Düzce Üniversitesi Bilim ve Teknoloji Dergisi

Derleme Makalesi

Bulut Bilişim Güvenliği İçin Kullanılan Makine Öğrenimi Yöntemleri Üzerine Bir Derleme

Bilge Kağan YAZAR [a,*], Sedat AKLEYLEK [b], Erdal KILIÇ [c]

[a] *Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Ondokuz Mayıs Üniversitesi, Samsun, TÜRKİYE*
** Sorumlu yazarın e-posta adresi: bilgekaganyazar@gmail.com*
DOI: 10.29130/dubited.979040

Öz

Son zamanlarda bulut bilişimin farklı amaçlar için kullanımı artmaktadır. Bu durum bulut üzerindeki bilgilerin çoğalmasına sebep olmaktadır ve daha yüksek güvenlik gereksinimlerinin olduğunu göstermektedir. Güvenliği sağlamanın yollarından bir tanesi makine öğrenmesi yöntemlerinin bulut sistemlerine adapte edilmesidir. Geleneksel yöntemler saldırılardaki çeşitlilik nedeniyle istenilen düzeyde başarı sağlayamamaktadır. Makine öğrenimi yaklaşımları, verileri daha etkin bir şekilde ele aldıklarından daha duyarlı ve otomatikleştirilmiş güvenlik çözümleri sunabilmektedir. Bulut üzerindeki verilerin gizliliği, bütünlüğü, bulut kaynaklarının kullanılabilirliği ve bulut platformu üzerindeki kimlik doğrulama işlemleri için makine öğrenimi tabanlı sistemlerin kullanımı son zamanlarda oldukça popülerdir. Genellikle izinsiz giriş tespit sistemi olarak adlandırılan bu sistemler, bulut uygulamalarındaki bilgileri yetkisiz erişimlerden korumak için kapsamlı yaklaşımlar kullanmaktadır. Bu çalışmada bulut bilişim güvenliği ve bu alanda kullanılan makine öğrenmesi yaklaşımları üzerine bir sistematik literatür taraması yapılmıştır. Kullanılan makine öğrenimi yöntemleri ve değerlendirme kriterleri, kullanılan veri kümeleri ve çalışmaların sağladıkları bilgi güvenliği kavramları baz alınarak, literatürde etkisi olan çalışmalar ele alınmıştır. Bazıları hibrit bazıları bağımsız şekilde 23 farklı makine öğrenimi yöntemi ve 17 farklı değerlendirme ölçütünün kullanıldığı görülmüştür. Toplamda 11 farklı hazır veri kümesi ve sekiz çalışmada ise oluşturulmuş olan veri kümelerinin kullanıldığı görülmüştür. Son olarak çalışmalar gizlilik, bütünlük, erişilebilirlik ve kimlik denetimi olacak şekilde bilgi güvenliği kavramları açısından değerlendirilmiştir.

Anahtar Kelimeler: *Bulut Bilişim Güvenliği, Makine Öğrenimi, Performans Ölçütleri, Veri Kümeleri*

A Review of Machine Learning Methods Used for Cloud Computing Security

Abstract

Recently, the use of cloud computing for different purposes has been increasing. This causes the proliferation of information on the cloud and indicates higher security requirements. One of the ways to ensure security is to adapt machine learning methods to cloud systems. Traditional methods cannot achieve the desired level of success due to the diversity in attacks. Machine learning approaches can offer more responsive and automated security solutions as they handle data more effectively. The use of machine learning-based systems for the confidentiality and integrity of data in the cloud, the availability of cloud resources, and authentication on the cloud platform have been very popular recently. These systems, often called intrusion detection systems, use comprehensive approaches to protect the information in cloud applications from attacks. In this study, a systematic literature review was conducted on cloud computing security and machine learning approaches used in this field. Based on the machine learning methods and evaluation criteria used, the datasets used and the information security concepts provided by the studies, the studies that have an impact on the literature are

Geliş: 05/08/2021, Düzeltme: 03/11/2021, Kabul: 08/11/2021

discussed. It has been observed that 23 different machine learning methods and 17 different evaluation criteria are used, some of the hybrid and some independently. In total, 11 different ready-made datasets and the datasets created in eight studies were used. Finally, the studies were evaluated in terms of information security concepts such as confidentiality, integrity, availability, and authentication.

Keywords: *Cloud Computing Security, Machine Learning, Performance Criteria, Datasets*

I. GİRİŞ

Bulut bilişim (Cloud Computing), internet üzerindeki hizmetleri kolaylaştırmak ve sunmak için son zamanlarda sıklıkla kullanılmakta olan yeni bir yöntemdir. Ulusal Standartlar ve Teknoloji Enstitüsü (National Institute of Standards and Technology) bulut bilişimi hızla sağlanabilen ve piyasaya sürülebilen yapılandırılabilir bilgi işlem kaynaklarının paylaşılan bir havuzda her yerde bulunabileceği, minimum yönetim çabası ve hizmet sağlayıcısı etkileşimi ile isteğe bağlı bir ağ erişimini mümkün kılan bir model olarak tanımlamaktadır [1]. Son yıllarda bulut bilişim üzerinde fazla sayıda çalışma yapılmaktadır ve bununla birlikte önemli yenilikler ortaya çıkmaktadır. Bulut bilişim, sağladığı pratiklik sayesinde çoğu kuruluş tarafından kullanılmaktadır ve çeşitli kolaylıklar sağlamaktadır. Bulut bilişim, altyapı olarak servis (Infastructure as a Service - IaaS), platform olarak servis (Platform as a Service - PaaS) ve yazılım olarak servis (Software as a Service - SaaS) gibi hizmet modellerine sahiptir ve genel, özel, topluluk ve hibrit bulut gibi dağıtım modelleri üzerinde çalışmalar gerçekleştirilmektedir.

Bulut bilişimin yaygınlaşmasıyla birlikte hem sağlanan hizmetlerin hem de hizmet kullanıcılarının güvenliği en önemli endişe kaynaklarından bir tanesi haline gelmiştir. Bulut bilişim güvenliği, bilgi güvenliğinin önemli alt dallarından bir tanesi olarak öne çıkmaktadır. Bulut bilişim hizmetlerinin temel olarak internet protokolleri üzerinden sağlanması ve sanallaştırma tekniklerinin kullanılması, veri güvenliği ihlali, kötü amaçlı yazılımların sisteme yüklenmesi, hizmet reddi (Denial of a Service – DoS) saldırıları gibi verileri ciddi anlamda etkileme potansiyeline sahip saldırılara ve diğer güvenlik tehditlerine karşı bazı zayıflıklar bulunmaktadır [2], [3]. Bulut üzerindeki tehditler, kötü amaçlı veya tesadüfen olabilecek olumsuz bir olaydır [4]. Bulut bilişimde güvenlik tehditleri genellikle gizlilik (confidentiality), bütünlük (integrity) ve erişilebilirlik (availability) olmak üzere üç ana başlık altında sınıflandırılmaktadır [5]:

- Gizlilik tehditleri, kullanıcı verilerine yönelik içeriden bir tehdit, dışarıdan gelecek saldırılar ve veriler ile ilgili sorunları kapsamaktadır.
- Bütünlük tehditleri, zayıf erişim kontrolü ve bilgi kalitesine yönelik riskler, güvenlik parametrelerinin anlamlarını yanlış bir şekilde birleştiren bilgi izolasyonu riski, sanal makinelerin (virtual machine - VM) yanlış tasarımı ve uzak istemci tarafındaki hiper yöneticilerin zayıflıklarını içerir.
- Erişilebilirlik tehditleri, servis sağlayıcı kuruluşun erişilebilir olmaması, varlıkların fiziksel olarak kesintiye uğraması ve verimsiz veri kurtarma, arıza giderme stratejilerini kapsamaktadır.

