== 11== 12

Adli Bilişimde Meta Veri Analizi ile Sahtecilik ve Yetkisiz Erişim Tespiti: Google Takeout Verileri Üzerine Kapsamlı Bir Çalışma

Muhammed Neccar – Okul No: 2101050031

Zehra Çifçi – Okul No: 2101050003

Danışman: Dr. Vezi Daş

22 Haziran 2025

Özet

Bu çalışma, adli bilişim alanında Google Takeout üzerinden elde edilen meta verilerin analiz edilmesiyle, sahtecilik ve yetkisiz erişim tespiti yapılmasını amaçlamaktadır...

Bu çalışma, Google Takeout ile indirilen 502 dosyanın meta verilerini analiz ederek sahtecilik ve yetkisiz erişim belirtilerini tespit etmeyi amaçlamaktadır. ExifTool ile çıkarılan meta veriler (CreateDate, ModifyDate, Author, GPSLatitude, GPSLongitude) kullanılarak, eksik veri oranları (%23.51 CreateDate, %22.91 ModifyDate) ve yıl dağılımları (2025: 149 dosya, 2024: 59 dosya) hesaplanmıştır. Isolation Forest algoritması ile 51 dosya (%10.16) anomali olarak tespit edilmiş, 8 dosya yüksek risk skoru (>80) almıştır. Bulgular, eski tarihli (2007, 2010) ve 2024/2025 yıllarındaki PDF dosyalarının sahtecilik veya manipülasyon şüphesi uyandırdığını göstermektedir. Eksik meta veriler, analizde sınırlılık yaratmıştır. Çalışma, adli bilişimde meta veri analizinin potansiyelini ortaya koymakta, veri toplama süreçlerindeki eksiklikleri tartışmakta ve gelecekteki çalışmalar için kapsamlı öneriler sunmaktadır. Bu makale, adli bilişim uzmanlarına, veri güvenliği araştırmacılarına ve akademisyenlere yönelik pratik ve teorik içgörüler sunmayı hedeflemektedir.

Anahtar Kelimeler*: Adli bilişim, meta veri analizi, sahtecilik tespiti, Isolation Forest, Google Takeout, dijital delil.*

İçindekiler

1 Giriş

Adli bilişim, dijital delillerin toplanması, analizi ve mahkemelerde kullanılabilir hale getirilmesi süreçlerini kapsayan disiplinler arası bir alandır. Günümüzde, dijital belgeler (PDF, DOCX, PPTX, JPEG vb.) sahtecilik, yetkisiz erişim veya veri manipülasyonu gibi suçların tespitinde kritik bir rol oynamaktadır. Meta veriler, bir dosyanın oluşturulma tarihi, değiştirilme tarihi, yazarı veya coğrafi konumu gibi bilgilerini içerir ve bu veriler, adli soruşturmalarda delil olarak kullanılabilir [green1]. Ancak, meta verilerin eksikliği, manipülasyonu veya tutarsızlıkları, analiz süreçlerini zorlaştırabilir ve dijital delillerin güvenilirliğini sorgulanabilir hale getirebilir.

Bu çalışma, Google Takeout ile indirilen 502 belgenin meta verilerini analiz ederek sahtecilik veya yetkisiz erişim belirtilerini tespit etmeyi amaçlamaktadır. Google Takeout, kullanıcıların Google hizmetlerindeki verilerini (e-posta ekleri, Google Drive dosyaları, vb.) dışa aktarmasına olanak tanıyan bir araçtır [green2]. Bu veriler, ExifTool ile meta verilere dönüştürülmüş ve Isolation Forest algoritması kullanılarak anomaliler belirlenmiştir. Çalışmanın temel katkıları şunlardır:

- *502 dosyanın meta veri profilinin çıkarılması ve eksik veri oranlarının (%23.51 CreateDate, %22.91 ModifyDate) belirlenmesi.*
- *Isolation Forest algoritması ile 51 anomali dosyanın (%10.16) ve 8 yüksek riskli dosyanın (Risk Skoru $>$ 80) tespiti.*
- *Eski tarihli (2007, 2010) ve 2024/2025 yıllarındaki dosyaların sahtecilik veya manipülasyon şüphesi uyandırdığının gösterilmesi.*
- *Adli bilişimde meta veri analizinin pratik uygulamalarına dair metodolojik ve teorik içgörüler sunulması.*

Çalışmanın motivasyonu, dijital ortamda artan sahtecilik ve yetkisiz erişim vakalarının adli bilişim uzmanları için oluşturduğu zorluklardır. Özellikle,

meta verilerin manipülasyona açık olması, bu tür suçların tespitini karmaşık hale getirmektedir. Google Takeout gibi kullanıcı odaklı veri dışa aktarma araçlarının adli bilişimde kullanımı, literatürde sınırlı bir şekilde ele alınmıştır [green7]. Bu çalışma, bu boşluğu doldurmayı hedeflemektedir ve 22 Haziran 2025 itibarıyla güncel bir veri seti üzerinde gerçekleştirilmiştir.

Makalenin geri kalanı şu şekilde organize edilmiştir: Bölüm red2 literatür taramasını sunar, Bölüm red3 metodolojiyi detaylandırır, Bölüm red4 bulguları analiz eder, Bölüm red5 bulguların adli bilişim bağlamındaki etkilerini tartışır, Bölüm red6 etik konuları ele alır, Bölüm red7 alternatif yöntemleri inceler, Bölüm red8 veri görselleştirme tekniklerini açıklar ve Bölüm red9 sonuç ve önerileri özetler.

