarda yapay zekâ, makine öğrenmesi ve büyük veri teknolojilerinin gelişmesi ve farklı alanlarda giderek artan kullanımı ile paralel olarak terörle mücadelede de bu teknolojilerinin kullanıldığını görmekteyiz. Bu çalışmada, terörle mücadele için erişime açık veri kaynakları, makine öğrenmesi ve büyük veri teknolojilerinin kullanımı incelenmiştir. Makalede ilk önce terör konusunda erişime açık veri kaynakları hakkında bilgi sunulmaktadır. Daha sonra literatürde son yıllarda bu veri kaynaklarını kullanan makaleler için yapılan tarama sonuçları paylaşılmıştır. İlgili veri kaynaklarına değinen oldukça fazla sayıda makale tespit edilmiş olmakla birlikte bunlardan sadece belirli sayıdaki çalışma, yapay zekâ ve büyük veri teknolojilerinin kullanımına odaklanmıştır. Özellikle bu konulara odaklanan çalışmalar detaylı olarak incelenerek, ele alınan problemler, kullanılan çözüm yaklaşımları ve alınan sonuçlar karşılaştırmalı olarak sunulmaktadır.

Anahtar Kelimeler: *Terörizmle Mücadele, Erişime Açık Veri Kümesi, Yapay Zekâ, Makine Öğrenmesi, Büyük Veri Mimarileri.*

* Yüksek Lisans Öğrencisi, Milli Savunma Üniversitesi, KBRN-P Savunma Ana Bilim Dalı, EYP ve Patlayıcılar İle Mücadele, caner_dereli@yahoo.com, ORCID: 0000-0001-9939-5783.

** Yüksek Lisans Öğrencisi, Milli Savunma Üniversitesi, KBRN-P Savunma Ana Bilim Dalı, EYP ve Patlayıcılar İle Mücadele, hfa5303@yahoo.com, ORCID:0000-0002-1781-3783.

*** Prof.Dr., Orta Doğu Teknik Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, karagöz@ceng.metu.edu.tr , ORCID: 0000-0003-1366-8395.

Geliş Tarihi/Received : 07.12.2021

Kabul Tarihi/Accepted : 07.09.2022

Araştırma Makalesi/ Research Article

DOI: 10.17134/khosbd.1031843

Prediction Approaches to Counter Terrorism with Machine Learning and Big Data Architectures Using Open Access Terrorism Datasets

Abstract

Terrorism continues as a security problem that humanity has been struggling with for centuries. In recent years, in parallel with the development of artificial intelligence, machine learning and big data technologies and their increasing use in different fields, it is seen that these technologies are used in the fight against terrorism as well. In this study, open access data sets relevant to terror events and the use of machine learning and big data technologies to combat terrorism are examined. In the article, firstly, information about the open access data sources on terrorism is presented. Then, the results of the literature search on the recent related studies using these data sources are shared. Although a large number of articles have been identified, only a few of them focused on the use of artificial intelligence and big data technologies. The studies focusing on these issues are examined in detail, and the problems addressed, the solution approaches used and the results obtained are presented comparatively.

Keywords: *Counterterrorism, Open Dataset, Artificial Intelligence, Machine Learning, Big Data Architectures.*

Giriş

Terör yüzyıllardır olduğu gibi günümüzde de en önemli güvenlik sorunlarından biri olarak ulusal, bölgesel ve uluslararası düzeyde önemli bir tehdit oluşturmaktadır. *Terör* sözcüğü, Türk Dil Kurumu sözlüğünde “yıldırma, cana kıyma ve malı yakıp yıkma, korkutma, tedhiş” şeklinde açıklanmaktadır (Türk Dil Kurumu, 2021). Bu tanımla paralel olarak *terörizm* sözcüğü ise “siyasal bir hedefe ulaşmak amacıyla devlete, halka veya bireylere karşı şiddet eylemlerine başvurma” (Türk Dil Kurumu, 2021) olarak ifade edilmektedir. Tanımlardan da görüldüğü üzere, terörizm, toplum düzenine farklı boyutlarda zarar vermekte ve tehdit oluşturmaktadır. Bu nedenle terör tehdidi altındaki ülkeler, terörle mücadeleye önem vermekte ve farklı stratejiler geliştirmektedirler (Güçlü, 2014; Çıtak, 2021).

Terör eylemlerinin yıkıcı etkiye neden olan önemli bir yönü de olayların ne zaman ve nasıl meydana geleceği konusundaki belirsizliktir. Olası terör eylemlerini öngörebilmek ve engellemek için istihbarat kaynakları önemli rol oynamakla birlikte (Çıtak, 2021), bu bilgilerin elde edilmesi ve değerlendirilmesi her zaman mümkün olmamaktadır. Terör eylemlerindeki bu belirsizlik, yeni yöntemlere ihtiyaç doğurmaktadır. Gelişen bilgi teknolojileri, terörle mücadele ile yeni olanaklar sunmakta, bu teknolojilerle, terör olayları ile ilgili geçmiş bilgilerin toplanması, saklanması ve analiz edilmesi mümkün olmaktadır (Bhatia, Chhabra ve Kumar, 2020). Özellikle büyük veri teknolojileri ve yapay zekâ tekniklerinin kullanımı ile terör olayları ile ilgili tahminlemelere yönelik çalışmalar giderek artmaktadır (Atsa'am, Wario ve Okpo, 2020; Vaibhav, 2020; Singh, Chaudhary ve Kaur, 2019). Bunun yanı sıra geçmiş terör olaylarına ilişkin verilerin analiz amaçlı olarak toplanarak terör olaylarına ilişkin veri tabanları oluşturulmaya başlandığı görülmektedir (GTD, 2021; AOAV, 2021). Bu veri koleksiyonları yapay zekâ ve büyük veri analizi çalışmaları için önemli bir kaynak oluşturmaktadır.

Yapay zekâ, makina öğrenmesi, büyük veri teknolojilerinin farklı alanlarda kullanımına ilişkin literatür incelemelerini sunan çeşitli makaleler bulunmaktadır. Ancak terörle mücadele konusunda bu tekniklerin kullanımına ilişkin bir literatür incelemesinin sunulmamış olduğu görülmektedir. Bu eksiklikten hareketle, özellikle son yıllarda (2019-2022) terör olaylarına ilişkin açık kaynak veri kaynakları ve bu verileri kullanarak terörle mücadele için yapay zekâ ve büyük veri tekniklerini kullanan çalışmaları içeren bir literatür incelemesi sunulmuştur.

Son yıllarda yapay zekâ ve büyük veri teknolojilerindeki gelişmelerle pek çok farklı alanda yenilikçi çözümlerin kullanıldığı görülmektedir. Yapay zekâ için pek çok farklı tanım yapılmakla birlikte (McCarthy ve Hayes, 1981; Fetzer, 1990), mevcut durumu algılama, anlamlandırma ve buna göre çözüm geliştirme konularında fark yarattığı söylenebilir. Yapay zekâ konusunu bir alt alanı olan makine öğrenmesi, geçmiş durum ve deneyimleri içeren verilerle olay ve durum örüntüleri bulmayı ve tahminleme yapmayı hedefleyen çözümler geliştirmeyi amaçlar. Makine öğrenmesi tekniklerini (Michalsky, Corbonell ve Mitchell, 2013) ana hatlarıyla *gözetimli öğrenme* ve *gözetimsiz öğrenme* olarak ikiye ayrılmaktadır. Gözetimli öğrenme tekniklerinde ilgili veri için bilinen gruplamalar mevcuttur ve yeni gelen bir veri için grup etiketinin tahmin edebilecek modeller geliştirilmesi

esastır. Örneğin terör olaylarının kullanılan silah tipine göre gruplanması ve silah tiplerinin önceden belli olması durumu için, gözetimli öğrenme modeli, yeni bir terör olayına ilişkin, eldeki mevcut verileri kullanarak, olayda kullanılan silah tipini tahmin edebilir. Oluşturulan matematiksel model ve izlenen adımlara göre çeşitlilik gösteren pek çok gözetimli öğrenme algoritması bulunmaktadır. Bu algoritmalara Karar Ağaçları, Naif Bayes Sınıflandırıcı, Destek Vektör Makinesi, K-En Yakın Komşuluk (KNN) gibi algoritmalar örnek olarak verilebilir. (Michalsky vd., 2013; Han, Pei ve Kamber, 2011). Gözetimsiz öğrenme tekniklerinde ise, hedef bir gruplama verilmeden eldeki verinin doğasından gelen gruplamaların otomatik olarak bulunması amaçlanır. Bu makina öğrenmesi yaklaşımı altındaki en bilinen yöntem *kümelemedir*. Eldeki terör olayları verilerinin olayların benzerliklerine göre kendi içinde gruplanması kümeleme için örnek olarak verilebilir. Gözetimsiz öğrenme yöntemlerinde de yine eldeki örnekler arasındaki benzerliklerin farklı şekilde belirlendiği çeşitli algoritmalar bulunmaktadır. Bunlar K-Ortalama, DBSCAN (Density-Based Spatial Clustering of Applications with Noise Algorithm) Algoritması gibi algoritmalar (Michalsky vd., 2013; Han vd., 2011).

Çeşitli kaynaklardan sağlanan verilerin artması bir yandan makine öğrenmesi için yeni olanaklar oluştururken, verilerin toplanması, saklanması, işlenmesi ve sorgulanması gibi pek çok aşamada büyüyen veri miktarı, karmaşıklaşan veri yapısı ve artan veri akış hızı ile birlikte yeni zorluklar ortaya çıkmıştır. *Büyük veri* olarak adlandırılan bu durum karşısında *büyük veri teknolojileri* olarak adlandırabileceğimiz yeni veri saklama, işleme ve sorgulama yapıları ve mimarileri ortaya çıkmıştır (Azarmi, 2016). Map-reduce yaklaşımı ve bu yaklaşımı kullanan bir çözüm yöntemi olan Hadoop, büyük veri mimarileri arasında en bilinen örnek olarak yer almaktadır (Dittrich ve Quiané-Ruiz, 2012). Map-reduce yaklaşımı, büyük miktarda belli bir prensibe göre farklı işlemci ve veri saklama birimlerine dağıtıp aynı şekilde ayrı ayrı işleme ve elde edilen sonuçları toplama temeline dayanır.

Sunulan makalede makine öğrenmesi ve büyük veri teknolojilerinin terörle mücadelede kullanımı incelenmiştir. Bu amaçla ilk önce açık erişimli terör olayları veri tabanları ve veri kümeleri araştırılmıştır. Yaptığımız incelemede bu niteliklere sahip dört veri kaynağı bulunmuştur. Bunlar; Global Terörizm veri tabanı (GTD, 2021), Action on Armed Violence veri tabanı (AOAV, 2021), Data World (DW, 2021) veri kümeleri ve Kaggle veri kümeleridir (Kaggle, 2021). Bunun devamında,

bu veri kümeleri üzerinde terörle mücadele odaklı yakın tarihli (2020-2022) akademik çalışmaları taranmıştır. Yapılan taramada çok sayıda makale künyesi tespit edilmiş olmasına rağmen, bu makaleler üzerinde yapılan daha detaylı incelemede, sınırlı sayıda çalışmada yapay zekâ ve büyük veri teknolojilerinin kullanılmış olduğunu ve veri olarak sadece GTD veri tabanının kullanılmış olduğu gözlemlenmiştir. Belirlenen diğer veri kaynaklarının makalelerde ismen ya da kaynakça olarak belirtilmiş olmasına rağmen bu veriler kullanılarak yapılan bir analiz sunulmamıştır. Yapay zekâ ve büyük veri teknolojileri konuları dışında kalan diğer makalelerde, terörizm başlığı altında terör probleminin sosyal, ekonomik, psikolojik boyutları gibi farklı hususlar ele alınmıştır. Makale kapsamında, yapay zekâ ve büyük veri teknolojilerinin kullanıldığı on adet çalışma detaylı olarak sunulmaktadır.