Bulut platformu üzerinde gerçekleştirilen saldırılar ise ağ tabanlı, sanal makine tabanlı, depolama tabanlı ve uygulama tabanlı olmak üzere dört kısma ayrılmaktadır [6], [7]:

- Ağ tabanlı saldırılar: Genellikle incelenen saldırılar bağlantı noktası taraması (port scan), botnetler ve sahtekarlık (spoofing) saldırılarıdır.
- VM tabanlı saldırılar: Farklı alt yapılardaki farklı VM'ler birden çok güvenlik sorununa neden olabilmektedir. Sanal makine görüntüsünün içine yerleştirilen kötü amaçlı kodlar veya yan kanal saldırıları bu kısım altında incelenmektedir.

- Depolama tabanlı saldırılar: Sistem üzerinde güçlü bir izleme mekanizması yoksa, saldırganlar bazı depolama cihazlarında depolanan önemli verileri elde edebilir. Veri süpürme ve veri tekilleştirme gibi durumlar bu kısım altında değerlendirilmektedir.
- Uygulama tabanlı saldırılar: Bulut üzerinde çalışan uygulamalar, performansı etkileyen ve kötü amaçlı amaçlarla bilgi sızmasına neden olan birçok saldırı ile karşı karşıya kalabilir. Üç ana uygulama tabanlı saldırı; kötü amaçlı yazılım yüklenmesi, stenografi saldırıları, web hizmetleri ve kurallara dayalı saldırılardır.

Bulut bilişim internet üzerinden sağlanan bir hizmet olduğundan çok geniş kapsama sahip bir alandır. Bulut bilişimde sistemi tehdit altına alabilecek bazı güvenlik açıkları; sanallaştırma, çok kullanıcılık, yetkisiz erişim gibi durumlardır [4], [8]. Bu açıklar özel verilerin güvenliğini tehlike altına alacak durumlar oluşturmaktadır ve bulut sisteminin güvenilirliği tehlike altına girmektedir. Bulut ortamındaki genel bilgilerin çoğalmasıyla, aynı şekilde buluttaki hassas bilgilerin de çoğalması söz konusudur ve bu da bulut bilişimde daha yüksek güvenlik gereksinimlerinin olduğunu göstermektedir. Yukarıda bahsedilen saldırılar, güvenlik açıkları ve bahsedilmeyen birçoğunun, tespit edilmesi ve önlenmesi için kullanılan geleneksel yöntemler, büyük veri akışları gerçekleştiğinde yeterli verimlilikte çalışmayabilir [2]. Makine öğrenimi (Machine Learning – ML) yöntemleri, bulut platformlarındaki güvenlik sorunlarını çözmek ve verileri daha etkin yönetmek için kullanılmaktadır. Makine öğrenimi, bilgisayar sistemlerinin modellere ve kabullere bağlı olarak belirli bir yöntemi uygulamak için kullandığı hesaplamaların ve ölçülebilir modellerin mantıksal incelemesi olarak tanımlanabilir. Makine öğrenimi yöntemleri denetimli, yarı-denetimli ve denetimsiz olmak üzere üç ana başlık altında sınıflandırılmaktadır. Her bir başlık için karar ağaçları, destek vektör makineleri, yapay sinir ağları ve k-means gibi çeşitli yöntemler bulunmaktadır. Bunlara ek olarak son zamanlarda popülerliği artan derin öğrenme modelleri de bulut bilişim güvenliği alanında son zamanlarda sıkça kullanılmaktadır.

Literatürde, makine öğrenimi yöntemlerinin bulut platformundaki saldırıların ve kötü amaçlı uygulamaların tespiti için farklı şekillerde kullanıldığı ve başarı oranlarının yüksek olduğu görülmektedir. ML ile güvenliği sağlamak için en çok kullanılan yöntem saldırının gerçekleştiği sırada tespitinin yapıldığı ve devamında hizmet kurtarmanın yapıldığı durumlardır. Diğer bir yöntem ise bulut sisteminin sürekli kendi kendini kontrol etmesi ile saldırı gerçekleşmeden önce önlenmesi yönündedir. Tablo 1'de literatürde yapılan bulut bilişim güvenliği ve makine öğrenimi tabanlı bazı derleme çalışmalarının kapsamları ve eksikliklerinden kısaca bahsedilmiştir. Bu tablo çalışmaların özeti ve bizim çalışmamızın farklılıklarını göstermektedir.

ML ile bulut bilişim güvenliği alanında yapılmış olan sistematik literatür taraması sayısının çok sınırlı olmasıyla birlikte, bu alandaki Türkçe kaynak eksikliği göze çarpmaktadır. Bu çalışmada makine öğrenmesi-derin öğrenme ile bulut bilişim güvenliği alanındaki diğer derleme çalışmalarına göre farklılıklar bulunmaktadır;

1. İncelenen çalışmalarda derin öğrenme modellerini kullanan çok sayıda çalışma seçilmiştir.
2. Bulut bilişim güvenliği alanında farklı güvenlik yönlerini ele alan çalışmalar incelemeye dahil edilmiştir.
3. 2016-2021 yılları aralığında yapılan incelemelerle birlikte, 2019-2021 yılları arasında yayınlanmış olan güncel çalışmalar incelenmiştir.
4. Çalışmalarda sıklıkla kullanılan veri kümeleri detaylı olarak ele alınmıştır.
5. Çalışmalar bilgi güvenliği kavramları açısından değerlendirilmiştir.

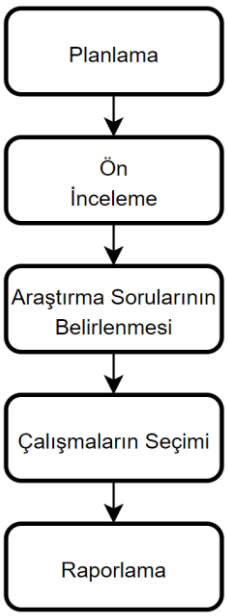
Çalışmanın devamında Bölüm 2'de çalışmaların seçilme aşaması, araştırma yöntemi detaylandırılmıştır ve araştırma soruları belirlenmiştir. Bölüm 3'de belirlenen araştırma sorularına cevaplar aranmıştır, toplanan çalışmaların özetleri verilmiştir ve çalışmalar üzerinden genel bir değerlendirme yapılmıştır. Bölüm 4'te ise elde edilen sonuçlar özetlenmiştir.