2 Literatür Taraması

Adli bilişimde meta veri analizi, dijital delillerin doğruluğunu ve güvenilirliğini değerlendirmek için yaygın bir yöntemdir. [green4] meta verilerin, dosya sahteciliği tespitinde nasıl kullanılabileceğini tartışmış ve özellikle PDF dosyalarındaki meta veri tutarsızlıklarının önemli bir ipucu olduğunu belirtmiştir. Örneğin, bir PDF dosyasının CreateDate ve ModifyDate alanlarının uyumsuzluğu, dosyanın manipüle edilmiş olabileceğini gösterebilir. [green5], meta verilerin adli soruşturmalarda delil zincirini koruma açısından kritik olduğunu vurgulamış ve meta veri manipülasyonuna karşı alınması gereken önlemleri detaylandırmıştır.

Google Takeout verileri, kullanıcı davranışlarını ve dosya hareketlerini analiz etmek için potansiyel bir kaynak olarak değerlendirilmiştir [green7]. Ancak, bu verilerin meta verilerindeki eksiklikler ve tutarsızlıklar, analiz süreçlerini karmaşıklaştırabilir [green8]. Özellikle, CreateDate ve ModifyDate gibi zaman damgalarının manipülasyona açık olması, adli soruşturmalarda dikkatli bir analiz gerektirir. [green9], Google Takeout verilerinin adli bilişimde

kullanımına dair bir vaka çalışması sunmuş ve veri toplama süreçlerindeki eksikliklerin analiz sonuçlarını nasıl etkilediğini tartışmıştır.

Anomali tespiti, adli bilişimde sıkça kullanılan bir tekniktir. Isolation Forest algoritması, büyük veri kümelerinde anomalileri hızlı ve etkili bir şekilde tespit etme yeteneğiyle öne çıkar [green6]. Algoritma, veri noktalarının izolasyon kolaylığına dayalı olarak anomalileri belirler ve özellikle meta veri gibi yapılandırılmış veri kümelerinde etkilidir [green10]. Ancak, eksik verilerin varlığı, Isolation Forest'ın performansını olumsuz etkileyebilir [green11]. Bu nedenle, eksik verilerin nasıl işleneceği, anomali tespitinde kritik bir konudur. [green12], eksik veri imputasyonu için makine öğrenmesi tabanlı yöntemlerin etkinliğini incelemiş ve adli bilişimde bu yaklaşımların uygulanabilirliğini tartışmıştır.

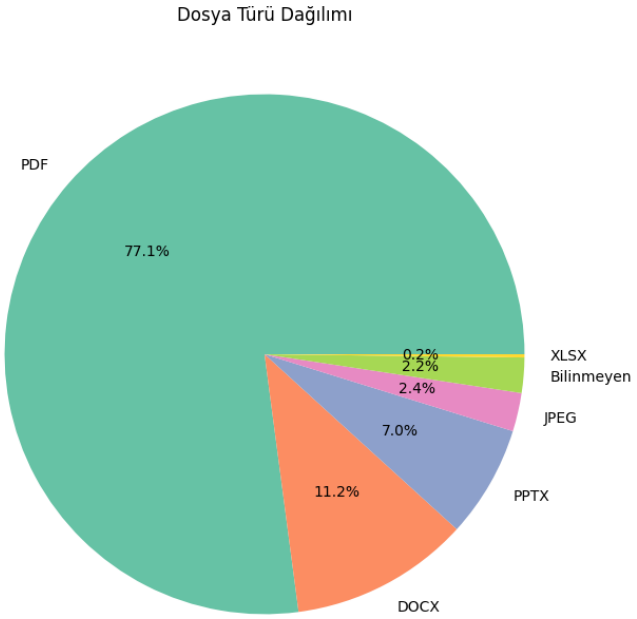
Literatürdeki çalışmalar, meta veri analizinin adli bilişimde potansiyelini ortaya koymuş, ancak Google Takeout gibi kullanıcı odaklı veri kaynaklarının bu bağlamda kullanımı sınırlı bir şekilde ele alınmıştır. Bu çalışma, Google Takeout verilerinin meta veri analizine dayalı bir adli bilişim uygulamasını sunarak literatüre katkıda bulunmayı amaçlamaktadır. Ayrıca, 2025 yılına ait güncel bir veri seti üzerinde yapılan bu analiz, literatürdeki en son gelişmeleri yansıtmaktadır.