Makalenin geri kalanı şu şekilde düzenlenmiştir. Birinci bölümde geçmiş terör olaylarını içeren veri kümeleri hakkında bilgi paylaşılmıştır. İkinci bölümde çalışma kapsamında uygulanan makale tarama yaklaşımı sunulmuş ve tarama sonucunda bulunan makalelere ilişkin bilgi verilmiştir. Üçüncü bölümde incelenen çalışmalarda kullanılan makine öğrenmesi algoritmaları tanıtılmıştır. Dördüncü bölümde makine öğrenmesi ve büyük veri teknolojileri kullanarak GTD üzerinde analiz ve tahminleme yapan çalışmalar, ele aldıkları problemler, kullandıkları yöntemler ve alınan sonuçlar açısından detaylandırılarak karşılaştırmalı olarak sunulmuştur. Sonuç bölümünde ise yapılan literatür tarama çalışmasından çıkartılan sonuçlar ve tartışma verilmektedir. Makale son olarak özet bir değerlendirme ile tamamlanmıştır.

1. Veri Kümeleri

Bu bölümde, meydana gelmiş terör olaylarına ilişkin verileri içeren açık erişimli dört veri kaynağı üzerine bilgi sunulmaktadır. Bu dört veri kaynağı için özet bilgi Tablo 1'de verilmiştir. Yapılan ön incelemede literatürdeki çalışmalarda öne çıktıklarının görülmesi nedeniyle bu dört veri kaynağı seçilmiştir. Bunların yanı sıra sadece tek ya da oldukça kısıtlı sayıda çalışmaya özel ve erişime açık olmayan veri kümeleri kapsam dışı tutulmuştur.

Tablo 1. Veri Kümeleri Özet Bilgileri

Veri Kümesi Adı	Erişim Adresi	İçerik
GTD	www.start.umd.edu/gtd	Dünya genelinde terör olayları ile çeşitli kaynaklardan toplanan bilgiler tasnif edilerek bir veri tabanı oluşturulmuştur.
AOAV	www.aoav.org.uk	Dünya genelinde patlayıcı silahlarla meydana gelen terör olaylarına ilişkin İngilizce haber kaynaklarından toplanan bilgiler tasnif edilerek bir veri tabanı oluşturulmuştur.
Data.world	https://data.world/	Farklı konularda verilerin bulunduğu bir veri portalı yapısındadır. Terör olaylarına ilişkin veriler de bulunmaktadır.
Kaggle	www.kaggle.com	Makine öğrenmesi uygulamaları ve yarışmalarına yönelik bir portal yapısındadır. Terör olaylarına ilişkin veriler de bulunmaktadır.

a. Global Terörizm Veri Tabanı

Global Terörizm Veri Tabanı (Global Terrorism Database-GTD) dünya genelinde meydana gelmiş terör olayları ile bilgileri bünyesinde toplayan erişime açık bir veri tabanıdır (GTD, 2021). Veri tabanı 1970-2019 yılları arasında meydana gelmiş olan terör olaylarını içermektedir. 2022 Ocak ayı itibariyle 220.000'i aşkın terör olayına ilişkin bilgiler içermektedir. Her olay için, olayın tarihi ve yeri, kullanılan silahlar ve hedefin niteliği, yaralı sayısı ve eğer biliniyorsa, sorumlu grup veya kişi hakkında bilgi mevcuttur. Veri tabanında kayıtlı her olay için, erişilebilir verinin durumuna göre, en az 40 öznitelik için veri sunulmaktadır. Yakın tarihli olaylar için öznitelik sayısı 120'yi geçebilmektedir. GTD veri tabanı 1970'ten bu yana meydana gelmiş olan 95.000'den fazla bombalama, 20.000 suikast ve 15.000 adam kaçırmaya ve rehin alma olayı hakkında bilgi içerir. Veri tabanındaki veriler farklı haber makaleleri ve haber kaynaklarından toplanmıştır.

Sunulan web uygulaması üzerinden çeşitli kriterlere göre arama yapılabildiği gibi verinin tamamının indirilmesine de imkân sağlamaktadır. Veri

üzerinde şu kriterlere göre arama yapılabilir: Tarih aralığı, bölge, şehir, terör olayı tipi, kullanılan silah tipi, hedef tipi, fail grubu, kayıplar. Örneğin 2017-2019 tarihleri arasında Türkiye’de patlayıcı olaylarının dökümünü almak için tarih aralığı, ülke ve kullanılan silah tipi kriterleri için ilgili bilgilerin seçilmesi ile sonuç alınabilmektedir.

b. Action on Armed Violence (AOAV) Veri Tabanı

AOAV web sitesi (AOAV, 2021) patlayıcı silahlarla gerçekleştirilmiş olan terör ve saldırı olayları hakkında bilgilerin yanı sıra üzerinde arama yapılabilen bir veri tabanı içermektedir. Veri tabanında, 2010-2021 yılları arasında meydana gelmiş 22.000’i aşkın terör ve saldırı olayına ilişkin veriler yer almaktadır. AOAV veri tabanı İngilizce haber kaynakları taranarak oluşturulmaktadır. GTD ile benzer bir yapısı olmakla birlikte, patlayıcı silahlarla gerçekleştirilen terör olaylarına odaklanması açısından farklılaşmaktadır.

GTD veri tabanı ile benzer şekilde arama yapılabilecek belirli kriterler bulunmaktadır. Bu kriterlerden üçü (tarih, ülke, kayıplar) zorunlu doldurulma alanı, dokuzu (saldırıda bulunan grup, saldırıda bulunanların durumu, meskûn yerleşim yerinde gerçekleşip gerçekleşmediği, olay yeri, hedef kitle, kayıplara ait yaş ve cinsiyet, silah tipi, saldırı tipi, saldırını nereden başlatıldığı) ise seçimli doldurulma alanıdır. Örneğin Türkiye’de 2017-2019 yılları arasında meydana gelen “El Yapımı Patlayıcı” olayları ve etkilenen askerî birimlerin sayısı alınabilmekte ve ilgili birimler bu kriterlerin belirtilmesi ile listelenebilmektedir. Ayrıca bütün sonuçlar Excel dosya formatında da alınabilmektedir.

c. Data.world Terörizm Veri Kümeleri

Data.world (DW, 2021) çeşitli alanda veri kümelerinin aranabildiği, indirilebildiği ve analiz edilebildiği bir veri depolama sitesidir. Bu sitenin avantajlarından biri bünyesinde birden fazla kaynaktan ve farklı konular için (finans, terörizm, ekonomi, Twitter, NASA (Ulusal Havacılık ve Uzay Dairesi) ve diğerleri) çeşitli veri kümelerini barındırmasıdır. Bunun yanı sıra, kullanıcıların kendi veri kümelerini yüklemeleri ve paylaşımlarına da olanak verilmektedir. Bu amaçla 53 farklı uygulama ve internet sitesi ile entegrasyonu bulunmaktadır. Veri üzerinde arama yapmak için site ara yüzünde SQL (Yapılandırılmış Sorgu Dili) ve SPARQL

(Protokol ve Sorgu Dili) sorguları yazılabilmektedir. Ayrıca, R ve Python programlama dilleri ile seçilen veriler üzerinde analiz yapmayı da desteklemektedir.

Makalemizin araştırma konusu olan *terörizm* sözcüğü ile arama yapıldığında 107 sonuç listelenmiştir. Bu sonuçların 84 tanesinin veri kümesi, 12 tanesinin proje ve analiz, 11 tanesinin de yorum olduğu gözlenmiştir. Sunulan 84 adet veri kümesi .xls, .csv gibi farklı veri formatlarında olabilmektedir. Bu veri kümeleri, genel bir terörizm araştırmasına ait olabildiği gibi, belirli yıllara, belirli bir ülkede ya da belirli bir konuda özelleşmiş olan terörizm faaliyetlerini de içerebilmektedir. Bu veri kümelerine şu örnekleri verebiliriz; Terrorism, Global Terrorism Index 2016, Indian Terrorism Deaths, Terrorism Cased 2001-2016.

Sitede paylaşılmış olan veri kümelerinden *Indian Terrorism Deaths* örnek olarak incelenmiş ve şu gözlemler yapılmıştır. İlgili veriler Hindistan'da terörizm kaynaklı olaylarda yaşanan kayıpların incelenmesi için hazırlanmış olup, bu olayların tarihleri, yaşandığı yerleri, kayıp sayısı, askerî personel kaybı olup olmadığı, olayın detaylı açıklaması vb. bilgileri içermektedir. Veri kümesinde toplam 37 öznitelik (sütun) ve 27.233 örnek (satır) bulunmaktadır.

Veri kümesi üzerinde arama yaparken belli bir zaman dilimine, kayıpların sivil olmasına, yaşanan şehre ve sayıya göre filtreleme desteklenmektedir. Örneğin, arama kriterlerimizi tek bir günde, 100 ve üzeri sivil ölüm olarak kısıtlarsak 711624882 ID numarasına sahip ve Dhubri şehrinde yaşanan olaya ait veri listelenmektedir.

Ayrıca indirilmiş olan veri kümesi üzerinde anahtar sözcük tabanlı arama yapılabilmektedir. Örneğin; “El Yapımı Patlayıcı” (Improvised Explosive Device (IED)) terimi ile yapılan aramada ilgili olaylar listelenmektedir.

ç. Kaggle Terörizm Veri Kümeleri

Kaggle web sitesinin ilk kurulma amacı makine öğrenmesi yarışmaları sunan bir platform sunmaktır (Kaggle, 2021). Bu platformda çeşitli veri kümelerine ve beklenen çıktıları tanımlı çeşitli problemlere erişilebilmektedir. Veri kümelerine ve değerlendirme ölçütlerine erişimi olan kullanıcılar, kendi modellerini oluşturup, bu modelleri Kaggle web sitesine yükleyebilirler. Modellerin performansı kullanıcıların erişimi olmadığı bir test veri kümesi üzerinden değerlendirilmekte ve kullanıcıların başarısı bu test kümesinde elde ettikleri metrik değeriyle belirlenmektedir. Web

sitesinde ‘terrorism’ kelimesi ile arama yapıldığında toplam 1339 adet sonuç çıkmakla birlikte 576 kod, 551 yorum, 149 tartışma konusu, 61 dataset ve 1 blog tespit edilmiştir.

Bu sitede de Data.world sitesinde olduğu gibi indirilmiş olan veri kümesi üzerinde belli bir zaman dilimine, kayıpların sivil olmasına, yaşanan şehre ve sayıya göre filtreleme yapılabilmektedir. Ayrıca veri kümesi üzerinde anahtar sözcük tabanlı arama da desteklenmektedir. Örneğin, IED sözcüğü ile yapılan aramada ilgili olayların verileri toplu halde listelenmiş olup bu olaylara dair istatistikler sunulmuştur.