Tablo 1. *Literatürdeki bulut bilişim güvenliği ve makine öğrenimi ile ilgili derleme çalışmaları ve bu çalışmaların eksiklikleri*

Çalışma	Özet	Eksikler	Yıl
A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing [8]	Bulut bilişimdeki zayıflıklara ve güvenlik endişelerine odaklanarak bulut bilişim üzerine kapsamlı bir inceleme sunulmaktadır. En önemli güvenlik tehditleri ve mevcut çözümlerin üzerinde durulmuştur. Literatürdeki ML yöntemleri karşılaştırmalı olarak incelenmiş ve sonuçları tartışılmıştır	Bulut bilişim güvenliği ve makine öğrenmesi tekniklerini kapsamasına rağmen yapılan incelemeler yazarların kendi elde ettikleri sonuçlar üzerinden yapılmıştır	2012
An intrusion detection and prevention system in cloud computing: A systematic review [9]	Bu çalışma, bulut bilişim sistemlerindeki izinsiz girişleri tespit etmek ve önlemek için olası çözümleri, en son geliştirilen modelleri ve alarm yönetimi teknikleri hakkında bilgiler içermektedir.	Sadece, bulut bilişimde izinsiz giriş konusu üzerine bir çalışma yapılmıştır	2013
A review on intrusion detection techniques for Cloud computing and security challanges [10]	Bu çalışma bulut üzerindeki izinsiz giriş saldırılarına, sistem türlerine ve içinde makine öğrenmesinin de bulunduğu tekniklerin analizi üzerinedir.	Sadece bulut üzerinde izinsiz giriş durumları incelenmiştir	2015
A survey of intrusion detection techniques in the Cloud [11]	Bu çalışmada, bulut kaynaklarını ve hizmetlerini bilgi güvenliği kavramları açısından etkileyen izinsiz giriş durumları araştırılmıştır. Bulut üzerindeki saldırı tespit ve saldırı önleme sistemleri üzerinde durulmuştur	Makine öğrenimi açısından yeterince incelemeler yapılmamıştır	2017
A survey of deep learning based network anomaly detection [12]	Bu çalışmada ağ anormallik tespitleri için kullanılan derin öğrenme teknikleri incelenmiştir. Çalışmaların karşılaştırmaları ve incelenen alan üzerinde bilgiler paylaşılmıştır	Sadece bulut üzerinde derin öğrenme yöntemlerini kullanan anormallik tespit durumları incelenmiştir.	2019
Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, andopen research issues [13]	Bu çalışmada, (mobil) bulut ortamında bilişsel zekâ yöntemlerini kullanan saldırı tespit sistemlerine ilişkin genel bir bakış açısı sunulmuştur. Saldırı tespiti için bir sınıflandırma tanımlanmıştır ve bilişsel zekâ tabanlı teknikler bağımsız ve hibrit yöntemler olarak sınıflandırılmıştır.	Genel kapsam izinsiz giriş tespit sistemleri üzerinedir	2020
A Review of Machine Learning Algorithms for Cloud Computing Security [7]	Bu çalışmada, bir veya birkaç ML algoritması kullanan bulut bilişim yöntemlerinin, güvenlik tehditlerinin, sorunlarının ve çözümlerinin bir analizi sunulmaktadır. Denetimli, denetimsiz, yarı denetimli ve pekiştirmeli öğrenme dahil olmak üzere bulut güvenliği sorunlarını çözmek için kullanılan farklı makine öğrenimi algoritmaları incelenmiştir. Her tekniğin performansı ve özellikleri, avantajlarına ve dezavantajlarına göre karşılaştırılmıştır.	Derin öğrenme modelleri üzerinde yeterince inceleme yapılmamıştır	2020
Machine Learning for Cloud Security: A Systematic Review [2]	Bu çalışmada, ML, Bulut güvenlik metodolojileri ve teknikleri için Sistematik Literatür İncelemesi (SLR) yapılmıştır. İlgili 63 çalışma analiz edilmiştir ve SLR sonuçları üç ana araştırma alanında kategorize edilmiştir: bunlar farklı bulut güvenlik tehdit türleri, kullanılan makine öğrenimi teknikleri ve performans sonuçlarıdır.	İncelenen çalışmalarda bilgi güvenliği kavramları üzerinde yeterince durulmamıştır	2021
Bu çalışma	Literatürdeki makine öğrenimi yöntemlerini kullanarak bulut güvenliğini sağlayan çalışmaların incelemesini yapmak için bir SLR çalışması yapılmıştır. Seçilen 15 çalışma incelenmiştir ve üç araştırma sorusu üzerinde durulmuştur. Bunlar; bulut bilişimde güvenliği sağlamak adına kullanılan makine öğrenmesi – derin öğrenme yöntemleri ve performans ölçütleri, kullanılan veri kümeleri ve seçilen çalışmaların sağladıkları bilgi güvenliği kavramlarıdır.	-	-

II. ARAŞTIRMA YÖNTEMİ

Bu çalışmada sistematik bir literatür incelemesi yapılmıştır. Çalışma yapılırken süreç belirli aşamalara ayrılmıştır. Bu aşamalar Şekil 1'de görülmektedir. Bu bölümün devamında çalışmada takip edilen adımlar detaylandırılacaktır.



Şekil 1. *SLR çalışması yapılırken takip edilen adımlar*

A. ARAŞTIRMA SORULARI

Bu çalışmadaki ana amaç 2016'dan günümüze kadar bulut bilişim güvenliğini sağlamak için kullanılan ML yöntemlerini ve tekniklerini incelemektir. Literatürdeki çalışmalar incelendikten sonra detaylı olarak irdelenmemiş üç araştırma sorusu ortaya çıkmıştır:

AS1: Bulut bilişim güvenliği uygulamalarında hangi makine öğrenimi, derin öğrenme yöntemleri tercih edilmektedir ve hangi değerlendirme kriterleri kullanılmaktadır?

AS2: Seçilen çalışmalarda kullanılan veri kümeleri ve özellikleri nelerdir?

AS3: Bu çalışmalar bilgi güvenliği kavramlarından hangileri üzerinde yoğunlaşmaktadır?

B. ARAMA YÖNTEMİ

Bu çalışma için kaynakların toplanması; IEEE Xplore, Scopus, Web of Science ve ScienceDirect olmak üzere dört veri tabanı üzerinden yapılmıştır. Bu veri tabanları mühendislik alanında sistematik literatür taramalarında sıklıkla kullanılmaktadır ve otomatikleştirilmiş arama araçlarına sahip olması bakımından büyük kolaylıklar sağlamaktadır. Bu veri tabanları üzerinden bilimsel çalışma araştırması şu işleyişe göre yapılmıştır:

- Araştırma sorularının oluşmasıyla birlikte yapılan aramalar "cloud security" ve "machine learning" anahtar kelimeleri üzerinde yoğunlaştırılmıştır.
- "cloud security" anahtar kelimesi ile eş anlamlı olabilecek "cloud data security", "cloud computing security" gibi ve "machine learning" anahtar kelimesine ek olarak kullanılabilecek diğer yöntemleri barındıran çalışmalara ulaşabilmek adına, "transfer learning", "ensemble

learning", "reinforcement learning" gibi anahtar kelimeler oluşturulmuştur ve bu kelimeler kullanılmıştır.

- Veri tabanlarında arama yapılırken mantıksal operatörler kullanılmıştır. Eş anlamlı anahtar kelimeler için OR ve anahtar kelimeleri birleştirmek için AND operatörleri kullanılmıştır.