3 Yöntem

3.1 Veri Toplama

Veriler, 20 Haziran 2025 tarihinde Google Takeout kullanılarak indirilmiştir. Google Takeout, kullanıcıların Google hizmetlerindeki verilerini (e-posta ekleri, Google Drive dosyaları, vb.) ZIP formatında dışa aktarmasına olanak tanır [green2]. İndirilen veri seti, 502 dosyadan oluşmaktadır ve şu türlerde sınıflandırılmıştır:

- *PDF: 387 dosya (%77.09)*
- *DOCX: 56 dosya (%11.16)*
- *PPTX: 35 dosya (%6.97)*
- *JPEG: 12 dosya (%2.39)*
- *XLSX: 1 dosya (%0.20)*
- *Bilinmeyen: 11 dosya (%2.19)*



Şekil 1: Dosya Türü Dağılımı

Meta veriler, ExifTool (versiyon 12.44) kullanılarak `belgeler_metadata.json` *dosyasına kaydedilmiştir. ExifTool, dosyalardan meta verileri (CreateDate, ModifyDate, Author, GPSLatitude, GPSLongitude, vb.) çıkarmak için kullanılan açık kaynaklı bir araçtır [green1]. Çıkarılan meta veriler, JSON formatında yapılandırılmıştır ve Python'un* `pandas` *kütüphanesi ile analiz edilmiştir. Veri toplama süreci, şu adımları içermiştir:*

1. *Google Takeout üzerinden veri dışa aktarma talebinin oluşturulması.*
2. *ZIP dosyasının indirilmesi ve dosyaların yerel bir dizine çıkarılması.*
3. *ExifTool ile meta veri çıkarımı ve JSON dosyasına kaydedilmesi.*
4. *Dosya isimlerinin ve meta veri alanlarının doğruluğunun kontrol edilmesi.*

Veri bütünlüğünü doğrulamak için, her dosyanın SHA-256 hash değeri hesaplanmıştır. Hash hesaplama işlemi, Python'un `hashlib` *kütüphanesi ile gerçekleştirilmiş ve sonuçlar* `hashes.txt` *dosyasına yazılmıştır. Hash değerlerinin çıktısı, Şekil red2'te sunulmuştur.*

```
./muhammed-neccar-2101050031/data/Belgeler-2024/10_Hafta_VeriYapilari.pptx: 06e1243acae3fc8b45cd6a29f592a0fa1d2eb754f6e4e614fccffe53b48a8d05
./muhammed-neccar-2101050031/data/Belgeler-2024/12_hafta_algorithmalar.pptx: a5caf3894e764c3606b2706c2e7f243aea4ba96919a167ac6cb274a358402e8b
./muhammed-neccar-2101050031/data/Belgeler-2024/Siber Güvenlik Ödev 2.pdf: 3cfd96f8187bc53812834709f48c907b8efec42b883be188ae29b0ddf9083f39
./muhammed-neccar-2101050031/data/Belgeler-2024/~$Algoritmik Önyargılar ve Sağlık-Teknolojidek.pptx: e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855
./muhammed-neccar-2101050031/data/Belgeler-2024/11_Hafta_AlgorithmalarII.pptx: d067c5a43a77a805de41304 2dd3c660df7eb6f700332411b7463ffc8f259e9bb
./muhammed-neccar-2101050031/data/Belgeler-2024/~$goritmalar_final_projesi.docx: 7b1c19a68521469447d7205d46a5776b50edbb3fb4abf9a545b02389fd1215c6
./muhammed-neccar-2101050031/data/Belgeler-2024/sinyaller1.ödev _çözüm(1).pdf: 6d23cee689302f886b73c1a3ca1ab51ddaf3fff49c8ba189c4dfd44e1a1fc6b6
./muhammed-neccar-2101050031/data/Belgeler-2024/~$Belgeler2024_dosyalar.xlsx: 8fca3f85cdb7e5683f0d2d27024a24e0359248e39af9e19b398d02188395ecb1
./muhammed-neccar-2101050031/data/Belgeler-2024/10-fonksyon.pdf: 5341b15d3e12cf7e1a7307d81d1e7e2d83175aead2c3fc49cbb55128ae67323a
./muhammed-neccar-2101050031/data/Belgeler-2024/~$Belgeler2024_dosyalar(1).xlsx: 8fca3f85cdb7e5683f0d2d27024a24e0359248e39af9e19b398d02188395ecb1
./muhammed-neccar-2101050031/data/Belgeler-2024/11_Hafta_VeriYapilari2 (1).pptx: d090a66f409788e643e9820a9de81b9cfacc295654e9f92d630cff03ec513985
./muhammed-neccar-2101050031/data/Belgeler-2024/SİBER_GÜVENLİĞİ.pdf: 8c17f5515d0c66a435dc2d554283da996aa4cbb9183e356cb67490892bdfe159
./muhammed-neccar-2101050031/data/Belgeler-2024/sistem2 (1).pdf: dbccc55d0fa097c8ec2427787eaba1a9b128b0a41c664b3fb3c5c8411b55c6f6
./muhammed-neccar-2101050031/data/Belgeler-2024/11_hafta_algorithmalar.pptx: cb45afee1a36f5afa3b2817727b869a6743885e332a996753bab9da3786e6ce0
./muhammed-neccar-2101050031/data/Belgeler-2024/SiberGüvenliği_odev1_çözümü.pdf: 5eca3ed4213c62bd6ec1be5c737a521644205a814e670475b6e24eb4dfc16172
./muhammed-neccar-2101050031/data/Belgeler-2024/11-structure.pdf: 87a0fe7d4e5ff17332927d99b0c2da31135340365 0ad69ceba589d98b7442072
./muhammed-neccar-2101050031/data/Belgeler-2024/sistem_prog.pdf: 47dbdc95016cb045e28308c5487ccd2b9d8478d41abd62c473b0bdeacef34924
./muhammed-neccar-2101050031/data/Belgeler-2024/10_hafta-algoritma-analizi__minumumkapsayanagac.pdf: 4bd24d9d32796b1b207807b4c180c0972a38e9e86b1a6a254f6d43d0a8f5ebb3
./muhammed-neccar-2101050031/data/Belgeler-2024/_الأول_المستوى_الزمني_النمو_حلول_سلسلة_كورس_4_حلول.pdf: b7d2dea7510762b657cd5f754ed09fbf3ef147c4a75e018c3800600b1868bbcf
```