2. Makale Seçim Yaklaşımı

Bu çalışmada, birinci bölümde sunulmuş olan veri kümeleri üzerinde terör ve terörle mücadele konusunda yapılan çalışmalar incelenmiştir. Makale taraması için Google Scholar* makale veri tabanı kullanılmıştır. Her bir veri kümesi için ayrı arama yapılmış, veri kümesinin adının yanına *terrorism* sözcüğü eklenerek, örneğin; *Kaggle terrorism* şeklinde, arama sözcük grupları oluşturulmuştur. Arama sonucunda dönen makale listeleri 2019-2022 yılları için filtrelenmiş, bu sonuçlar için ilk 5 sonuç sayfası taranmıştır. Seçilen makalelerin hakemli bir dergide yayınlanmasına dikkat edilmiş olup, dergi makalesi olarak yayınlanmayan yüksek lisans ve doktora tezleri incelemeye dâhil edilmemiştir.

Arama sonucu listelenen çalışmalar için şöyle bir eleme yöntemi kullanılmıştır: Arama motorunda listelenen bazı makalelerin veri kümelerinden yararlanmadığı ve genel bir konu üzerine derleme olduğu görülmüş olduğundan bu makaleler çalışmaya dâhil edilmemiştir. Özet bilgisi ve detay içeriğine ulaşamayan makaleler de dışarda bırakılmıştır. Bunların yanı sıra, hakemli dergilerde yayınlanmamış olan çalışmalar da elenmiştir.

GTD için yapılan taramada 2019-2022 yılları için oluşturulan toplam 14000 sonuç listelenmiştir. İlk beş sayfada bulunan toplam 50 başlık incelenmiştir. Bu 50 başlıktan ilgili kelimelerin geçtiği ve konumuzla ilgili olduğunu değerlendirilen 26 makale, AOAV veri tabanı için yapılan taramada ilgili 6 makale, Data.world veri sitesi için ilgili 8 makale, Kaggle sitesi için ilgili 17 makale belirlenmiştir.

* <https://scholar.google.com.tr/>

Listelenen yayınlar göz önüne alındığında, yayınların büyük çoğunluğunun verileri GTD (Global Terrorism Database) sitesinde yayınlanan veriler olduğu görülmektedir. Bunun yanı sıra ele alınan konu ile ilgili çalışmaların GTD'den alınan verileri kullandıkları, ele aldığımız diğer veri kümelerine makalelerde sadece atıfta bulunulduğu görülmektedir.

Taranan ve ilgili bulunan çalışmalar tekrar daha detaylı incelenmiş ve GTD verileri üzerinde yapılan çalışmaların büyük bir kısmının yapay zekâ ve büyük veri tekniklerini kullanmadıkları, daha temel istatistiksel veri analizi sonuçları sundukları görülmüştür. Bunların arasında, elle alınan konuya odaklanan 10 adet çalışma detaylı inceleme kapsamına alınarak makaledeki çalışmaya dâhil edilmiştir. Bu makalelerin detayları 4'üncü bölümde sunulmaktadır.

3. Makine Öğrenmesi Tekniklerine Genel Bir Bakış

Bu kısımda incelediğimiz 10 çalışmada kullanılmış olan makina öğrenmesi teknikleri kısaca tanıtılmaktadır. Bunun yanı sıra, makine öğrenmesi modellerinin doğruluk analizi yöntemleri için temel yaklaşımlar da özetlenmektedir.

a. Gözetimli Öğrenme Algoritmaları

Makine öğrenmesi yöntemleri altında yer alan gözetimli öğrenme algoritmaları, temel olarak, etiketli (grubu belli olan) verileri kullanarak yeni verileri etiketleyebilecek, diğer bir deyişle, verinin ait olduğu grubu tahmin edecek modeller oluşturur. Etiketli veri grubunu büyük bir kısmı modeli eğitmek için kullanılır, bu veriler modelin eğitimi için, bir nevi gözetmen/öğretmen rolü üstlenir (etiketli verinin kalan kısmı modeli iyileştirmek ve test etmek için kullanılır) (Michalsky vd., 2013; Han, Pei ve Kamber, 2011).

Etiketli verilerin kullanımı ile model eğitimi için farklı algoritmalar geliştirilmiştir. *Karar Ağacı* (Decision Tree) yaklaşımları, veri etiketi tahmini için bir ağaç yapısı oluşturur. Eğitim verisi üzerinde farklı kriterler (Information Gain, Gini Index vb.) kullanılarak en belirleyici öznitelikler belirlenir. Bunlar oluşturulan karar ağacının düğümleri olarak yer alırlar. Karar ağacının yaprak düğümleri ise etiket değerlerini taşır. Yeni gelen bir verinin öznitelik değerleri Karar Ağacının yapısı ile karşılaştırılarak veri için etiket belirlenir.

Rastgele Orman (Random Forest), farklı karar ağaçlarını bir arada kullanarak tahmin doğruluğunu iyileştirmeyi hedefler. Esasen birden çok gözetimli

öğrenme (sınıflandırma) modelini bir arada kullanan bu tip yaklaşımlar, *Topluluk* (Ensemble) model olarak adlandırılır. Bir diğer topluluk modeli olan AdaBoost, zayıf öğrenici (weak learner) adı verilen farklı tipte sınıflandırıcılar üzerinde uygulanabilmekle birlikte, klasik olarak karar ağacı modelleri ile birlikte kullanılmaktadır.

Bir diğer gözetimli öğrenme algoritması olan *Naif Bayes Sınıflandırıcı* Bayes Teoremini kullanarak bir verinin eldeki gruplara ait olma olasılığını hesaplar. Olasılığı en yüksek grup, veri etiket tahmini olarak sunulur. Naif Bayes Sınıflandırıcılar altında yer alan Gauss Naif Bayes Sınıflandırıcı (Gaussian Naive Bayes (GNB)) ise sayısal değer alan öznitelikler için Gauss (normal) dağılım varsayımı ile olasılık hesaplar.

K-En Yakın Komşuluk (K-Nearest Neighbor) algoritması ise etiket tahmini için eldeki verinin komşularının etiketlerini değerlendirir. Komşuluk kavramı öznitelik uzayında veriye yakın (benzer) olma durumunu ifade eder. K sayısı tahminlemede kullanılacak komşu sayısını belirtir.

Yapay sinir ağları, doğal nöron ağlarından benzetilerek oluşturulan yapay zekâ modelleridir. Sınıflandırma başta olmak üzere pek çok problem için kullanılmaktadırlar. Biyolojik nöronun bağlantıları, yapay sinir ağlarında düğümler arasındaki ağırlıklar olarak modellenir. Pozitif bir ağırlık, uyarıcı bir bağlantıyı yansıtırken, negatif değerler, engelleyici bağlantılar anlamına gelir. Tüm girdiler bir ağırlıkla değiştirilir ve toplanır. Bu yapı lineer bir kombinasyon oluşturur. Gözetimli öğrenme problemleri için bu lineer kombinasyonun çıktısı tahmini oluşturur. Doğru tahmin elde edebilmek için tahmin hatası sinir ağına geri bildirim olarak verilerek ağırlıklar iteratif olarak güncellenir. *Derin öğrenme modellerinde* sinir ağı yapısı katman sayısı artırılarak daha karmaşık durumları öğrenebilmesi amaçlanır. Bunun yanı sıra nöronların yapısı ve nöronlar arası bağlantıların yapısı da farklılaşmaktadır.

Destek Vektör Makinaları (Support Vector Machines (SVM)), temel olarak ikili sınıflandırma için geliştirilmiş bir algoritmadır. Eldeki iki kategoriden biri ile etiketli eğitim verileri ile SVM algoritması, iki sınıf arasındaki sınırı bularak ikili doğrusal sınıflandırıcı modeli oluşturur. Doğrusal olarak ayrılmayan veriler için girdileri örtük olarak yüksek boyutlu özellik uzaylarına eşleyerek sınıflandırma modeli sağlar.

Doğrusal Ayrım Analizi (Lineer Diskriminant Analizi (LDA)), iki veya daha fazla sınıfı karakterize eden veya ayıran özelliklerin lineer bir kombinasyonunu bulmak için istatistiksel bir yöntemdir. Özniteliklerin lineer kombinasyonu ile oluşturulan model, doğrusal bir sınıflandırıcı olarak kullanılmaktadır.

İstatistiksel modellemede, *Regresyon Analizi*, bir bağımlı değişken (etiket/sonuç) ile bir veya daha fazla bağımsız değişken arasındaki ilişkileri tahmin etmeye yönelik bir dizi istatistiksel süreçtir. Regresyon analizinin en yaygın biçimi, belirli bir matematiksel kritere göre verilere en yakından uyan doğrunun (veya daha karmaşık bir lineer kombinasyonun) bulunduğu *Lineer Regresyon*dur. *Lojistik Regresyon* ise, temel biçiminde, ikili bağımlı değişkeni modellemek için bir lojistik fonksiyon kullanan bir modeldir. Genellikle geçti-kaldı, sağlıklı-hasta gibi belirli bir sınıfın veya olayın var olma olasılığını modellemek için kullanılır.

b. Gözetimsiz Öğrenme Algoritmaları

Gözetimsiz öğrenme algoritmalarında yukardaki özetlenen algoritmalarından farklı olarak etiketli eğitim verisi kullanmaz. Veri içindeki mevcut gruplanmaları ve örüntüleri bir eğitmen olmaksızın ortaya çıkartmayı hedefler (Michalsky vd., 2013; Han, Pei ve Kamber, 2011).

Gözetimsiz öğrenme tekniklerinden en bilinen kümelemedir. Kümeleme algoritmaları veri örnekleri arasındaki benzerlik ölçümlerini kullanarak en benzer örneklerin bir araya geldiği kümeleri oluşturmayı hedefler. Öte yandan bir diğer hedef de oluşan kümelerin (kümelere ait elemanların) birbirinden oldukça farklı olmasıdır. Bu hedefi sağlamak üzere çeşitli kümeleme algoritmaları geliştirilmiştir. Bunlardan *Agglomeratif Hiyerarşik Kümeleme* (Agglomerative Hierarchical Clustering) her veri örneğinin tek başına bir küme oluşturduğu varsayımından başlayarak verilen veri ve küme benzerliği metriklerini kullanarak en benzer kümeleri seviye birleştirerek bir küme hiyerarşisi oluşturur. Oluşturulan hiyerarşik yapıda eldeki problemin/verinin doğasına uygun bir seviye seçilerek kümeler oluşturulur. Diğer bir kümeleme algoritması olan *K-Ortalama* (K-Means) *Kümelemede* ise elde edilmesi istenen küme sayısı k parametresi ile belirtilir. Buna göre her verinin kendi kümesinin uzaklığının toplamını minimize edecek bir küme yapısı oluşturulması hedeflenir. Bu hedefi sağlamak üzere sezgisel bir yöntemle iteratif olarak verilerin ait olduğu kümeler gözden geçirilerek optimal küme yapısına varılması amaçlanır.