Tablo 2. *Veri tabanlarında arama yapılırken kullanılan sorgular*

Veritabanı	Sorgu
IEEE Xplore	(("cloud security" OR "cloud data security" OR "cloud server security" OR "cloud attack security" OR "cloud computing security "OR "cloud service security" OR "cloud malware detection" OR "cloud attack detection") AND ("machine learning" OR "deep learning" OR "transfer learning" OR "ensemble learning" OR "reinforcement learning"))
Web of Science	TS = (("cloud security" OR "cloud data security" OR "cloud server security" OR "cloud attack security" OR "cloud computin gsecurity" OR "cloud service security" OR "cloud malware detection" OR "cloud attack detection") AND ("machine learning" OR "deep learning" OR "transfer learning" OR "ensemble learning" OR" reinforcement learning"))
Scopus	TITLE-ABS-KEY (("cloud security" OR "cloud data security" OR "cloud server security OR "cloud attack security" OR "cloud computing security" OR "cloud service security" OR "cloud malware detection" OR"cloud attack detection") AND ("machine learning" OR "deep learning" OR "transfer learning" OR "ensemble learning" OR "reinforcement learning")) AND PUBYEAR AFT 2015
ScienceDirect	(("cloud security" OR "cloud data security" OR "cloud attack security" OR "cloud computing security" OR "cloud service security" OR "cloud malware detection" OR "cloud attack detection") AND ("machine learning" OR "deep learning"))

Tablo 2'de veri tabanlarında arama yapmak için kullanılan sorgular verilmiştir. ScienceDirect veritabanı için yazılan sorgunun diğerlerine göre kısa olmasının sebebi, kullanılacak sorgular için sekiz tane mantıksal operatör sınırı bulunmasıdır. Google Scholar, Springer vb. diğer veritabanları üzerinde Tablo 2'de yazıldığı gibi sorgular yazılamadığından veya yazılsa bile anlamlı sonuçlar elde edilemediğinden (çok fazla ilgisiz sonuç çıkması, incelenemeyecek kadar çok sayıda çalışma çıkması) veri tabanı olarak sadece sonuç alınabilen dört veri tabanı kullanılmıştır. Yapılan bu sorguların sonucunda, 2016-2021 yılları arasında yayınlanmış IEEE Xplore'da 80 tane, WOS'da 59 tane, Scopus da 120 tane ve ScienceDirect de 501 tane olmak üzere toplam 760 tane çalışma (arama yapılan tarih itibariyle) görülmüştür. Bu çalışmalarla ilgili sayısal bilgiler Tablo 3'te verilmiştir.

Tablo 3. *Veri tabanlarında yapılan aramalar sonucunda elde edilen çalışma sayıları*

Veritabanı	Arama Sonucu	Konferans Bildirisi	Dergi Makalesi	Diğer (Derleme, Kitap bölümü vs.)
IEEE Xplore	80	64	15	1
Web of Science	59	30	26	3
Scopus	120	62	39	19
ScienceDirect	501	-	390	111
Toplam	760	156	470	134

C. ÇALIŞMA SEÇİMİ

Yapılan aramalar sonucu 760 tane çalışma bulunmuş olsa da, bu çalışmaların büyük bir çoğunluğu kapsam dışında kalmaktadır. Özellikle ScienceDirect veri tabanında çalışmaların tümü üzerinden arama yapıldığından konu ile ilgisi olmayan birçok çalışma çıkmıştır. Bunlara ek olarak bazı çalışmalar arama sonuçlarında birden çok veri tabanında çıkabilmektedir. Çalışma seçimi işlemine

başlanmadan önce bu şekilde tekrarlayan çalışmalar tek bir veri tabanından alınacak şekilde düzenlenmiştir. Devamında bu çalışmanın amacına uygun olan çalışmaların dahil olması için bazı ekleme ve çıkarma kriterleri belirlenmiştir. Ekleme kriterleri şunlardır;

- Bulut bilişim güvenliği alanındaki çalışmalar
- Bulut bilişim güvenliği alanında makine öğrenmesi teknikleri kullanılan çalışmalar
- Hakemli dergilerde yayınlanmış olan makaleler
- 2016-2021 aralığında yayınlanmış olan çalışmalar

Çıkarma kriterleri ise;

- Makine öğrenimi yöntemlerini kullanmadan bulut bilişim güvenliğini ele alan çalışmalar
- Bulut bilişim güvenliği ve makine öğrenimi yöntemlerini içermeyen çalışmalar

şeklindedir. Ön inceleme aşamasında genellikle çalışmaların özetleri üzerinden incelemeler yapılmıştır. Ancak bazı çalışmaların tamamının okunmasının gerektiği durumlarda olmuştur. Belirlenen kriterler doğrultusunda 15 tane çalışma incelenmek üzere seçilmiştir. Bu çalışmalar Tablo 4'te görülmektedir. Yapılan incelemelerde makine öğrenmesi ile bulut bilişim güvenliği alanına özellikle 2018 yılından sonra ilginin arttığı görülmektedir ve Tablo 4'te görülüğü üzere seçilen çalışmaların hepsi 2019-2021 yılları arasında yayınlanmıştır. Bu açıdan bakıldığında yapmış olduğumuz inceleme güncel bir yapıdadır.

Tablo 4. *İncelenmek üzere seçilen çalışmalar*

Çalışma Adı	**Yıl**	**Referans**
A deep learning approach for proactive multi-cloud cooperative intrusion detection systems	2019	[14]
A focus on future cloud: machine learning-based cloud security	2019	[15]
A hybrid machine learning approach for malicious behavior detection and recognition in cloud computing	2020	[16]
A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogenous client networks	2019	[17]
An intrusion detection system for connected vehicles in smart cities	2019	[18]
DeMETER in clouds: detection of malicious external thread execution in runtime with machine learning in PaaS clouds	2020	[19]
Detection and mitigation of fraudulent resource consumption attacks in cloud using deep learning approach	2021	[20]
Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing	2020	[21]
Intelligent approach to build a Deep Neural Network based IDS for cloud environment using combination of machine learning algorithms	2019	[22]
KVMInspector: KVM Based introspection approach to detect malware in cloud environment	2020	[23]
Multilayer Self-Defense System to Protect Enterprise Cloud	2020	[24]
Secure data transmission and detection of anti-forensic attacks in cloud environment using MECC and DLMNN	2020	[25]
TIDCS: A Dynamic Intrusion Detection and Classification System Based Feature Selection	2020	[26]
Cryptomining Detection in Container Clouds Using System Calls and Explainable Machine Learning	2021	[27]
A machine learning application for recducing the security risks in hybrid cloud networks	2020	[28]

III. ELDE EDİLEN BULGULAR VE TARTIŞMA

Çalışmanın bu kısmında literatürdeki ML ile bulut bilişim güvenliği alanında literatürdeki çalışmaları ve kullanılan yöntemleri incelemek için yapılan sistematik tarama sonucunda araştırma soruları (AS) için elde edilen çıktılar alt bölümlerde paylaşılmış ve detaylandırılmıştır.

A. AS1: MAKİNE ÖĞRENMESİ YÖNTEMLERİ VE DEĞERLENDİRME KRİTERLERİ

Geleneksel yöntemlerin kullanımı ile bulut üzerindeki tehditlerin engellenmesi artan saldırı sayısı ve saldırı çeşitliliği nedeniyle yetersiz kalmaktadır. Makine öğrenimi yaklaşımları, verileri daha etkin bir şekilde ele aldıklarından daha duyarlı ve otomatikleştirilmiş güvenlik çözümleri sunabilmektedir. Bulut üzerindeki verilerin gizliliği ve bütünlüğü, bulut kaynaklarının kullanılabilirliği ve bulut platformu üzerindeki kimlik doğrulama işlemleri için makine öğrenimi tabanlı sistemler son zamanlarda oldukça popülerdir. Genellikle izinsiz giriş tespit sistemi (Intrusion Detection System - IDS) olarak adlandırılan bu sistemler, bulut uygulamalarındaki bilgileri saldırılardan korumak için kapsamlı yaklaşımlar kullanmaktadır. Bu bölümün devamında incelenen çalışmalar kısaca özetlenmiştir ve araştırma sorusu için elde edilen sonuçlar paylaşılmıştır.