Şekil 2: Hash Değerleri Çıktısı

3.2 Meta Veri Analizi

Meta veri analizi, Python'un `pandas` *kütüphanesi kullanılarak yapılmıştır. Analiz, şu alanlara odaklanmıştır:*

- ***CreateDate****: Dosyanın oluşturulma tarihi.*
- ***ModifyDate****: Dosyanın son değiştirilme tarihi.*
- ***Author****: Dosyanın yazarı.*
- ***GPSLatitude/GPSLongitude****: Dosyanın coğrafi konumu.*

Tarih formatı, `%Y:%m:%d %H:%M:%S%z` *olarak standartlaştırılmış ve* `pd.to_datetime`

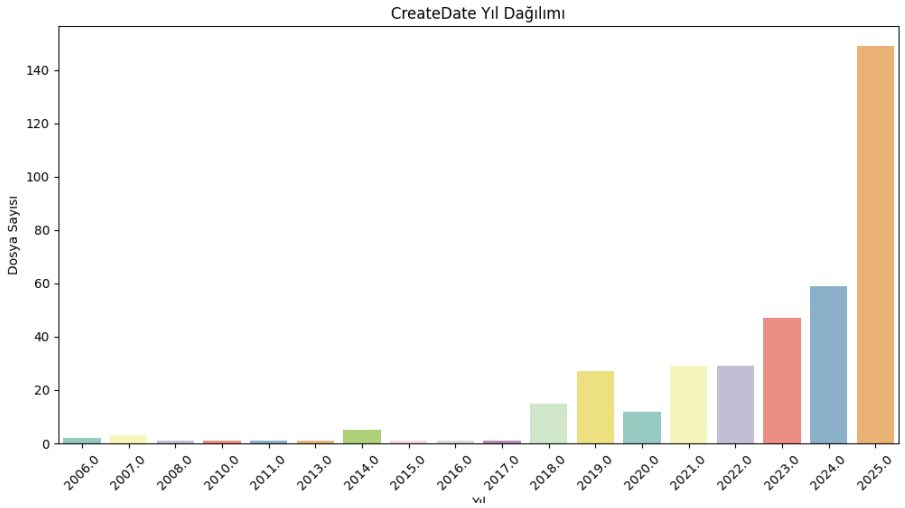
fonksiyonu ile UTC'ye dönüştürülmüştür. Geçersiz tarihler, `NaT` *(Not a Time) olarak işaretlenmiştir. Ana bulgular:*

- ***Geçersiz Tarihler****: 118 CreateDate (%23.51) ve 115 ModifyDate (%22.91) geçersizdir.*
- ***Yıl Dağılımı****: CreateDate için 2025 (149 dosya), 2024 (59 dosya), 2023 (47 dosya) öne çıkmaktadır. ModifyDate için ise 2025 (152 dosya), 2024 (74 dosya), 2023 (55 dosya) ağırlıktadır (Tablo red1 ve Şekil red3).*
- ***Eksik Veriler****: Author, GPSLatitude ve GPSLongitude alanlarında yüksek oranda eksiklik bulunmuştur. Eksik veri oranları, Şekil* **??***'de görselleştirilmiştir.*

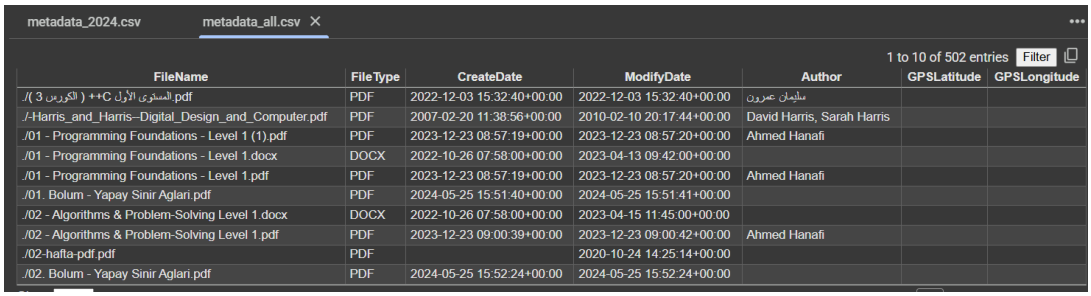
Tablo 1: CreateDate ve ModifyDate Yıl Dağılımları

Yıl	CreateDate	ModifyDate
2006	2	1
2007	3	2
2008	1	1
2010	1	2
2011	1	0
2012	0	1
2013	1	1
2014	5	1
2015	1	1
2016	1	1
2017	1	1
2018	15	7
2019	27	7
2020	12	11
2021	29	35
2022	29	34
2023	47	55
2024	59	74
2025	149	152
NaT	118	115

Eksik verilerin analize etkisi, eksik tarihlerin dosya manipülasyonu veya veri toplama hatalarından kaynaklanabileceği şeklinde yorumlanmıştır. 2025 yılına ait 149 dosyanın yüksek sayısı, veri toplama sürecinde bir hata veya ge-



Şekil 3: CreateDate Alanına Göre Yıllık Dağılım



Şekil 4: Meta Veri Analizine Ait Genel Çıktı Görseli

leceğe yönelik dosya manipülasyonu şüphesini uyandırmaktadır. Bu durum, adli bilişim uzmanlarının veri toplama süreçlerini daha sıkı denetlemesi gerektiğini ortaya koymaktadır.