Gözetimsiz öğrenme başlığı altında yer alan bir diğer metot da *Sık Örüntü (ya da Birliktelik Kural) Analizidir*. Bu metotta veri içinde birlikte görülme sıklığı verilen bir eşik değerinin üzerinde olan örüntülerin bulunması hedeflenir. En bilinen uygulama örneği, alışveriş sepetleri içinde sıklıkla birlikte alınan ürünlerin (alt ürün sepetlerinin) bulunmasıdır. Bulunan bu sık örüntüler üzerinden ilişkisel kurallar oluşturulur.

c. Model Doğruluğu Ölçme Yöntemleri

Gözetimli öğrenme modellerinin doğruluk analizi model tahminlerinin bir test veri grubunun gerçek etiketleri ile karşılaştırılması ile sağlanır. Bu karşılaştırma ile her bir örnek için tahminlemenin doğru yapıp yapılmadığı bulunur. Bu karşılaştırma tüm test verisi için uygulanarak oransal bir doğruluk bulunur. Bu kapsamda en bilinen tahmin metriği *doğruluk* (accuracy) ölçümüdür. Doğruluk eldeki test örneklerinin etiketlerinin yüzde olarak ne kadarının doğru tahmin edildiğini belirtir. Kimi zaman eldeki tüm etiketler yerine tek bir etiket için tahmin ölçümü yapılabilir. Bu durumda *doğru tespit oranı* (recognition rate / precision) sıklıkla kullanılan bir ölçüm yöntemidir. Bu ölçüt, eldeki veri grubu için doğru yapılan tahminlerin, aynı etiket için yapılan tüm tahminlere oranı şeklinde hesaplanır (Tan vd., 2016; Han, Pei ve Kamber, 2011).

İncelediğimiz çalışmalarda da kullanılmış olan bir diğer ölçüm yöntemi de *Alıcı İşletim Karakteristik Eğrisi Altında Kalan Alan* (Area Under the Receiver Operating Characteristic curve (AUROC)) ölçütüdür. ROC eğrisi, çeşitli eşik ayarları altında sınıflandırma performansını gösteren bir olasılık eğrisidir. Bu eğri altında kalan alan oluşturulan modelin sınıfları ne kadar ayırt edebildiğini söyler (Tan vd., 2016; Han, Pei ve Kamber, 2011).

4. İncelenen Çalışmalar

Literatür taraması sonucu ilgili bulunan çalışmalar üzerinde tekrar yapılan incelemede, doğrudan GTD verileri üzerinde yapılan çalışmaların pek çoğunda makalede odaklandığımız şekilde yapay zekâ ve büyük veri tekniklerinin kullanılmadığı, daha temel veri analizi sonuçlarının paylaşıldığı görülmüştür. Bu makaleler arasında, yapay zekâ teknikleri kullanarak analiz ve örüntü bulma konularını ele alan çalışmaları belirlenmiştir. Bu kısımda, bu çalışmalar detaylı olarak sunulmaktadır. Tablo 2’de incelenen bu çalışmaların künyeleri verilmiştir.

Tablo 2. Detaylı İncelenen Çalışmaların Künyeleri

Atıf Künyesi	Çalışma Künyesi
Atsa'am, Wario ve Okpo, 2020	A New Terrorism Categorization Based On Casualties And Consequences Using Hierarchical Clustering (Hiyerarşik Kümelemeyi Kullanan Zayıt ve Sonuçlarına Dayalı Yeni Bir Terörizm Kategorizasyonu)
Vaibhav, 2020	Predicting Success Of Terrorist Attack And Extent Of Its Economic Impact Using Data Mining (Veri Madenciliğini Kullanarak Terörist Saldırının Başarısını ve Ekonomik Etkisinin Kapsamını Tahmin Etme)
Singh, Chaudhary ve Kaur, 2019	A Machine Learning Approach For Enhancing Defence Against Global Terrorism (Küresel Terörizme Karşı Savunmayı Geliştirmek İçin Makine Öğrenimi Yaklaşımı)
Huamaní, Mantari ve Roman-Gonzalez, 2020	Machine Learning Techniques To Visualize And Predict Terrorist Attacks Worldwide Using The Global Terrorism Database (Küresel Terör Veri Tabanını Kullanarak Dünya Çapında Terörist Saldırıları Görselleştirmek ve Tahmin Etme İçin Makine Öğrenme Teknikleri)
Bhatia, Chhabra, ve Kumar, 2020	Data Analysis Of Various Terrorism Activities Using Big Data Approaches On Global Terrorism Database (Küresel Terörizm Veri Tabanında Büyük Veri Yaklaşımlarını Kullanarak Çeşitli Terörizm Faaliyetlerinin Veri Analizi)
Narula, Kumar, Kaul ve KS, 2020	Predictive Analysis On Global Terrorism (Küresel Terörizm Üzerine Öngörü Analizi)
Iorliam vd., 2021	An Investigation And Insight Into Terrorism In Nigeria (Nijerya'daki Terörizm Üzerine Bir İnceleme ve İlgörü Çalışması)
Salem, Naouali ve Chtourou, 2020	Clustering Terrorist Groups Using Their Activity And Lethality Measures (Terörist Gruplarının Aktivite ve Ölümcüllük Ölçütlerini Kullanarak Kümelenmesi)
Krieg vd., 2022	Predicting Terrorist Attacks In The United States Using Localized News Data (Yerleştirilmiş Haber Verilerini Kullanarak ABD'deki Terör Saldırıların Tahmin Etme)
Wang, Cui ve He, 2022	Winning the War On Terror: Using "Top-K" Algorithm And CNN To Assess The Risk Of Terrorists (Teröre Karşı Savaşı Kazanmak: Teröristlerin Riskini Değerlendirmek İçin "Top-K" Algoritması ve CNN'i Kullanan Bir Çözüm)

a. İncelenen Çalışmalarda Ele Alınan Problemler ve Kullanılan Yöntemler

Çalışma kapsamında taranmış olan makaleler arasında ele alınan problemle ilgili olarak belirlenmiş olan Tablo 2'deki çalışmalar, bu kısımda detaylı olarak sunulmaktadır.

Atsa'am, Wario, ve Okpo (2020) tarafından yapılan çalışmada, GTD veri kümesi üzerinde veri madenciliği ve makine öğrenmesinin temel örüntü bulma yöntemlerinden bir olan kümeleme tekniği kullanılarak terör olaylarına ilişkin örüntülerin tespiti amaçlanmıştır. Kümeleme tekniklerinde, veri için belli gruplar belirtmeden verinin doğasından gelen kümelenmelerin bulunmasına odaklanılır. Çalışmada, terör olayları, meydana gelen zayıat (casualty) ve ortaya çıkan sonuçlar üzerinden analiz edilmiştir. Yani, her bir olay, zayıat ve olayın sonuçlara karşılık gelen öznelilikler seçilerek bir öznelilik vektörü olarak tanımlanmıştır. GTD verisinde, terör olayları için öldürülen faillerin sayısı (nkill), yaralı sayısı (nwound), rehine sayısı (nhostkid), talep edilen fidye miktarı (ransomant), ödenen fidye miktarı (propvalue) gibi zayıat ve sonuca ilişkin öznelilikler incelenmiştir. Eksik veri girişleri nedeniyle yapılan filtreleme sonucunda nkill, nwound, propvalue, nhostkid öznelilikleri seçilmiş ve bu öznelilikler için eksik veri bulunmayan (tam gözleme sahip) 168 terör olayı için öznelilik vektörü oluşturulmuştur.

Veri kümeleme için Aglomeratif Hiyerarşik Kümeleme tekniğinin kullanıldığı belirtilmiştir. Bu kümeleme algoritması her bir veri örneğinden başlayarak gruplama yapar. Öncelikle benzerliği en yüksek olan veriler küçük kümeler oluşturur. Daha sonra veriler benzer kümelere katılır, ya da küçük kümeler benzerliklerine göre birleşerek daha büyük bir küme oluştururlar. Böylece kümeleme adımları hiyerarşik bir yapı oluşturur. Çalışma kapsamında Aglomeratif Hiyerarşik Kümeleme algoritması R istatistik programlama dili kütüphaneleri kullanılarak uygulanmıştır (Thomas, Vaughan ve Lello, 2013).

Vaibhav (2020) tarafından sunulan çalışma, terör saldırılarında maddi hasarın boyutunu tahmin problemi üzerinde çalışılmıştır. Veri kaynağı olarak yine GTD veri tabanı kullanılmıştır. Maddi hasar tahmin problemi bir gözetimli öğrenme problemi olarak tanımlanmıştır. Gözetimli öğrenme tekniklerinde bilinen sınıf etiketleri bulunur ve bu etiketlere sahip eğitim verileri ile sınıflandırma modeli oluşturulur. Bir diğer deyişle, bilinen hasar seviyeleri ile terör olaylarının

öznitelikleri üzerinden bir eşleştirme yapılmaktadır. Gözetimli öğrenme tekniği olan Sınıflandırma Yaklaşımı kullanılmış, bu amaçla Karar Ağacı (Decision Tree), Rastgele Orman (Random Forest) algoritmalarının tahmin performansı incelenmiştir. Terör olaylarını vektör uzay modelinde ifade etmek üzere GTD veri tabanında bulunan 135 öznitelik arasında eleme yapılarak şu öznitelikler seçilmiştir: Olayın meydana geldiği yıl, ay, gün, ülke, bölge, olayın konum koordinatları (enlem ve boylamı), öldürülen faillerin sayısı (nkill), saldırı türü, kullanılan silah türü, bireysel saldırı olup olmadığı, intihar saldırısı olup olmadığı. Makalede ayrıca, Sentetik Azınlık Aşırı Örnekleme tekniğinin (SMOTE) sınıflandırma modellerinin sonuçları üzerindeki etkisi de incelenmiştir.

Singh, Chaudhary ve Kaur (2019) tarafından yapılan çalışmada amaç makine öğrenimi yaklaşımlarını kullanarak bir terör saldırısının bölge ve ülkesini tahmin etmektir. Çalışma, GTD veri tabanında yer alan 1970-2017 yılları arasındaki terör olaylarının verilerini kullanmıştır. Çalışma kapsamında 88000 bombalama, 19000 suikast ve 11000 adam kaçırmaya olayı da dahil olmak üzere toplam 180000 terör olayı incelenmiştir. Önceki çalışmalarda olduğu gibi her bir terör olayı vektör uzay modeli kullanılarak ifade edilmiş, bunun için GTD veri tabanında yer alan olay tarihi, saldırı ayı, olayın yeri ve olayın olduğu ülke, olay bölgesi, olayda kullanılan silahlar, hedefin niteliği, saldırı türü, yaralı sayısı, olaydan sorumlu grup veya kişi bilgileri arasında filtreleme yapılarak saldırı ayı, hedef türü ve saldırı türü öznitelik olarak kullanılmıştır. Bu çalışmada da ele alınan tahminleme problemi gözetimli öğrenme problemi olarak tanımlanmış ve sınıflandırma teknikleri kullanılmıştır. Tahminleme için altı gözetimli makine öğrenimi modeli Gauss Naif Bayes (Gaussian Naive Bayes (GNB)), Doğrusal Ayrımcılık Analizi (Linear Discriminant Analysis (LDA)), K-En Yakın Komşuluk (K-Nearest Neighborhood), Destek Vektör Makinaları (Support Vector Machines (SVM)), Karar Ağacı ve Lojistik Regresyon uygulanmış ve ardından performansları değerlendirilmiştir.

Huamaní, Mantari ve Roman-Gonzalez (2020)'in araştırmasında, GTD veri tabanında yer alan ve 1970'ten günümüze kadar dünyada meydana gelen terör saldırılarının görselleştirilmesi ve tahmin edilmesi amaçlanmıştır. Çalışmada, bölgeye ve terör saldırısı türüne göre saldırı sayısının tahmini problemi ele alınmış, tahmin modeli oluşturmak için Karar Ağacı ve Rastgele Orman sınıflandırma algoritmaları kullanılmıştır. Makalede tahminleme modeli oluşturmak üzere izlenen

veri temizleme, ön işleme ve parametre ayarlama ile tahmin başarımının iyileştirilmesi için izlenen veri madenciliği adımları ana hatlarıyla sunulmuştur.