[14] nolu çalışmada yazarlar, farklı bulut sağlayıcılarına ait olan IDS'ler arasındaki iş birliğinin karşılıklı yararlar sağlayarak model performanslarını artırabileceği düşüncesi üzerinde durmuşlardır. Ancak, bu tarz sistemlerde IDS'ler arasındaki bilgi paylaşımında ve bilgi toplanırken gecikmeler olduğundan bahsedilmiştir ve bunun üzerine IDS'lerden kısmi veya eksik bilgi geldiği düşünülerek bir sistem oluşturulmuştur. Önerilen model diğer IDS'lerden gelen eksik bilgilere karşı gerekli özellikleri nasıl çıkartacağını öğrenebilen bir yapıdadır. Bu sayede modelin gerçek zamanlı ortamlarda karar verme işlemini hızlandıracağı belirtilmektedir. Çalışmada eksik veriyi yeniden yapılandırmak için bir derin öğrenme yöntemi olan yığın gürültü giderici otokodlayıcı (Stacked Denoising Autoencoder) modeli kullanılmıştır. Sınıflandırma işlemi içinse derin sinir ağları (Deep Neural Networks – DNN) kullanılmıştır. Önerilen model IDS'lerin eksik geri bildirimlerine karşı, kısmi veriler üzerinden tahminler yaparak gerçek zamanlı ortamlarda daha iyi bir algılama doğruluğu sağlamaktadır. Elde edilen sonuçlar çok katmanlı algılayıcı (Multilayer Perceptron) ve yığın otokodlayıcı (Stacked Autoencoder) modelleri ile karşılaştırılmıştır ve daha iyi sonuçlar elde ettiği görülmüştür.

[15] nolu çalışmada önerilen sistem, makine öğrenimi tekniklerini kullanarak dağıtık bulut tabanlı ortamlarda güvenliği artırmak amacıyla önerilmiştir. Yazarlar doğrusal regresyon, destek vektör makinesi gibi yöntemlerin statik yapıları sebebiyle güvenlik sunabilecekleri çözümlerinin sınırlı olduklarını söylemektedir. Çalışmada ele alınan temel durum evrişimsel sinir ağı (Convolutional Neural Network - CNN) kullanarak bulut ağındaki trafik analizini gerçekleştirmek üzerinedir. Önerilen modelde CNN'den elde edilen çıktıları çok sınıflı destek vektör makinesine (Multiclass Support Vector Machine - MSVM) girdi olarak verilmektedir ve sınıflandırma işlemi bu şekilde yapılmaktadır. Önerilen model karar ağaçları, Naive Bayes, lojistik regresyon, destek vektör makinesi (doğrusal ve farklı çekirdekler ile) yöntemleri ile karşılaştırılmıştır. Önerilen şekilde bir kullanım ile geleneksel modellere daha güçlü temsiller sağlanabildiğinden daha iyi sonuçların alındığı görülmektedir.

[16] nolu çalışmada bulut ortamındaki güvenliği artırmak amacıyla, kullanıcıların kimlik bilgilerini kullanmak yerine kullanıcı davranışlarını gözlemlemeye dayalı bir yaklaşım önerilmiştir. Ağ trafiğindeki anlamlı bilgileri ve kullanıcıların davranış kalıplarını kendi kendine optimize edilmiş bir şekilde bulabilen ve bunları kötü niyetli davranışları tespit etmek için kullanan bir makine öğrenimi yaklaşımı sunulmuştur. Önerilen model ilk olarak eldeki verilere temel bileşen analizi (principal component analysis) uygulayarak verilerin boyutunu indirgemektedir. Kendi kendini optimize eden bir yapı oluşturabilmek içinse parçacık sürü optimizasyonu (Particle Swarm Optimization) yöntemini kullanan olasılıksal sinir ağları (Probabilistic Neural Networks) kullanılmaktadır. Çok sınıflı ve iki

sınıflı olmak üzere model iki farklı şekilde kullanılmaktadır. Genel olarak bu çalışmada, kötü niyetli davranış örneklerini otomatik olarak tanımlamak için kullanılabilecek etkili bir çözüm sunulmuştur.

[17] nolu çalışmada heterojen istemci ağlarını içeren mobil bulut ortamlarında dağıtılmış hizmet reddi (Distributed Denial of Service - DDoS) ve aradaki adam (Man in the Middle - MITM) saldırılarına karşı güvenlik sağlayan makine öğrenimi tabanlı bir sistem önerilmiştir. Önerilen şema, bulutun bir kısmı tehlikeye girdiğinde hızlı hizmet kurtarmaya olanak tanıyan bulut bulutu olarak adlandırılan bir bulut modelini benimsemektedir. Teknik olarak önerilen yöntem, toplanan konum ve işletim sistemi bilgilerine göre trafik taraması yapan ve uygun sanal makineyi seçen bilişsel bir sistemdir. Bu işleme ek olarak dış bulut katmanında kimlik doğrulaması işlemi de yapılmaktadır. Çalışmada literatürdeki iki kümeleme algoritması olan k-means ve DBSCAN algoritmalarının bir birleşimi önerilmiştir. k-means ile DBSCAN algoritmalarının birlikte kullanımları ile elde edilen kümeler daha az gürültüye sahip olmaktadır. Önerilen şema kullanım şekline göre özelleştirilebilir yapıda olmasıyla, sanal makinelerdeki güvenlik sorunlarını çözebilme kapasitesi ve kümeleme yaklaşımını kullanan tek çalışma olmasıyla öne çıkmaktadır.

[18] nolu çalışmada yazarlar bulut aracılığıyla birbirine bağlı olan araçlar arasındaki bilgi paylaşımının gizlilik, bütünlük ve erişilebilirlik problemlerine sebep olabileceğini söylemektedir. Bu durumdan dolayı bu çalışmada, bağlı araçlar için gerçekleşebilecek saldırılara karşı bir IDS mekanizması sağlayan sürekli bir bulut hizmeti kullanılabilirliği çerçevesi sunulmaktadır. Sunulan çalışmada servis talep eden akıllı araçların her biri bir kümeye dahildir ve küme başları ile bu taleplerini servis sağlayıcılarına (güvenilir üçüncü taraf) iletmektedirler. Ortamda gerçekleşebilecek bir saldırının tespitinde ise, veri boyutunu indirgemek için bir derin öğrenme yöntemi olan kısıtlanmış boltzmann makinası (Restricted Boltzmann Machine) ve sınıflandırma içinde ID3 algoritması kullanılmaktadır. Önerilen yöntem olasılıksal sinir ağları, genetik algoritma ve derin inanç ağlarını kullanan yöntemlerle kıyaslanmıştır. Önerilen yöntem veri boyutu indirgerken aynı anda gerekli özellikleri seçerek daha düzgün bir sınıflandırma yapılmasını sağlamaktadır ve karşılaştırma yapılan yöntemlerden daha iyi sonuçlar elde etmiştir.

[19] nolu çalışmada çok kullanıcılı bulut ortamlarında bir bulut uygulamasına yapılan talebin iyi niyetlimi kötü niyetlimi olduğunu anlamak için bir yöntem önerilmiştir. İşlemci (CPU) kullanım ölçümleri ve kaynak erişimlerinin izlenmesi ile birlikte makine öğrenimi teknikleri kullanılmaktadır. Önerilen mekanizma çok çeşitli senaryoları kapsamaktadır. Veri tabanlarındaki CRUD (Create - Read - Update - Delete) işlemlerine denk gelebilecek bir çok saldırı simule edilmiştir. Çalışmada birçok makine öğrenimi yöntemi kullanılmış ve birbirleriyle kıyaslanmıştır. Kullanılan yöntemler; k-NN, rastgele orman, yığma (stacking), Bayes ağları, torbalama (bagging), AdaBoost yöntemleridir. Bu yöntemler arasından en iyi sonuçlar rastgele orman yöntemi ile elde edilmiştir.