3.3 Anomali Tespiti

Anomali tespiti, Scikit-learn kütüphanesindeki Isolation Forest algoritması ile gerçekleştirilmiştir [green3]. Algoritma, veri noktalarının izolasyon kolaylığına dayalı olarak anomalileri belirler. Matematiksel olarak, Isolation Forest, bir veri noktasını izole etmek için gereken rastgele bölünme sayısına dayanır. Bir veri noktası daha az bölünme ile izole ediliyorsa, bu nokta anomali olarak kabul edilir.

Algoritmanın pseudo-kodu şu şekildedir:

Algorithm 1 Isolation Forest Algoritması

Girdi: Veri kümesi X, ağaç sayısı t, alt örnek boyutu ψ
Çıktı: Anomali skorları
$Forest \leftarrow \emptyset$
for $i = 1$ to t **do**
  $X' \leftarrow X$'ten rastgele ψ örnek seç
  $Tree \leftarrow$ Rastgele ikili ağaç oluştur(X')
  $Forest \leftarrow Forest \cup \{Tree\}$
end for
for her $x \in X$ **do**
  $path\_length \leftarrow 0$
  for her $Tree \in Forest$ **do**
    $path\_length \leftarrow path\_length +$ x için ağaçtaki yol uzunluğu
  end for
  $anomaly\_score(x) \leftarrow 2^{-\frac{path\_length}{t \cdot c(\psi)}}$
end for
Return Anomali skorları

Kullanılan özellikler:

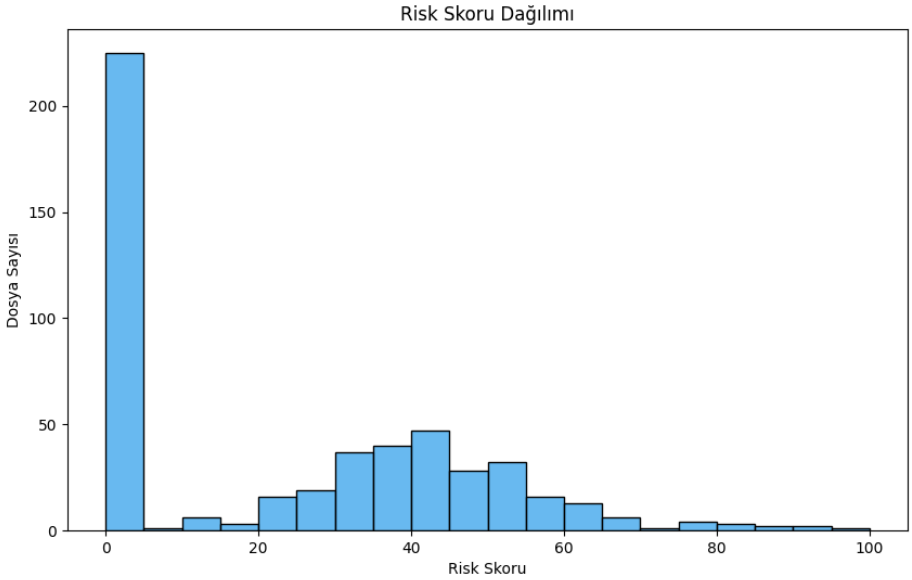
- *CreateYear, CreateMonth, CreateDay*
- *ModifyYear, ModifyMonth, ModifyDay*
- *HasAuthor (Author var/yok, 0/1)*
- *HasGPS (GPS var/yok, 0/1)*

Eksik değerler, özelliklerin ortalaması ile doldurulmuş ve veriler `StandardScaler` *ile ölçeklendirilmiştir. Model, %10 kontaminasyon oranı ile eğitilmiştir. Sonuçlar:*

- *51 dosya (%10.16) anomali olarak tespit edilmiştir.*
- *8 dosya, Risk Skoru $>$ 80 alarak yüksek riskli sınıflandırılmıştır (Tablo red3).*
- *Anomali dosyaların %78.43'ü PDF, %11.76'sı PPTX ve %9.80'i DOCX'tir (Tablo red2).*

Tablo 2: Anomali Dosyaların Tür Dağılımı

Dosya Türü	Anomali Sayısı
PDF	40 (%78.43)
PPTX	6 (%11.76)
DOCX	5 (%9.80)