Bhatia, Chhabra ve Kumar (2020), GTD veri tabanı üzerinde yaptıkları analizlerde, incelediğimiz diğer çalışmalardan farklı olarak büyük veri işleme teknolojilerine odaklanmışlardır. Bu amaçla, büyük veri işleme teknolojilerinden Hadoop modeli kullanılmıştır (Dittrich ve Quiané-Ruiz, 2012). Böylece GTD veri tabanındaki veriler üzerinde farklı sorguların cevapları hızlı bir şekilde alınarak veri içindeki ilginç bulguların tespit edilmesi amaçlanmıştır. Çalışma yöntemi açısından, incelediğimiz makine öğrenmesi odaklı makalelerden bu noktada farklılaşmaktadır. Veri içindeki örüntülerin makine öğrenmesi teknikleri yerine hızlı ve farklı sorgularla elde edilmesi hedeflenmiştir. GTD verileri eksik girdilerin temizlenmesi sonrası Hadoop HIVE ortamında saklanmış, hazırlanarak bir Linux sistemi üzerine kurulu olan Hadoop ortamı ile bütünleşmiş bir şekilde analiz edilmiştir. Veri üzerinde önceden belirli sorguları çalıştırmak için Hive Sorgu Dili (Hive Query Language (HiveQL)) kullanılmıştır. Büyük veri işleme altyapısı üzerinde sorgulama yoluyla elde edilen analiz sonuçları, tablolar ve grafikler halinde sunulmuştur.

Narula, Kumar, Kaul ve KS (2020) tarafından yapılan çalışmada GTD verileri üzerinde terör saldırıları analizi için iki alt problem ele alınmıştır. Bunlar: terör olayında kullanılan silah tahmini ve terör olayı failinin tahminidir. Her iki problem de gözetimli öğrenme problemi olarak tanımlanarak bilinen sınıflandırma algoritmaları olarak Karar Ağacı, Rastgele Orman ve K-En Yakın Komşuluk kullanılmıştır. Ele alınan problemler özelinde düşünüldüğünde, örneğin terör olayı faili tahmini için, ele alınan terör olayının bilinen terör gruplarından hangisi tarafından gerçekleştirildiğinin tahmini yapılmaktadır. Çalışmada her iki alt problem için de Rastgele Orman sınıflandırma algoritmasının diğerlerine göre daha başarılı tahmin performansı sunduğu belirtilmiştir. Ancak veri üzerinde yapılan ön işlemler, kullanılan verinin nihai büyüklüğü ve sayısal tahmin sonuçları çalışmada sunulmamıştır.

Iorliam vd. (2021) tarafından yapılan çalışmada, terörizmin, dünya çapında insanlığın karşı karşıya olduğu en önemli tehditlerden biri olduğu, terör örgütlerinin faaliyetlerinin barışı tehdit eden, ülkelerin gelişmesini olumsuz etkileyen bir unsur olduğu vurgulanmaktadır. Örnek olarak Nijerya kullanılmıştır. Bu ülkedeki terör faaliyetlerinin son on yılda ekonomik büyümeyi olumsuz etkilediği ve ülkeye

yapılan yabancı yatırım olanaklarını büyük ölçüde azalttığı belirtilmektedir. Vurgulanan konunun sayısal sonuçlarla desteklenmesi amacıyla, çalışmada, 1970'ten 2019'a kadar Nijerya'da meydana gelen terörist faaliyetler incelenmiştir. GTD verileri kullanılarak, terör saldırılarının oluşumları, hedeflerin yerleri, başarılı veya başarısız terör olayları hakkında tespitler yapılmıştır. Bu tespitler yapılırken sık örüntü analizi tekniklerinden biri olan Apriori algoritması kullanılmıştır. Bu algoritma ile veri içindeki öznitelikler arasındaki birlikte görülme ilişkileri ortaya konmuştur. Böylece terörist faaliyetlerin Nijerya'da durdurulmasına yönelik doğru kararları vermede güvenlik güçlerinin teröristlerden bir adım önde olmaları için gerekli bilgileri ve örüntülerin elde edilmesi amaçlanmıştır.

Salem, Naouali ve Chtourou (2020)'nin çalışmasında, GTD verileri kullanılarak 1970-2016 yılları arasında Kuzey Afrika ve Ortadoğu'da meydana gelen terörist saldırı incelenmiştir. Analiz yaklaşımı olarak kümeleme yöntemi kullanılmıştır. K-Ortalama algoritması ile yapılan olay kümelemesi için her olay aktiflik ve öldürücülük öznitelikleri ile ifade edilmiştir. Bu amaçla, terör gruplarının aktifliğini ve öldürücülüğünü gösteren iki parametre önerilmiştir. Aktiflik parametresi, terörist grubun kurulduğu zamandan itibaren her yıl için ortalama saldırı sayısı olarak tanımlanmıştır. Öldürücülük parametresi ise belirlenen zaman periyodu içinde bir terörist grubunun saldırılarının zayıflığının aynı dönemde meydana gelen bütün terör olayları zayıflığına oranlanması şeklinde tanımlanmıştır.

Krieg, Smith, Chatterjee ve Chawla (2022)'nin sunduğu çalışmada, Amerika Birleşik Devletleri'ndeki terör olaylarını tahmin etmek için yerel haber verilerini kullanan gözetimli öğrenme modeli oluşturulmuştur. Oluşturulan model belirli bir tarihte ve belirli bir eyalette terör olayı olup olmayacağını tahmin etmektedir. Model, GTD veri tabanından yerel haber verilerini kullanarak eğitilmiştir. Olayların tarih ve eyalet etiketleri GTD verilerinden elde edilmiştir. Çalışma en çok saldırının gerçekleştirildiği eyaletler olan New York, California, Texas, Florida ve Washington için 2015-2018 yıllarının verilerini kullanmıştır. GTD altındaki olay veri tabanından alınan olayın aktörü, tipi gibi özniteliklerin yanı sıra duygu analizi ile elde edilen öznitelikler de kullanılmıştır. Gözetimli öğrenme modeli olarak Rastgele Orman ve AdaBoost topluluk (ensemble) modellerinin yanı sıra Klasik İleri Beslemeli Yapay Sinir Ağı (Feed Forward Neural Network) ve Kısa Uzun Vadeli Hafıza (Long Short Term Memory (LSTM)) sinir ağı modelleri kullanılmıştır.

Wang, Cui ve He (2022) tarafından sunulan çalışmada, terörle mücadelede erken uyarı stratejileri geliştirmek için kullanılmak üzere teröristlerin saldırı riski değerlendirmesi üzerinde çalışılmıştır. Risk değerlendirmesi bir tahmin problemi olarak ele alınmış ve risk seviyeleri kırmızı, turuncu ve sarı olarak belirtilmiştir. Buna göre kırmızı risk, gerçekleşmesi halinde en yüksek zararı ve etkiyi vermesi beklenen durumu ifade etmektedir. Çözüm için yapay sinir ağı tabanlı bir risk tahmin modeli önerilmiştir. Yapay sinir ağı mimarisi olarak, Evrişimli Sinir Ağları (Convolutional Neural Network (CNN)) temel alınmıştır. CNN, derin öğrenme kategorisinde yer alan, evrişim hesaplamalarını içeren, ileri beslemeli sinir ağı modelidir. Görüntü analizi ve öznitelik oluşturma başta olmak üzere pek çok problemde başarılı sonuçlar sağlamaktadır. Sunulan yöntemde yazarlar tarafından oluşturulan “Top-K” algoritması ve CNN mimarisi birlikte kullanılmıştır. Bu yöntemde öncelikle bilinen terör örgütlerinin öznitelikleri ikili (binary) bir kod ile ifade edilmiştir. İkinci adım olarak, "Top-K" algoritması ile terörle ilgili en belirleyici k öznitelik ilgi derecesine göre sıralanarak bir öznitelik serisi oluşturulmuştur. Matris yapısındaki öznitelik verileri ikili bir görüntüye benzer şekilde CNN modeline girdi olarak verilerek, risk tahmini, modelin çıktısı olarak elde edilir.

b. Çalışmalar Kapsamında Alınan Sonuçlar

Atsa'am, Wario ve Okpo (2020) tarafından yapılan çalışma kapsamında 168 terör olayı verisi üzerinde Aglomeratif Hiyerarşik Kümeleme uygulanmış ve Tablo 3'te sunulduğu şekilde 4 küme elde edilmiştir.

Tablo 3. Aglomeratif Hiyerarşik Kümeleme Sonucu

Kümelere	Küme 1	Küme 2	Küme 3	Küme 4	Toplam
Veri sayısı	136	5	25	2	168

(Kaynak: Atsa'am vd., 2020).

Bu çalışmanın öne çıkan yönü, terör saldırılarının gerçekleştiği coğrafi bölge, kullanılan saldırı aracı gibi bilinen ve grupların önceden belli olduğu

gruplamaların dışında bir yaklaşım kullanmasıdır. Atsa'am vd. (2020) tarafından sunulan, meydana gelen zayıata ve oluşan sonuca göre yapılan bu kümelemelerin, bütçeleme ve kaynakların dağıtımını açısından küresel terörle mücadele stratejisine katkı sağladığı vurgulanmıştır. Yapılan zayıat ve sonuç tabanlı bu analiz, terör gruplarının yıkıcılık düzeylerinin belirlenmesine ve her bir grup için uygulanacak mücadele çabalarının planlanması açısından kullanım potansiyeli taşımaktadır. Kullanılan teknik açısından baktığımızda gözetimsiz öğrenme yöntemi olan kümeleme, grupların önceden tanımlanmış olmasını gerektirmediği için bir ön bilgi ihtiyacı içermez. Bu açıdan, kümeleme ile terör olaylarının doğasından gelen gruplamaları daha uygun bir şekilde bulunabilir. Çalışmanın farklı özneliliklerle ve farklı kümeleme algoritmaları kullanılarak kümeleme kalitesi açısından geliştirilmesi olasıdır.

Vaibhav (2020) çalışmasında terör olaylarının maddi hasar tahmini için oluşturulan Karar Ağacı ve Rastgele Orman modelleri üç farklı bölge için ayrı ayrı incelemiştir. Çalışmada incelenen bölgeler şunlardır: Kuzey Amerika, Güney Asya, Orta Doğu ve Kuzey Afrika. Olaylara yol açan faktörler ve senaryoların bölgelere göre farklılık göstermesi nedeniyle seçilen bölgelerin verileri için ayrı modeller oluşturularak tahmin analizi ortaya konulmuştur. Kullanılan iki sınıflandırma algoritması arasında Rastgele Orman algoritmasının daha yüksek tahmin başarıları sunduğu belirtilmiştir. 1970 - 2018 yılları arasındaki veriler için bu üç bölgedeki terör olaylarının maddi hasar tahmini, elde edilen en yüksek başarı oranları ve her bölge için kullanılan terör olayı örneği sayısı Tablo 4'te sunulmuştur.

Tablo 4. Maddi Hasar Tahmini Sonuçları

Bölge	Tahmin Başarısı Oranı	Örnek Sayısı
Kuzey Amerika	%83	3579
Güney Asya	%87	48266
Orta Doğu ve Kuzey Afrika	%87	53110

(Kaynak: Vaibhav, 2020).