[20] nolu çalışmada sahte kaynak tüketimi (fraudulent resource consumption - FRC) saldırılarının tespiti üzerinde durulmuştur. Bu saldırı çeşidi bulut sağlayıcılarının ekonomik kaynaklarını hedefleyen bir ekonomik DoS (EDoS) saldırısı olarak sınıflandırılmaktadır. Bu saldırılar DDoS saldırılarından farklı olarak bulut sağlayıcılarına mali yük getirmektedir ve yapılan istek oranı düşük olduğundan IDS'ler tarafından tespit edilememektedir. Bu çalışmada FRC davranışının tespiti için web sunucu günlükleri (log) kullanılarak bir LSTM (Long-Short Term Memory) ağı kullanımı önerilmiştir. Zaman serisi verileri oluşturulmuştur ve dalgacık dönüşümü tabanlı ön işleme adımından geçirildikten sonra veriler modele beslenmiştir. Çeşitli ağlar farklı saldırı yüzdeleri ile eğitilmiştir ve yapılan saldırı yüzdesi tahminine göre sınıflandırıcılardan birisi kullanılmaktadır. Önerilen model rastgele orman, lojistik regresyon, destek vektör makinesi, k-NN, yapay sinir ağları gibi literatürdeki birçok yöntem ile kıyaslanmıştır. Önerilen model bütün yöntemlerden iyi sonuçlar elde ederken sadece yapay sinir ağları yakın sonuçlar elde etmiştir.

[21] nolu çalışmada topluluk öğrenmesi (ensemble learning) yöntemi kullanılarak yararlı özellikleri bulmak ve ağ saldırılarını sınıflandırmak amaçlanmıştır. Saldırı tespiti için tek değişkenli bir toplu filtre özellik seçimi yöntemi önerilmiştir. Model özellik çıkarım yönteminden sonra destek vektör makinesi, lojistik regresyon, Naive Bayes ve karar ağaçları sınıflandırıcıları ile çoğunluk oylama

yöntemi (majority voting) kullanılarak sınıflandırma yapmaktadır. Veri kümeleri içerisinden daha düzgün temsillerin daha düşük boyutlar ile elde edilmesi ve çoğunluk oylama yöntemi kullanılarak sınıflandırma yapılması başarılı sonuçlar alınmasını sağlamıştır.

[22] nolu çalışmada bulut ortamları için DNN, genetik algoritma (GA) ve benzetimli tavlama algoritması (Simulated Annealing Algorithm - SAA) kullanılarak, uygun hesaplama maliyetine sahip olan ve saldırı tespiti gerçekleştiren bir model oluşturulması amaçlanmıştır. Önerilen sistem bulut ortamının ön (front-end) ve arka (back-end) yüzünde çalışmaktadır. Veriler DNN'ye verilirken modelin katmanlarındaki düğüm sayıları GA kullanılarak hesaplanmaktadır. GA'yı optimize etmek içinse SAA kullanılmaktadır. Ek olarak bu işlemlerden önce öznitelik seçim işlemi de yapılmaktadır. Bu çalışmanın diğer bir özelliği de kullanılan veri kümelerinden çok detaylı olarak bahsetmesidir. Önerilen yöntem birçok model ile kıyaslanmıştır ve kıyaslanan modellerden daha iyi sonuçlar elde etmekle birlikte neredeyse %100 başarı ile çalışmaktadır.

[23] nolu çalışmada yazarlar bulut ortamındaki anahtar teknolojinin, temel donanım ve yazılım üzerinde bir soyutlaştırma katmanı oluşturmaya yardımcı olan sanallaştırma işlemi olduğundan bahsetmektedir. Kötü amaçlı yazılımları tespit etmek için ve çekirdek sanal makine (Kernel Virtual Machine - KVM) tabanlı bulut ortamında güvenlik açısından kritik süreçlerin varlığını devam ettirmek için iç gözlem tabanlı bir güvenlik mekanizması önerilmiştir. KVM tabanlı bulut ortamında gelişmiş saldırıları tespit etmek için süreçler üzerinde ayrıntılı analizler ile gelişmiş bir güvenlik çözümü sunulmaktadır. Çalışmada veri olarak bulut içi ve bulut dışı sistem çağrıları gerekli bilgileri elde etmek için kullanılmaktadır. Özellik çıkarımı is N-gram algoritması ile yapılmaktadır. Çıkarılan özellikler içinden gerekli olanları seçmek içinse özyinelemeli özellik çıkarımı (Recursive Feature Extraction) yöntemi kullanılmaktadır. Tespit kısmında ise karar ağacı, rastgele orman, AdaBoost, k-NN, Naive Bayes, destek vektör makinesi ve bu yöntemleri oylama yöntemi ile birleştirerek topluluk öğrenimi şeklinde modeller kullanılmıştır. En iyi sonuç rastgele orman yöntemi ile elde edilmiştir.

[24] nolu çalışmada DDoS saldırılarını tespit etmek, azaltmak ve bulut ortamını korumak için yazılım tanımlı ağ (Software Defined Network) tabanlı çok katmanı bir kendini savunma sistemi önerilmiştir. DDoS saldırıları simule edilerek veriler oluşturulmuştur. Önerilen sistemde ağ trafiği sınıflandırması için destek vektör makinesi kullanılmaktadır. Servis olarak Spark (Apache) adı verilen bir yapı kullanılarak gerçek zamanlı verileri daha hızlı ve düzgün işleme kapasitesine sahip bir model önerilmiştir.

[25] nolu çalışmada bulut üzerinde güvenli veri iletişimi ve anti adli saldırıların tespiti için bir model önerilmiştir. Önerilen yaklaşım eliptik eğri kriptografi ve DNN kullanmaktadır. DNN burada veri şifreleme işlemi yapıldıktan sonra iletilirken saldırıya uğramış mı uğramamış mı kontrolü için kullanılmaktadır. Ancak, DNN kullanılırken standart kullanımdan farklı olarak optimizasyon işlemi için mürekkepbalığı algoritması (Cuttle Fish Algorithm) kullanılmaktadır. Kullanılan derin öğrenme yöntemi ile birlikte önerilen şifreleme yönteminin kapsamlı bir güvenlik analizi yapılmıştır. Önerilen derin öğrene modeli ise k-NN, DNN yöntemleri ile farklı ölçütler kullanılarak kıyaslanmıştır ve daha iyi sonuçlar elde edilmiştir.

[26] nolu çalışmada önerilen sitem, bulut ağındaki elemanlar arasındaki güven ilişkisinin periyodik olarak değerlendirildiği ve yenilendiği bir fikirdir. Önerilen model periyodik olarak yenilenen düğümler arasındaki güven ilişkisine dayalı olarak izinsiz girişleri tespit etmektedir. Daha az özellik ile daha iyi sonuçlar elde etmek ve makine öğrenimi algoritmaları için en iyi özelliklerin seçilmesi amacıyla bir özellik seçim algoritması önerilmiştir. Önerilen yöntem zaman kullanımını azaltarak özelliklerin bir alt kümesini rastgele olarak oluşturmaktadır. Çıkartılan özellikler devamında karar ağacı ve rastgele orman yöntemleri ile test edilmiştir. Önerilen şekilde öznitelik seçimi yapıldığında kullanılan geleneksel modellerin performanslarının %20'ye kadar gelişim gösterdiği görülmektedir.