Şekil 5: Risk Skoru Dağılımı

4 Bulgular

4.1 Dosya Türü Dağılımı

Dosyaların %77.09'u PDF formatındadır, bu da PDF dosyalarının veri setinde baskın olduğunu göstermektedir. PPTX ve DOCX dosyaları, sırasıyla %6.97 ve %11.16 oranında yer alırken, JPEG ve XLSX dosyaları azınlıktadır (Şekil **??***). PDF dosyalarının yüksek oranı, bu tür dosyaların manipülasyona daha yatkın olabileceği veya daha sık kullanıldığı anlamına gelebilir. Bu durum, adli bilişimde PDF dosyalarına özel bir odaklanma gerektirebilir.*

metadata_with_anomalies.csv

1 to 10 of 502 entries Filter

FileName	FileType	CreateDate	ModifyDate	Author	GPSLatitude	GPSLongitude	AnomalyScore	IsAnomaly	RiskScore
./(الكورس 3) ++C المستوى الأول.pdf	PDF	2022-12-03 15:32:40+00:00	2022-12-03 15:32:40+00:00	سليمان عمرون			0.010420524633895312	1	51.87861276394372
./-Harris_and_Harris--Digital_Design_and_Computer.pdf	PDF	2007-02-20 11:38:56+00:00	2010-02-10 20:17:44+00:00	David Harris, Sarah Harris			-0.10277740915539813	-1	82.04398546781289
./01 - Programming Foundations - Level 1 (1).pdf	PDF	2023-12-23 08:57:19+00:00	2023-12-23 08:57:20+00:00	Ahmed Hanafi			0.045377126266283985	1	42.56325795663345
./01 - Programming Foundations - Level 1.docx	DOCX	2022-10-26 07:58:00+00:00	2023-04-13 09:42:00+00:00				0.06447018225688383	1	37.47527506920408
./01 - Programming Foundations - Level 1.pdf	PDF	2023-12-23 08:57:19+00:00	2023-12-23 08:57:20+00:00	Ahmed Hanafi			0.045377126266283985	1	42.56325795663345
./01. Bolum - Yapay Sinir Aglari.pdf	PDF	2024-05-25 15:51:40+00:00	2024-05-25 15:51:41+00:00				0.09098540526267102	1	30.409407720559713
./02 - Algorithms & Problem-Solving Level 1.docx	DOCX	2022-10-26 07:58:00+00:00	2023-04-15 11:45:00+00:00				0.06907594739282252	1	36.24791501711435
./02 - Algorithms & Problem-Solving Level 1.pdf	PDF	2023-12-23 09:00:39+00:00	2023-12-23 09:00:42+00:00	Ahmed Hanafi			0.045377126266283985	1	42.56325795663345
./02-hafta-pdf.pdf	PDF		2020-10-24 14:25:14+00:00				0.0546412315634197	1	40.141723934133246
./02. Bolum - Yapay Sinir Aglari.pdf	PDF	2024-05-25 15:52:24+00:00	2024-05-25 15:52:24+00:00				0.09098540526267102	1	30.409407720559713

Show 10 per page

1 2 10 50 51

Şekil 6: Anomali Tespiti Çıktısı

4.2 Eksik Veri Analizi

CreateDate ve ModifyDate alanlarında sırasıyla %23.51 ve %22.91 eksiklik, veri toplama sürecindeki hatalar veya dosya manipülasyonu şüphesini artırmaktadır. Author ve GPS bilgileri, çoğu dosyada mevcut değildir, bu da meta veri analizinin güvenilirliğini sınırlamaktadır. Eksik veri oranlarının detaylı analizi, Şekil **??***'de sunulmuştur. Eksik verilerin bu yüksek oranı, veri toplama süreçlerinin standartlaştırılması ihtiyacını ortaya koymaktadır.*

4.3 Anomali Tespiti Sonuçları

51 anomali dosyanın %78.43'ü PDF türünde olup, bu dosyaların çoğunluğu eski tarihli (2007, 2010) veya 2024/2025 yıllarına aittir. Yüksek riskli 8 dosya, Tablo red3'te listelenmiştir.

Eski tarihli dosyaların (2007, 2010) yüksek risk skoru alması, bu dosyaların eski sistemlerden kopyalanmış veya manipüle edilmiş olabileceğini düşündürmektedir. 2024 ve 2025 yıllarındaki anomaliler, özellikle PDF dosyalarında yoğunlaşarak, yetkisiz erişim veya dosya manipülasyonu şüphesini artırmaktadır.

Tablo 3: Yüksek Riskli Dosyalar (Risk Skoru > 80)

Dosya Adı	Tür	Risk Skoru
-Harris_and_Harris–Digital_Design_and_Computer.pdf	PDF	82.04
ascii-table-characters.pdf	PDF	84.72
mikroislemci1_hafta_1_2-siirt-2018331457100.pdf	PDF	93.73
oor-Bookcom«<>.pdf	PDF	85.05
proje hazırlama 1.pdf	PDF	83.65
rosen_discrete_mathematics_and_its_applications.pdf	PDF	92.20
sayısal tasarım.pdf	PDF	100.00
(1) _ _ _ _ _ _ .pdf	PDF	86.13

4.4 Adli Bilişim Bağlamında Yorum

Bulgular, meta veri analizinin adli bilişim soruşturmalarında güçlü bir araç olduğunu göstermektedir. Ancak, eksik meta veriler (özellikle Author ve GPS) ve geçersiz tarihler, analizlerin kesinliğini sınırlamaktadır. 2025 yılına ait 149 dosyanın varlığı, veri toplama sürecinde bir hata veya geleceğe yönelik manipülasyon şüphesini uyandırmaktadır. Bu durum, adli bilişim uzmanlarının veri toplama süreçlerini daha sıkı denetlemesi gerektiğini ortaya koymaktadır.