Singh vd. (2019) tarafından yapılan çalışmada terör saldırısı bölgesi ve ülkesinin tahmini için altı ayrı sınıflandırma algoritmasının tahmin performansı karşılaştırılmıştır. Kullanılan tüm makine öğrenimi modellerinin tahmin doğruluğuna ilişkin veriler Tablo 5'te sunulmuştur. Yapılan karşılaştırmalı analizde Lojistik Regresyon, LDA, GNB ve SVM modellerinin terör saldırısının gerçekleştiği ülke ve bölge tahmininde %82'lik doğruluk sağladığı ve diğer tahminleme modellerinden daha başarılı sonuç sunduğu görülmektedir.

Tablo 5. Terör Saldırısı Ülke ve Bölge Tahmin Sonuçları

Model Adı	Tahmin Başarısı Oranı
Lojistik Regresyon, LDA, GNB, SVM	%82
Karar Ağacı, K-En Yakın Komşuluk	%73

(Kaynak: Singh vd., 2019).

Huamaní vd. (2020) çalışmasında, GTD verileri üzerine Karar Ağacı ve Rastgele Orman algoritmaları ile bölge ve saldırı türüne göre tahminleme sonuçları incelenmiş ve sonuçların görselleştirilmesi üzerinde durulmuştur. Yapılan analizlerde, terör saldırılarında 2012'den 2018'e kadar bir artış olduğu belirlenmiştir. Son üç yılda sayıda azalma olmasına rağmen, önceki on yıllara kıyasla terör saldırılarının sayısının halen yüksek olduğu belirtilmiştir. Analizlerden, en fazla terör saldırısının Orta Doğu, Kuzey Afrika ve Güney Asya'da olduğu anlaşılmaktadır. Saldırı türü olarak, silahlı saldırıyı takiben, bombalı ve patlamalı terör saldırılarının sayıca en yüksek terör saldırısı türleri olduğu belirtilmiştir. Bölgesel terör saldırısı sayısı tahminlemesi için dünyadaki ülkeler on iki bölgeye ayrılmıştır (Kuzey Amerika, Orta Amerika ve Karayipler, Güney Amerika, Doğu Asya, Güneydoğu Asya, Orta Asya, Doğu Avrupa, Batı Avrupa, Orta Doğu ve Kuzey Afrika, Sahra-altı Afrika, Avustralya ve Okyanusya). Makalede her bölgede yer alan ülkeler detaylı olarak belirtilmiştir. Belirtilen bölgeler için ayrı ayrı saldırı sayılarının tahmini ortaya konmuştur. Saldırı türü için GTD verilerinden hareketle 9 tür belirlenmiştir (suikast, silahlı saldırı, bombalama/patlama, uçak/gemi vs.

kaçırma, rehin alma, adam kaçırma, tesise saldırı, silahsız saldırı, diğer). Her iki problem için de elde edilen ortalama tahmin doğruluğu Tablo 6'da sunulmuştur. Bu çalışmada da incelenen diğer çalışmalarda olduğu gibi Rastgele Orman algoritması ile daha yüksek bir tahmin doğruluğu elde edilmiştir. Bölgeler için saldırı sayılarının ve tahminlerin görselleştirilmesi için haritalarda bölgelerin renklendirilmesine dayalı bir görselleştirme yaklaşımı kullanılmıştır.

Tablo 6. Saldırı Sayısı ve Saldırı Türü Tahmin Sonuçları

	Tahmin Başarı Oranı	
	Karar Ağacı	Rastgele Orman
Bölgesel Terör Saldırısı Sayısı	%75,45	%89,54
Saldırı Türüne Göre Terör Saldırısı Sayısı	%79,24	%90,41

Kaynak: (Huamani vd., 2020).

Bhatia, Chhabra, ve Kumar (2020), Hadoop sistemine aktarılmış olan GTD verileri üzerinde, HiveQL sorgu dili ile tanımlanmış çeşitli sorgular çalıştırarak terör olaylarına ilişkin yeni bilgiler elde etmeyi amaçlamıştır. Elde edilen sorgu cevapları incelendiğinde, terör olaylarından en çok etkilenen ülkenin Irak olması nedeniyle, meydana gelen kayıpların en çok Ortadoğu'da olduğu sonucuna varılabilmektedir. Ayrıca, ilgili sorgularda, terör olaylarında en yüksek hasarın İngiltere'ye verildiği tespit edilmiştir. Bununla birlikte, saldırıların sıklığının verilen hasarın büyüklüğünü ortaya koymadığı sonucuna varılmıştır. Saldırıları için en çok kullanılan silahın bilinmeyen patlayıcılar (el yapımı patlayıcılar) olduğu tespit edilmiş ve ilgili bilginin internette kolayca bulunabileceği gerçeği göz önüne alındığında, elde edilen sorgu sonucu gerçekçi bulunmuştur. Yıllar geçtikçe terör oranında herhangi bir azalmanın görülmemesi, terör krizinde ve dünyadaki durumda bir iyileşme olmadığını

göstermektedir. Makalede, gelecek saldırı konumlarının ve saldırı türü tahmini için makine öğrenmesi algoritmalarından faydalanılması önerilmiştir (Bahita vd., 2020).

Iorliam vd. (2021) tarafından yapılan çalışmada Nijerya'daki terörist faaliyetleri Apriori algoritması kullanılarak incelenmiş ve terör olaylarına ilişkin çeşitli kural örüntüleri oluşturulmuştur. Yapılan çalışmada 2011-2019 yılları arasında terörist faaliyetlerin zirve noktasına ulaştığı görülmektedir. Elde edilen kural örüntülerine göre, şahıslara ve özel mülkiyete yapılan saldırılar yüksek oranda yer almaktadır ve meskûn yerlerde meydana gelme olasılıkları yüksektir. Bir diğer ilişki kural da Borno eyaletinde meydana gelen terör saldırıları yüksek oranda tekil olaylar şeklinde meydana gelmektedir. Gelecekteki çalışmalarda, Nijerya'daki terör saldırılarının araştırılmasında Benford yasası ve Zipf yasası kullanılarak analizler yapılması planlanmaktadır.

Salem, Naouali ve Chtourou (2020)'nin çalışmasında, Kuzey Afrika ve Orta Doğu'da bulunan terörist grupların aktivite ve öldürücülük parametrelerine göre K-Ortalama algoritması ile kümeleme sonuçları, verideki mevcut özniteliklerle kümeleme sonuçları ile karşılaştırılmıştır. Manuel olarak yapılan karşılaştırmaya göre önerilen parametrelerle yapılan kümelemenin birbirine benzer terör gruplarını daha iyi belirlediği belirtilmektedir.

Krieg vd. (2022) sunduğu çalışmada, Rastgele Orman ve AdaBoost topluluk modellerinin eğitimi için 3000 tahminci ve ağaç oluşturma aşamasında öznitelik seçme ölçütü olarak Gini Index kullanılmıştır. İleri Beslemeli Sinir Ağı için her seviyede 8000 nöron bulunan iki seviyeli bir mimari oluşturulmuştur. LSTM modeli içinse her seviyede 1024 nöron bulunmaktadır. Analizlerde kullanılan veri kümesinde, seçilen eyaletlerde 18 Şubat 2015 sonrası meydana gelmiş 208 olay yer almaktadır. Analizler 5-kat çapraz doğrulama ile Alıcı İşletim Karakteristik Eğrisi Altında Kalan Alan (Area Under the Receiver Operating Characteristic Curve (AUROC)) ölçütü kullanılarak yapılmıştır. Her model için alınan en iyi sonuçlar Tablo 7'de özetlenmektedir.

Tablo 7. Terör Olayı Tahmin Doğruluğu Oranları

Eyalet Model	New York	California	Texas	Florida	Washington
Rastgele Orman	% 68,50	% 50,40	% 72,10	% 68,50	% 66,70
AdaBoost	% 66,80	% 46,30	% 58,90	% 47,20	% 54,40
İleri Beslemeli Yapay Sinir Ağı	% 42,80	% 68,00	% 65,00	% 66,70	% 64,80
LSTM	% 37,40	% 58,20	% 72,50	% 69,60	% 56,30

Kaynak: (Krieg vd., 2022).

Alınan sonuçlara göre rastgele orman modelinin incelenen eyaletler için ortalamada en yüksek tahmin doğruluğu sağladığı görülmüştür. Bununla birlikte Texas ve Florida için LSTM modelinin en yüksek performans sağlamış olması daha fazla örnek bulunması ile sonraki çalışmalarda kullanılma potansiyeli sunmaktadır.

Wang, Cui ve He (2022) tarafından sunulan çalışmada teröristlerin risk tahmin için “Top-K Algoritması” ve CNN yapay sinir ağı modeli birleştirilmiştir. Tahmin analizi anonimleştirilmiş 1200 örnek kullanılarak yapılmıştır. 1000 örnekle risk tahmin modeli eğitildikten sonra 200 örnek üzerinde yapılan testlerde kırmızı seviye risklerin %95, turuncu risklerin %91, sarı risklerin ise %93 oranında doğru tahmin edildiği raporlanmıştır. Alınan sonuçlar Tablo 8’de özetlenmiştir.

Tablo 8. Risk Tahmin Oranları

	Kırmızı Seviye Risk	Turuncu Seviye Risk	Sarı Seviye Risk
Doğru Tanıma Oranı (Recognition Rate)	%95	%91	%93

Kaynak: (Wang, Cui ve He, 2022)

5. Gözlemler ve Tartışma

Terör olaylarına ilişkin açık kaynak kodlu veri kaynakları ve büyük veri teknolojilerini kullanan çalışmalar üzerine yaptığımız incelemede göze çarpan gözlemlerimiz şunlardır:

Terörle mücadeleye yönelik yapılan veri analiz ve tahminleme çalışmalarında GTD veri tabanına kullanımının öne çıktığı görülmektedir. Terör olaylarını gruplamak, dünyanın hangi bölgesinde olayların daha çok olduğunu araştırmak, terör olaylarının nerede, ne zaman olabileceğini tahmin etmek gibi problemler için GTD'nin faydalı bir açık kaynak olduğu görülmektedir. Bu veri tabanı zengin bir içeriğe sahip olması ve alanda bilinen bir veri kaynağı olması itibarıyla veri analizi çalışmaları için tercih edilmiş olması beklenen bir sonuçtur. Ancak incelemiş olduğumuz diğer veri kaynakları da terörle mücadeleye yönelik farklı ve faydalı örüntülerin bulunabilmesi için potansiyel sunmaktadır.

GTD, 1970 ve sonraki terör olaylarını içermektedir. Terör olaylarının tarihi çok eskiye dayanmakla birlikte analiz ve tahminleme açısından, oluşturulan verinin büyüklüğü oldukça yeterlidir. Verilerin çeşitli resmi kaynaklardan ve ulusal gazetelerden toplandığı görülmektedir. Sunulan veriler terör olayının yer ve zamanının yanı sıra tipi, kullanılan silahlar, verilen zayıat ve kayıplar, olayı üstlenen suç örgütü gibi pek çok farklı bilgiyi içermektedir.

İncelen çalışmalarda çoğunlukla gözetimli makine öğrenmesi yöntemleri tercih edilmiş ve farklı tahminleme problemlerine odaklanılmıştır. Terör olayının zayıatı, bölgesi veya ülkesi ve saldırı türüne göre olay sayısının tahmini problemlerinin çalışıldığı görülmektedir.