[27] nolu çalışmada bulut kaynaklarının sınırlı olduğu ve birçok kullanıcı tarafından kullanıldığı vurgusu yapılmıştır. Asıl amaç ise bulut üzerindeki kaynak kullanımının kontrolünü sağlamaktır. Kripto madencilik uygulamaları çok fazla kaynak tükettiklerinden bulutta erişilebilirlik sorunlarına yol

açabilmektedir. Çalışmada madencilik uygulamalarının tespiti için Linux sistem çağrıları özellik olarak kullanılmıştır. Tespit aşamasında karar ağacı, XgBoost, yapay sinir ağları ve LSTM yöntemleri kullanılmıştır. Karar ağaçları kullanımında en iyi sonuçlar elde edilmiştir.

[28] nolu çalışmada verilerin bulut üzerinde depolanması sırasında ve verileri bulut veri tabanından alırken veya verilere erişirken güvenliği sağlamak için bir yöntem önerilmiştir. Ayrıca verilerin güvenli bir şekilde depolanması ve geri çağırma işlemi sırasında veri tekrarının önlenmesi için bir veri tekilleştirme süreci de önerilmiştir. Önerilen mekanizmada geliştirilmiş C4.5 algoritması kullanılarak dinamik bir erişim kontrol mekanizması kullanılmaktadır. Kontrol mekanizmasının buradaki amacı bulut kullanıcılarını sınıflandırarak izinsiz girişleri engellemektir.

Tablo 5. *İncelenen çalışmalarda kullanılan yöntemler ve kullanım şekilleri*

Referans	Kullanılan Modeller	Model Kullanımı	Sınıf/Küme Sayısı	Model Türü
[14]	Yığın Gürültü Giderici Otokodlayıcı + Derin Sinir Ağı	Hibrit	İki	Sınıflandırma
[15]	CNN + MSVM	Hibrit	Çoklu	Sınıflandırma
[16]	Olasılıksal Sinir Ağı	Bağımsız	İkili ve Çoklu	Sınıflandırma
[17]	K-Means + DBSCAN	Hibrit	Çoklu	Kümeleme
[18]	Kısıtlanmış Boltzman Makinesi + ID3	Hibrit	Çoklu	Sınıflandırma
[19]	K-NN Rastgele Orman Yığma Bayes Ağları Torbalama AdaBoost	Bağımsız ve Topluluk	Çoklu	Sınıflandırma
[20]	LSTM	Bağımsız	İki	Sınıflandırma
[21]	Destek Vektör Makinesi Lojistik Regresyon Naive Bayes Karar Ağacı	Bağımsız ve Topluluk	Çoklu	Sınıflandırma
[22]	Genetik Algoritma + Derin Sinir Ağı	Hibrit	İki	Sınıflandırma
[23]	Karar Ağacı Rastgele Orman AdaBoost K-NN Naive Bayes Destek Vektör Makinesi	Bağımsız ve Topluluk	Çoklu	Sınıflandırma
[24]	Destek Vektör Makinesi	Bağımsız	İki	Sınıflandırma
[25]	Derin Sinir Ağı	Bağımsız	İki	Sınıflandırma
[26]	Karar Ağacı Rastgele Orman	Bağımsız	Çoklu	Sınıflandırma
[27]	Karar Ağacı XgBoost Yapay Sinir Ağları LSTM	Bağımsız ve Topluluk	İki	Sınıflandırma
[28]	C4.5	Bağımsız	-	Sınıflandırma

Tablo 5'te incelenen çalışmalar ve kullanılan yöntemlerle ilgili bazı bilgiler paylaşılmıştır. Çalışmalar incelendiğinde bulut bilişim alanında güveliği sağlamak için toplamda 23 farklı makine öğrenimi yönteminin kullanıldığı görülmüştür. Kullanılan yöntemlerin genellikle birbirinden bağımsız şekilde kullanıldığı görülmüştür. Ancak farklı yöntemleri birleştirerek hibrit ve topluluk öğrenimi şeklinde kullanan çalışmalarda mevcuttur. Tablo 5'te görüldüğü üzere önerilen modeller genellikler

sınıflandırma yapmak üzerinedir. Kümeleme yaklaşımını kullanan sadece bir tane çalışma bulunmaktadır. Kullanılan makine öğrenimi yöntemleri arasında en çok tercih edilenler; altı tane ile karar ağaçları, beş tane ile sinir ağları (yapay sinir ağları ve derin sinir ağları) ve dört tane ile destek vektör makinesi yöntemleridir. 15 tane çalışma içerisinde yedi tane çalışma derin öğrenme yöntemlerini kullanırken, sekiz tanesi geleneksel makine öğrenimi yöntemlerini kullanmaktadır. Farklı yöntemleri birleştirerek elde edilen hibrit yöntemlerini kullanan beş çalışmadan dört tanesi derin öğrenme yöntemlerini kullanmaktadır. Geleneksel makine öğreniminde ise modelleri bağımsız kullanmak yerine topluluk öğrenimi kullanmak daha iyi sonuçlar alınmasını sağlayabilmektedir ve dört çalışmada sadece bu şekilde bir kullanım mevcuttur.

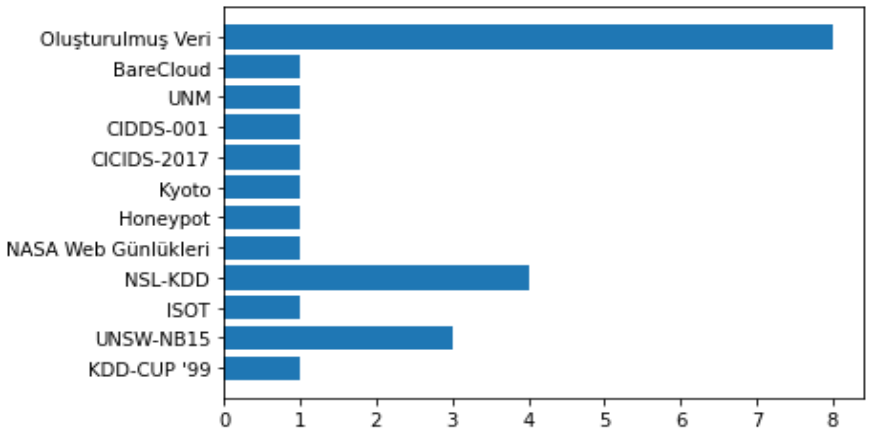
Tablo 6. *İncelenen çalışmalarda kullanılan performans ölçüm metrikleri*

	Çalışma													
Performans Ölçüm Yöntemi	[14]	[15]	[16]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]
Doğruluk (Accuracy)	✓	✓		✓		✓	✓	✓	✓		✓	✓	✓	✓
Gerçek Pozitif Oranı (True Positive Rate – TP, Sensitivity)		✓	✓		✓				✓	✓	✓			
Yanlış Negatif Oranı (False Negative Rate – FN)		✓		✓	✓	✓		✓	✓					
Gerçek Negatif Oranı (True Negative Rate – TN, Specivity)		✓			✓			✓	✓	✓	✓			
Yanlış Pozitif Oranı (False Positive Rate – FP)		✓	✓	✓	✓	✓		✓	✓			✓		
Kesinlik (Precision, Decision Rate)			✓	✓	✓	✓	✓	✓		✓	✓		✓	
Duyarlılık (Recall)					✓	✓	✓			✓	✓		✓	
F1 - Puanı			✓		✓			✓		✓	✓		✓	
ROC - AUC							✓	✓					✓	
Algılama Oranı (Detection Rate)				✓				✓				✓		
Yanlış Alarm Oranı (False Alarm Rate)							✓							
Mathew Korelasyon Katsayısı											✓			
Negatif Tahmin Oranı											✓			
Pozitif Tahmin Oranı											✓			
Cohen's Kappa													✓	
Test Sınıflandırma Hatası	✓													
Yanlış Sınıflandırma Doğruluğu					✓									