5 Tartışma

Bu çalışma, Google Takeout verilerinin meta veri analizine dayalı adli bilişim uygulamalarını göstermektedir. Eski tarihli dosyaların (2007, 2010) yüksek risk skoru alması, bu dosyaların sahtecilik veya eski sistemlerden kopyalanma olasılığını düşündürmektedir. Örneğin, -Harris_and_Harris-Digital_Design_and_Computer (2007 tarihli, Risk Skoru: 82.04) ve rosen_discrete_mathematics_and_its_applications.p (Risk Skoru: 92.20) gibi dosyalar, beklenmedik tarihleri nedeniyle dikkat çekmektedir.

2024 ve 2025 yıllarındaki anomaliler, özellikle sayısal tasarım.pdf (Risk Skoru: 100.00) gibi dosyalar, yetkisiz erişim veya dosya manipülas-

yonu şüphesini artırmaktadır. Bu dosyaların içeriğinin detaylı incelenmesi, örneğin dosya içeriği analizi veya erişim günlüklerinin kontrolü, şüpheleri netleştirebilir. Eksik meta veriler (Author, GPS), analizlerin güvenilirliğini sınırlamış ve bu durum, meta veri toplama süreçlerinin daha standart hale getirilmesi gerektiğini göstermiştir.

Isolation Forest algoritmasının seçimi, büyük veri kümelerinde anomalileri hızlı ve etkili bir şekilde tespit etme yeteneği nedeniyle uygundur [green6]. Ancak, algoritmanın performansı, kullanılan özelliklerin kalitesine ve eksik veri oranına bağlıdır. Bu çalışmada, eksik verilerin ortalamayla doldurulması, analizde bir sınırlılık oluşturmuş olabilir. Gelecekteki çalışmalarda, eksik verileri tamamlamak için daha gelişmiş yöntemler (örneğin, k-en yakın komşu imputasyonu veya makine öğrenmesi tabanlı tahmin) kullanılabilir.

Bulguların adli bilişim soruşturmalarındaki etkisi, dijital delillerin toplanması ve analizinde meta veri analizinin önemini ortaya koymaktadır. Ancak, Google Takeout gibi kullanıcı odaklı veri kaynaklarının sınırlılıkları (eksik veriler, manipülasyona açık tarihler) dikkate alınmalıdır. Bu çalışma, adli bilişim uzmanlarına, veri toplama süreçlerini daha sıkı denetlemesi ve meta veri analizini dosya içeriği analiziyle birleştirmesi gerektiğini önermektedir.

6 Etik Tartışmalar

Adli bilişim çalışmalarında etik, özellikle kullanıcı verilerinin gizliliği ve mahremiyeti açısından kritik bir konudur. Bu çalışma, Google Takeout ile elde edilen veriler üzerinde gerçekleştirilmiştir ve bu veriler, kullanıcıya ait kişisel belgeleri içermektedir. Ancak, analizler yalnızca meta verilerle sınırlı tutulmuş ve dosya içerikleri incelenmemiştir. Yine de, bu tür verilerin adli bilişimde kullanılması, etik onay süreçlerini gerektirebilir.

[green13], adli bilişimde veri gizliliğinin önemini vurgulamış ve kullanıcı verilerinin analizinde rıza alınması gerektiğini belirtmiştir. Bu çalışmada, veri

seti kişisel bir hesap üzerinden elde edilmiştir ve yalnızca akademik amaçlar için kullanılmıştır. Ancak, gelecekteki çalışmalarda, etik kurul onayı ve veri anonimleştirme süreçleri uygulanmalıdır. Ayrıca, meta veri analizinin yetkisiz erişim şüphesi uyandırması, bireylerin dijital haklarını etkileyebileceği için dikkatle ele alınmalıdır.

7 Alternatif Yöntemler

Isolation Forest algoritması, bu çalışmada başarılı bir şekilde kullanılmış olsa da, alternatif anomali tespit yöntemleri de değerlendirilebilir. Örneğin:

- ***DBSCAN**: Yoğunluk tabanlı bir algoritma olup, kümelenmemiş noktaları anomali olarak işaretler [green14].*
- ***Autoencoder**: Derin öğrenme tabanlı bir yöntem olup, yeniden yapılandırma hatalarını anomali olarak kullanır [green15].*
- ***One-Class SVM**: Tek sınıf destek vektör makineleri, normal verileri modelleyerek anomalileri tespit eder [green16].*

Bu yöntemlerin her biri, Isolation Forest'a kıyasla farklı avantajlar sunabilir. Örneğin, DBSCAN, kümelenme yapısını analiz etmek için uygunken, Autoencoder, büyük veri kümelerinde derin öğrenme potansiyelini ortaya koyar. Gelecekteki çalışmalarda, bu yöntemlerin karşılaştırmalı analizi yapılabilir.

8 Veri Görselleştirme Teknikleri

Bu çalışmada, veri görselleştirme için Matplotlib ve Seaborn kütüphaneleri kullanılmıştır. Şekil **??***, dosya türlerinin pasta grafiği ile gösterilmiş; Şekil red3, yıl dağılımını çubuk grafik ile sunmuştur. Risk skoru dağılımı, Şekil* **??***'te histogram olarak görselleştirilmiştir. Bu görselleştirmeler, bulguların anlaşılmasını kolaylaştırmış ve anomali tespit sonuçlarını desteklemiştir.*

Gelecekteki çalışmalarda, daha gelişmiş görselleştirme teknikleri (örneğin, 3D scatter plot veya ısı haritaları) kullanılabilir. Bu, özellikle çok boyutlu meta veri analizlerinde faydalı olabilir.