Tahminleme problemleri için Karar Ağacı, DVM, Rastgele Orman gibi farklı yöntemlerin tahmin doğruluğu performansı incelenmiştir. Ele alınan tahmin problemlerinin büyük çoğunluğunda Rastgele Orman algoritmasının diğerlerine göre daha başarılı sonuç verdiği raporlanmıştır (Huamaní, Mantari ve Roman-Gonzalez, 2020). Karar Ağacı ile aynı prensipte çalışan ancak farklı durumları daha iyi modelleme yeteneğine sahip olan bu algoritmanın, farklı alanlardaki tahmin problemleri için başarılı sonuç verdiği görülmektedir.

2022 tarihli son çalışmalarda CNN ve LSTM gibi derin öğrenme mimarilerinin kullanıldığı görülmektedir. İncelediğimiz iki çalışmada bu mimarilerin sadece temel modelleri kullanılmıştır, alınan tahmin oranları da henüz sınırlıdır. Ancak, daha yeni ve hibrit modeller kullanılarak, sonraki çalışmalarda daha yüksek tahmin başarısı elde edilmesi olasıdır. Terörizm konusunda derin öğrenme tekniklerinin kullanılması, yapılan mücadelede kolluk kuvvetlerine önemli bir yetenek sağlayarak, yeni karar destek sistemleri ve yazılımlarının oluşmasını sağlayacaktır.

Gözetimli öğrenme kullanılan makalelerde tahmin doğruluğu tek değerlendirme metriği kullanılarak sunulmuştur. Bu çalışmaların çoğunluğu doğruluk (accuracy) metriğini kullanmıştır. Sadece iki çalışmada doğruluk dışında metrikler ile ölçüm yapılmıştır. Krieg vd. (2022) çalışmasında AUROC, Wang, Cui ve He (2022)'nin çalışmasında ise metrik olarak Doğru Tanıma Oranı (Recognition Rate) kullanılmıştır. Yöntemlerin birden çok metrik altında değerlendirilmesi, doğruluk, AUROC ve doğru tanıma oranının yanı sıra geri çağırma (recall), f1-metrik gibi daha farklı ölçüm kriterleri ile de değerlendirilmesi, hangi yöntemin kullanımına karar verilmesi açısından bizlere daha fazla veri sağlayacaktır.

Gözetimli öğrenme ile tahmin yöntemi kullanan çalışmalar ele alınan problem, kullanılan algoritmalar ve GTD'den alınan veri alt kümeleri açısından farklılık taşımaktadır. Bu açıdan çalışmalarda sunulan sonuçlar karşılaştırılabilir yapıda değildir. Ancak alınan sonuçları şöyle özetleyebiliriz: Terör olayı hasar tahmini için en yüksek tahmin başarısı Rastgele Orman algoritması ile Güney Asya, Orta Doğu ve Akdeniz için %87 olarak raporlanmıştır (Vaibhav, 2020). Terör olayının ülke ve bölge tahmini için SVB, GNB gibi çeşitli sınıflandırma algoritmalarıyla %82 tahmin başarısı elde edilmiştir (Singh vd., 2019). Terör saldırısı türüne göre saldırı sayısı tahmini için Rastgele Orman sınıflandırıcı ile %90

civarı tahmin başarısı bulunmaktadır (Huamaní vd., 2020). ABD'deki terör olaylarının yer ve zaman tahmini için farklı eyaletler için farklı algoritmalar en iyi tahminleme başarısı vermekle birlikte, AUROC metriğine göre ortalamada en yüksek tahmin doğruluğu, yaklaşık %70'lerde tahmin başarısı ile Rastgele Orman algoritması ile sağlanmıştır (Krieg vd., 2022). Terörist risk seviyesi tahmini için araştırmacıların sunduğu CNN tabanlı tahminleme modeli ile ortalama %93 doğru tanıma başarısı sunulmuştur (Wang, Cui ve He, 2022).

İncelediğimiz çalışmalar çoğunlukla gözetimli öğrenme tekniklerine odaklanmış olmakla birlikte, iki çalışmada (Atsa'am, Wario ve Okpo, 2020 ve Salem, Naouali ve Chtourou, 2020) gözetimsiz makine öğrenmesi teknikleri (Hiyerarşik Kümeleme ve K-Ortalama algoritmaları) kullanılmıştır. Bu çalışmalar, bilinen terör olayları gruplamaları dışında bir gruplama ortaya koyması ve ele alınan boyutlar açısından benzerlik gösteren olayları tespit etmiş olması açısından değer taşımaktadır. Ancak bu çalışmalarda elde edilen kümelerin kalite analizine yönelik bir ölçüm sunulmamıştır.

Gerek gözetimli, gerekse gözetimsiz öğrenme yöntemleri ile yapılan çalışmalar, terörle mücadele için fayda sağlayabilecek örüntü ve bilgilerin elde edilmesi için bu tekniklerin etkinliğini göstermektedir.

Terör olaylarına ilişkin verilerin hızlı bir şekilde sorgulanabilmesi için büyük veri mimarilerinin kullanımı da yapılan çalışmalar içinde yer almaktadır. Artan veri boyutları altında farklı sorguları hızlı çalıştırabilmek ve terörle mücadele için fayda potansiyeli taşıyan bilgilerin verimli bir şekilde elde edebilmek açısından bu teknolojilerin nasıl kullanılabilceği ve faydası ortaya konmuştur.

Sunulan açık kaynak verileri oldukça yüksek miktarda geçmiş veri içermekle birlikte, bu verilerin güncellenmesi, haber kaynaklarının sürekli olarak taranarak meydana gelen olaylarla ilgili verilerin aktarılması gayret ve maliyet isteyen işlemlerdir. Verilerin güncelliği çalışmaların devamı açısından önem taşımaktadır. Bu açıdan ilgili veri kaynaklarının güncel içeriğinin devamı zorluk ve risk unsuru oluşturmaktadır.

Zaman içinde terör gruplarında ve terör olaylarında meydana gelen yapısal değişikliklerin takip edilerek, örneğin; siber saldırılar gibi yeni terör olayı tiplerinin ve bunlara özgü özniteliklerin veri tabanlarına aktarılması önemlidir. Bu tip

değişiklik ve bunların veri kaynaklarına yansıtılıp yansıtılmaması sonraki çalışmalar için zorluk oluşturabilir.

İncelediğimiz çalışmalarda daha çok klasik makine öğrenmesi yaklaşımlarının kullanıldığı görülmüştür. Derin öğrenme tabanlı çözümlerin yakın zamanda kullanılmaya başlanmış olması bu yönde ilerlemeye işaret etmektedir. Bu açıdan yeni makina öğrenmesi algoritma ve modellerinin kullanımı için araştırma alanı bulunmaktadır.

İlgili çalışmalarda büyük veri yaklaşımlarının kullanılmaya başlandığı görülse de artan veri miktarı ve yapay zekâ tekniklerinin daha verimli uygulanabilmesi için bu alanda da yeni çalışmalara ihtiyaç olduğu değerlendirilmektedir. Gerçek zamana yakın (near real-time) analiz ve tahminler elde etmek için akan veri mimarilerinin kullanımının da fayda potansiyeli bulunmaktadır.

İncelemiş olduğumuz makalelerin özet bilgileri Tablo 9’da sunulmaktadır.

Tablo 9. İncelenen Çalışmalar

Yazarlar	Yıl Aralığı	İncelenen Problem	Kullanılan Yöntem	Alınan Sonuçlar
Atsa'am, Wario ve Okpo, 2020	1970- 2017	Terör olaylarının meydana gelen zayıf ve ortaya çıkan sonuçlar üzerinden gruplanması	Kümeleme (Aglomeratif Hiyerarşik Kümeleme)	Elde edilen kümelerin kompakt yapıda olduğu, küme içindeki olay örneklerinin uyumlu olduğu belirtilmiştir.
Vaibhav, 2020	1970- 2018	Terör olaylarının maddi hasar tahmini	Gözetimli Öğrenme (Karar Ağacı ve Rastgele Orman Algoritmaları)	Rastgele Orman sınıflandırıcısının Karar Ağacından daha iyi tahminleme performansı raporlanmıştır.
Singh, Chaudhary ve Kaur, 2019	1970- 2017	Terör saldırısının bölge ve ülkesini tahmini	Gözetimli Öğrenme (Lojistik Regresyon, Karar Ağacı, K-NN, LDA, Naif Bayes, SVM Algoritmaları)	Lojistik regresyon, LDA ve SVM ile diğerlerine göre daha başarılı tahmin doğruluğu (%82) raporlanmıştır

Tablo 9'un Devamı. İncelenen Çalışmalar

Huamaní, Mantari, Roman- Gonzalez, 2020	1970- 2017	Bölgeye ve terör saldırısı türüne göre saldırı sayısının tahmini.	Gözetimli Öğrenme (Karar Ağacı ve Rastgele Orman Algoritmaları)	Her iki tahminleme problemi için de Rastgele Orman algoritması ile daha yüksek tahminleme doğruluğuna ulaşıldığı belirtilmiştir.
Bhatia, Chhabra ve Kumar, 2020	1970- 2018	Veri tabanı sorguları ile terör saldırılarına ilişkin farklı bilgilerin çıkarımı.	Büyük Veri Mimarisi (Hadoop, HIVE, HiveQL)	Hadoop sistemi ve HiveQL sorgu dili kullanılarak terör olaylarına ilişkin çeşitli sorgular tanımlanmıştır. Buna göre terör saldırılarından en çok etkilenen ülke Irak, en çok etkilenen bölge Orta Doğu olarak tespit edilmiştir. En çok yapılan saldırı tipi ise bomba/patlayıcılarla yapılan saldırı olarak bulunmuştur.
Narula, Kumar, Kaul ve KS, 2020	1970-2020	GTD verileri üzerinde terör saldırıları analizi için iki alt problem ele alınmıştır. Terör olayında kullanılan silah tahmini ve terör olayı failinin tahmini.	Karar Ağacı, Rastgele Orman ve K-En Yakın Komşuluk Algoritmaları	Çalışmada her iki alt problem için de Rastgele Orman sınıflandırma algoritmasının diğerlerine göre daha başarılı tahmin performansı sunduğu belirtilmiştir. Ancak kullanılan verinin nihai büyüklüğü ve sayısal tahmin sonuçları çalışmada sunulmamıştır.

Tablo 9'un Devamı. İncelenen Çalışmalar

Iorliam, Dugeri, Akumba, Otor ve Shehu, 2021	1970-2019	Terör olaylarıyla ilgili ilişkisel kural örüntülerinin bulunması ve bunların incelenmesi.	Apriori Algoritması	Yapılan çalışmada elde edilen ilişkisel örüntülere göre kişilere ve kişisel mülkiye yapılan saldırıların yüksek oranda meskûn yerlerde meydana geldiği, 2011-2019 yılları arasında terörist faaliyetlerin zirve noktasına ulaştığı, vb. tespitler sunulmuştur.
Salem, Naouali ve Chtourou, 2020	1970-2016	Terörist gruplarının faaliyetleri (aktivite oranı) ve saldırganlık seviyelerine (ölümcüllük parametresi) göre karakterize edilmesi.	Gözetimsiz Öğrenme (K-Ortalama Algoritması)	Terörist Grupların faaliyetleri ve saldırganlıkları, 'aktivite' ve 'öldürücülük' ölçütleri ile kümelendirilmiştir.
Krieg, Smith, Chatterjee ve Chawla, 2022	2015-2021	Terör saldırılarının tarih ve yer (eyalet) tahmini.	Gözetimli Öğrenme (Rastgele Orman, AdaBoost, İleri Beslemeli sinir ağı, LSTM)	A.B.D'deki beş eyalet için yerel haberlerden elde edilen öznitelikler üzerinden belli bir tarih ve eyalette terör saldırısı olup olmayacağını tahminlemek üzere gözetimli öğrenme modelleri oluşturulmuştur.
Wang, Cui, ve He, 2022	2010- 2019	Teröristlerin saldırı riski seviyesi tahmini.	Gözetimli Öğrenme (CNN Yapay Sinir Ağı Mimarisi)	Teröristlere ait öznitelikler ikili olarak kodlanmış ve Top-K Algoritması ile en ilgili öznitelikler önceliklendirilerek sıralanmıştır.

Sonuç

Terör saldırıları cana ve mala zarar vererek toplum düzenini hedef alması itibariyle ülkeler için önlenmesi önde gelen olaylardır. Bu amaçla ulusal ve uluslararası mücadele için çeşitli stratejiler geliştirilmektedir. Ancak terör saldırıları genellikle basit silahlar kullanılarak, küçük ve izole gruplar tarafından rastgele saldırılar şeklinde gelişebildiği için öngörülemezlik unsuru barındırmaktadır. Son yıllarda veri saklama ve işleme teknolojilerinin ve yapay zekâ tekniklerinin büyük bir hızla ilerlemesi sonucunda bu teknolojilerin terörle mücadele için, özellikle belirsizliği azaltarak tahminleme problemleri için kullanıldığını görmekteyiz. Bu çalışmada terör olaylarına ilişkin verileri içeren erişime açık veri kaynaklarını ve 2019-2022 yıllarında bu veri kaynakları üzerinde yapay zekâ ve büyük veri teknolojileri kullanarak terörle mücadele problemlerini ele alan akademik çalışmalar incelenmiştir.

İncelenen çalışmalarda mevcut dört açık veri kaynağından sadece bir tanesinin (GTD) akademik çalışmalarda kullanıldığı gözlemlenmiştir. Ancak bu zengin veri tabanı üzerinde yapılan makine öğrenmesi uygulamaları, terör olaylarına ilişkin kimi tahminlemeler için oldukça başarılı sonuçlar ortaya koyarak (örneğin terör saldırısının verdiği zayıat için %87 doğrulukla tahmin) yapay zekâ çözümlerinin kullanılabilirliğini göstermiştir. Buna ek olarak büyük veri mimarilerinin kullanımı ile artan veri miktarına rağmen, terör olaylarına ilişkin veri tabanı sorgularının ve veri analizlerinin hızlı bir şekilde yapılabileceği gösterilmiştir.

İncelemiş olduğumuz makalelerde kullanılan makine öğrenmesi teknikleri tahminleme için kullanılabilir performanslar vermekle birlikte, sadece klasik teknik ve algoritmaların kullanıldığı görülmektedir. Sinir ağları ve özellikle derin sinir ağları modelleri gibi daha yeni ve hibrit modeller gibi daha karmaşık ve yetenekli çözümler kullanılarak sonraki çalışmalarda daha yüksek tahmin başarısı elde edilmesi olasıdır.

İlgili akademik makalelerde ele alınmış olan tahmin problemleri, saldırının zayıat büyüklüğü, bölgesi, ya da verilen bir bölgedeki terör saldırısı sayısı gibi statik tahminlere odaklanmıştır. Bunların yanı sıra geleceğe yönelik tahminleme problemlerinin ele alınması, örneğin gelecek bir saldırının zamanı ve saldırı tipinin tahmini, olası saldırılar için öngörü geliştirebilmek açısından fayda sağlayacaktır.

Terör olayları ile ilgili veri tabanları için, veri miktarındaki artışın yanı sıra, mobil araçlar, sensörler ve benzer teknolojik çözümlerle ilgili verilerin akış hızında da yükselme olması beklenebilir. Bu durumda akan veri işleme mimarilerinin kullanımı ile terör olaylarına ilişkin verilerin daha etkin olarak sorgulanması, analiz edilmesi ve gerçek zamanlı tahminler yapılması mümkün olacaktır.

Extended Summary

Terrorism poses an important threat at the national, regional, and international levels as one of the most important security problems today as it has been for centuries. The word terrorism is defined in the dictionary of the Turkish Language Institution as “intimidation, killing, burning and destroying property, terror”.

Another important aspect of terrorist acts that causes destructive effects is the uncertainty about when and how events will occur. Although intelligence sources play an important role in predicting and preventing possible terrorist acts (Çıtak, 2021), it is not always possible to obtain and evaluate this information. This uncertainty in terrorist acts necessitates new methods. The developing information technologies provide new opportunities with the fight against terrorism, and it is possible to collect, store and analyze historical information about terrorist incidents (Bhatia, Chhabra ve Kumar, 2020). Especially with the use of big data technologies and artificial intelligence techniques, studies on forecasting terrorist incidents are increasing (Atsa'am, Wario ve Okpo, 2020; Vaibhav, 2020; Singh, Chaudhary ve Kaur, 2019). In addition, it is seen that data on past terrorist incidents have been collected for analysis, and terror databases have begun to be established (GTD, 2021; AOAV, 2021). These data collections constitute an important resource for artificial intelligence and big data analysis studies.

In the presented article, the use of machine learning and big data technologies in the fight against terrorism is examined. For this purpose, firstly, open-access databases and datasets of terrorist incidents were searched. In our review, we found four data sources with these attributes: Global Terrorism Database (GTD, 2021), Action on Armed Violence Database (AOAV, 2021), Data World (DW, 2021) datasets, and Kaggle datasets (Kaggle, 2021). Subsequently, we searched recent (2020 and 2021) academic studies on counter-terrorism on these datasets. Although we detected many article tags in our search, in our more detailed

analysis of these articles, we observed that artificial intelligence and big data technologies were used in a limited number of studies and only the Global Terrorism Database was used as data. Although the other data sources we have identified are mentioned in the articles by name or bibliography, an analysis using these data has not been presented. In other articles, other than artificial intelligence and big data technologies, different issues such as the social, economic, and psychological dimensions of the terror problem are discussed under the title of terrorism. In the last stage of the study, six studies that we determined to use artificial intelligence and big data technologies were examined in detail and presented comparatively.

The remainder of the article is organized as follows. In the first part, information about datasets containing past terrorist incidents is shared. In the second part, the article scanning approach applied within the scope of the study is presented and information about the articles found as a result of the scanning is given. In the third chapter, six studies that analyze and predict GTD data using machine learning and big data technologies are presented in detail in terms of the problems they deal with, the methods they use, and the results obtained. In the fourth chapter, the conclusions drawn from the study and the discussion are given. Finally, the article is completed with a summary evaluation.

The data sources examined include terrorist incidents of 1970 and later. Although the history of terrorist incidents is very old, the size of the data generated is quite sufficient in terms of analysis and estimation. It is seen that the data are collected from various official sources and national newspapers. The presented data includes much different information such as the location and time of the terrorist incident, as well as its type, weapons used, casualties and losses, and the criminal organization that took responsibility for the incident.

Across the studies we examined, mostly supervised machine learning methods were preferred and they focused on different estimation problems.

Kaynakça

Kitaplar

Azarmi, B. (2016). *Scalable Big Data Architecture. A Practitioner's Guide to Choosing Relevant Big Data Architecture (s.1-16)*. California: Apress.

- Fetzer, J. H. (1990). *What is Artificial Intelligence?. In Artificial Intelligence: Its Scope and Limits (pp. 3-27)*. Berlin: Springer Dordrecht.
- Güçlü, İ. (2014). *Terörle Mücadele Stratejileri*. Harmanlı, F. M, Gözübenli, M. ve Zengin, C. (Ed.), "Güvenlik Sektöründe Temel Stratejiler" içinde (s.99-120). İstanbul: Nobel Yayınevi
- Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). *Machine learning: An artificial intelligence approach*. Springer Science and Business Media.
- Tan, P., Steinbach, M. ve Kumar, V. (2016). *Introduction to Data Mining*. Boston: Pearson Addison Wesley.
- Thomas, R., Vaughan, I. ve Lello, J. (2013). *Data analysis with R statistical software: A guidebook for scientists*. Newport: Eco-explore.

Makaleler

- Atsa'am, D. D., Wario, R., Okpo, F. E. (2020). A new terrorism categorization based on casualties and consequences using hierarchical clustering. *Journal of applied security research*, 15 (3), 369-384.
- Çıtak, E. (2021). Terörle Mücadelede İstihbarata Başvuru: Açık Kaynak İstihbaratının Kullanımı. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (46), 163-179.
- Dittrich, J. ve Quiané-Ruiz, J. A. (2012). Efficient big data processing in Hadoop MapReduce. *Proceedings of the VLDB Endowment*, 5 (12), 2014-2015.
- Han, J., Pei, J. Ve Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Huamaní, E.L., Mantari, A. ve Roman-Gonzalez, A. (2020). Machine learning techniques to visualize and predict terrorist attacks worldwide using the global terrorism database. *International Journal of Advanced Computer Science and Applications*, 11.
- Iorliam A., Dugeri, R. U., Akumba, B. O. ve Otor, S. (2021). An Investigation And Insight Into Terrorism In Nigeria. *Journal of Information Security*, 12 (4).

Krieg S. J., Smith, C. W., Chatterjee, R. ve Chawla, N. V. (2022). Predicting Terrorist Attacks in the United States using Localized News Data. *Journal of Plos One* 17 (6), 1-26.

McCarthy, J. ve Hayes, P. J. (1981). Some philosophical problems from the standpoint of artificial intelligence. *In Readings in artificial intelligence*. 431-450.

Vaibhav, S. (2020). Predicting success of terrorist attack and extent of its economic impact using data mining. *International Journal for Research in Applied Science and Engineering Technology*, 8 (5), 1965-1972.

Wang, Y., Cui, X. ve He, P. (2022). Winning the War on Terror: Using “Top-K” Algorithm and CNN to Assess the Risk of Terrorists. *International Journal of Information Technology and Web Engineering (IJITWE)*, 17 (1), 1-15.

Sempozyum, Kongre ve Konferans Bildirileri

Bhatia, K., Chhabra, B. ve Kumar, M. (2020). Data analysis of various terrorism activities using big data approaches on global terrorism database. *In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 137-140. IEEE.

Narula, M., Kumar, S., Kaul, S. ve KS, S. (2020). Predictive analysis on global terrorism. *In Proceedings of the International Conference on Innovative Computing and Communications (ICICC)*.

Salem, S. B., Naouali, S. ve Chtourou, Z. (2020). Clustering Terrorist Groups Using Their activity and Lethality Measures. *Conference on International Business Information Management Association (IBIMA)*.

Singh, K., Chaudhary, A. S. ve Kaur, P. (2020). A machine learning approach for enhancing defence against global terrorism. *In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-5)*. IEEE.

İnternet Kaynakları

AOAV, Action on Armed Violence, Eylül 2021 tarihinde www.aoav.org.uk adresinden alınmıştır.

Data.world, Eylül 2021 tarihinde <https://data.world/> adresinden alınmıştır.

GTD, Global Terrorism Database, Eylül 2021 www.start.umd.edu/gtd adresinden alınmıştır.

Kaggle, Eylül 2021 tarihinde <https://www.kaggle.com/> adresinden alınmıştır.

TDK, Türk Dil Kurumu, *Türkçe'de Batı Kökenli Kelimeler Sözlüğü*, <https://sozluk.gov.tr/> adresinden alınmıştır.

Vikipedi, Terörizm, Eylül 2021 tarihinde <https://tr.wikipedia.org/wiki/Terörizm>, adresinden alınmıştır.