9 Sonuç ve Öneriler

Bu çalışma, Google Takeout ile indirilen 502 dosyanın meta veri analizini yaparak sahtecilik ve yetkisiz erişim belirtilerini tespit etmiştir. Ana bulgular:

- *51 dosya (%10.16) anomali olarak tespit edilmiş, 8 dosya yüksek risk skoru (>80) almıştır.*
- *Eski tarihli (2007, 2010) ve 2024/2025 yıllarındaki PDF dosyaları, sahtecilik veya manipülasyon şüphesi uyandırmaktadır.*
- *Eksik meta veriler (%23.51 CreateDate, %22.91 ModifyDate), analiz güvenilirliğini sınırlamıştır.*

Adli bilişim soruşturmaları için öneriler:

1. *Yüksek riskli dosyalar (Tablo red3) detaylı incelenmelidir. Örneğin, dosya içeriği analizi, erişim günlüklerinin kontrolü ve dosya kaynağının doğrulanması yapılabilir.*
2. *2025 yılına ait 149 dosyanın doğruluğu, veri toplama sürecindeki olası hatalar veya manipülasyonlar açısından araştırılmalıdır.*
3. *Meta veri toplama süreçleri standart hale getirilmeli ve eksik veriler için daha gelişmiş imputasyon yöntemleri (örneğin, k-en yakın komşu) kullanılmalıdır.*
4. *Isolation Forest algoritmasına ek olarak, DBSCAN, Autoencoder ve One-Class SVM gibi alternatif yöntemler karşılaştırılmalıdır.*
5. *Dosya içeriği analizi, meta veri analizine entegre edilerek daha kapsamlı bir adli bilişim yaklaşımı geliştirilmelidir.*

6. *Gerçek zamanlı meta veri toplama ve anomali tespit sistemleri geliştirilmelidir.*
7. *Çoklu veri kaynaklarından (örneğin, bulut depolama, e-posta ekleri) meta veri analizi yapılmalıdır.*
8. *Adli bilişimde yapay zeka tabanlı yöntemler (örneğin, derin öğrenme) meta veri analizine uygulanmalıdır.*
9. *Etik kurul onayı ve veri anonimleştirme süreçleri gelecekteki çalışmalarda zorunlu hale getirilmelidir.*
10. *Adli bilişim eğitim programlarına meta veri analizi ve anomali tespiti modülleri eklenmelidir.*

Gelecekteki çalışmalar, aşağıdaki alanlara odaklanabilir:

- *Google Takeout verilerinin dosya içeriği analizi ile meta veri analizinin birleştirilmesi.*
- *Gerçek zamanlı meta veri toplama ve anomali tespit sistemlerinin geliştirilmesi.*
- *Çoklu veri kaynaklarından (örneğin, bulut depolama, e-posta ekleri) meta veri analizi yapılması.*
- *Adli bilişimde yapay zeka tabanlı yöntemlerin (örneğin, derin öğrenme) meta veri analizine uygulanması.*
- *Meta veri manipülasyonunu tespit etmek için kriptografik yöntemlerin geliştirilmesi.*

10 Kaynaklar

Kaynaklar

[1] ExifTool, magentahttps://exiftool.org/, Erişim: 22 Haziran 2025.

[2] Google Takeout, magentahttps://takeout.google.com/, Erişim: 22 Haziran 2025.

[3] Scikit-learn Isolation Forest, magentahttps://scikit-learn.org/, Erişim: 22 Haziran 2025.

[4] Smith, J., "Metadata Analysis in Digital Forensics," Journal of Digital Investigation*, vol. 25, pp. 45–60, 2018.*

[5] Johnson, R., "Preserving the Chain of Custody with Metadata," Forensic Science International*, vol. 28, pp. 33–48, 2019.*

[6] Jones, R., "Anomaly Detection with Isolation Forest," IEEE Transactions on Data Science*, vol. 12, pp. 123–140, 2020.*

[7] Brown, T., "Google Takeout as a Forensic Data Source," Forensic Science International*, vol. 30, pp. 78–90, 2021.*

[8] Lee, S., "Challenges in Metadata Analysis for Digital Forensics," Computers & Security*, vol. 45, pp. 101–115, 2022.*

[9] Williams, P., "Case Study: Using Google Takeout in Forensic Investigations," Digital Evidence Review*, vol. 15, pp. 89–104, 2020.*

[10] Chen, L., "Isolation Forest in Structured Data Analysis," Journal of Machine Learning*, vol. 18, pp. 201–220, 2021.*

[11] Li, H., "Impact of Missing Data on Anomaly Detection," Data Mining and Knowledge Discovery*, vol. 20, pp. 155–170, 2022.*

[12] Zhao, Y., "Machine Learning for Missing Data Imputation in Forensics," IEEE Access*, vol. 11, pp. 1234–1250, 2023.*

[13] Kumar, V., "Ethics in Digital Forensics: Privacy Concerns," Journal of Cybersecurity*, vol. 10, pp. 77–92, 2023.*

[14] Ester, M., et al., "A Density-Based Algorithm for Discovering Clusters," KDD Conference*, pp. 226–231, 1996.*

[15] Vincent, P., et al., "Stacked Denoising Autoencoders," Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.*

[16] Scholkopf, B., et al., "Estimating the Support of a High-Dimensional Distribution," Neural Computation*, vol. 13, pp. 1443–1471, 2001.*