İncelenen çalışmalarda doğru makine öğrenimi yöntemini seçmek, bu yöntemin etkili ve doğru çalıştığını göstermek için önerilen modeller diğer modeller ile kıyaslanmaktadır. Bazı çalışmalarda birden fazla ML yöntemi ile bu kıyaslamalar yapılırken, bazı çalışmalarda ise önerilen modele yakın modeller ile kıyaslamalar yapılmaktadır. Bu karşılaştırmalar için bazı performans ölçüm metrikleri kullanılmaktadır. İncelenen çalışmalar doğrultusunda toplamda 17 tane değerlendirme ölçütünün kullanıldığı belirlenmiştir. Bunlarla ilgili bilgiler Tablo 6'da detaylandırılmıştır. Tablo 6 incelendiğinde önerilen modellerin ne kadar efektif çalıştığını göstermek adına 11 çalışmada doğruluk oranının kullanıldığı görülmektedir. Kesinlik ölçütü 2. en çok kullanılan ölçüttür ve bir modelin olumlu bir sonucu ne sıklıkta doğru tahmin ettiğini ifade etmektedir. Toplamda dokuz çalışmada kullanılmıştır. Yanlış pozitif oranı normal bir verinin yanlış sınıflandırılma oranını ifade etmektedir ve toplamda sekiz çalışmada kullanılmıştır. Gerçek pozitif oranı normal bir verinin doğru sınıflandırılma oranını ifade etmektedir ve altı çalışmada kullanılmıştır. Bu ölçütleri yine altı çalışmada kullanılmış olmaları ile yanlış negatif oranı, duyarlılık ve F1-puanı takip etmektedir. [17] nolu çalışmada kümeleme işlemi yapıldığından küme sayısı gibi farklı bilgiler paylaşılmaktadır ve karşılaştırma yapabilecek herhangi bir ölçüt kullanılmamıştır.

B. AS2: KULLANILAN VERİ KÜMELERİ VE ÖZELLİKLERİ

Veri kümeleri, bir modelin değerlendirmesini yapabilmek için çok önemlidir ve en iyi sonucu elde etmede önemli bir rol oynamaktadır. İncelenen çalışmalarda toplamda 11 tane hazır veri kümesinin ve bunlara ek olarak bazı çalışmalarda yazarların kendi oluşturdukları veri kümelerinin kullanıldığı görülmektedir. Kullanılan veri kümelerinin çalışmalara göre sayısı ile ilgili bir grafik Şekil 2'de verilmiştir. Şekil 2'de görüldüğü üzere hazır veri kümeleri olarak en çok tercih edilenler NSL-KDD ve UNSW-NB15 veri kümeleridir ve bundan dolayı bu veri kümeleri detaylandırılmıştır.



Şekil 2. *Çalışmalarda kullanılan veri kümeleri ve kullanım sayıları*

NSL-KDD veri kümesi KDD CUP '99 veri kümesinin gelişmiş bir versiyonu olarak 2009 yılında hazırlanmıştır [29]. Kayıt başına 43 özellik içerir, özelliklerin 41'i trafik girdisi ve son ikisi etiketler (normal veya saldırı) ve skor (trafik girdisinin önem derecesi) şeklindedir. En önemli özelliklerinden bir tanesi, eğitim ve test aşamasında kayıt sayısı arasındaki dengesizlik sorununu ortadan kaldırarak, yanlış alarm oranlarını azaltmasıdır. Veri kümesi içerisinde dört farklı saldırı sınıfı bulunmaktadır [29]:

- Hizmet Reddi (Denial of Service - DoS): Bu saldırı çeşidi, saldırganın bazı bilgi işlem veya bellek kaynaklarını meşru istekleri yerine getiremeyecek kadar meşgul, dolu hale getirdiği veya meşru kullanıcıların bir makineye erişimini engellediği bir saldırıdır. Bu, veri kümesindeki en yaygın saldırıdır. IDS, sistemin idare edemeyeceği kadar anormal miktarda trafikle dolup taşar ve kendini korumak için kapanır.
- Probe: Bu saldırıdaki amaç, bir hırsız gibi davranmak ve müşterilerle ilgili kişisel bilgiler veya bankacılık bilgileri gibi önemli bilgileri çalmaktır.
- User to Root (U2R): U2R, normal bir kullanıcı hesabıyla başlayan ve bir süper kullanıcı (kök) olarak sisteme veya ağa erişmeye çalışan bir saldırıdır. Saldırgan, kök ayrıcalıkları/erişim elde etmek için bir sistemdeki güvenlik açıklarından yararlanmaya çalışır.
- Remote to Local (R2L): Ağ üzerinden bir makineye paket gönderme yeteneğine sahip olan ancak o makine üzerinde bir hesabı olmayan saldırganın, makinenin kullanıcısı olarak yerel erişim elde etmek için bazı güvenlik açıklarından yararlanması durumunda ortaya çıkar.

Yukarıdaki saldırı çeşitleri veri kümesinde mevcut olmasına rağmen, saldırıların dağılımı oldukça çarpıktır. Veri kümesinde bulunan kayıtların yarısından fazlası normal trafiktir ve U2R, R2L dağılımları son derece düşüktür. Ancak, en yaygın saldırı çeşidinin DoS olduğu düşünüldüğünde, günümüz internet trafiği saldırılarının dağılımının yaklaşık bir temsilidir.

UNSW-NB15 veri kümesi, çok sayıda kullanıcı kaydı biçiminde gerçek modern normal kanıtlar ve sentetik çağdaş saldırı örneklerinin bir karışımından oluşur; normal gözlemleri ve dokuz saldırı sınıfı içermektedir [30]:

- Fuzzers: Rastgele oluşturulmuş verileri besleyerek bir programın veya ağın askıya alınmasına neden olmaya çalışan saldırıdır.
- Analysis: Farklı port tarama, spam ve html dosya penetrasyon saldırılarını içerir.
- Backdoors: Bir bilgisayara veya verilerine erişmek için bir sistem güvenlik mekanizmasının gizlice atlandığı bir teknik.
- DoS: Genellikle internete bağlı bir ana bilgisayarın hizmetlerini geçici olarak kesintiye uğratarak veya askıya alarak bir sunucuyu veya ağ kaynağını kullanıcılar tarafından kullanılamaz hale getirmeye yönelik kötü niyetli bir girişim.
- Exploits: Saldırganın, bir işletim sistemi veya bir yazılım parçası içindeki bir güvenlik sorununu bildiği ve bu güvenlik açığından yararlandığı bir saldırı çeşidi.
- Generic: Bu, esasen gizli anahtarlar üzerinde yapılan bir çakışma saldırısıdır. Bir teknik, blok şifrenin yapısı dikkate alınmadan tüm blok şifrelerine (belirli bir blok ve anahtar boyutu ile) karşı çalıştırılır.
- Reconnaissance: Bilgi toplayan saldırıların benzetimini yapabilen tüm saldırılar.
- Shellcode: Yazılım güvenlik açığından yararlanmada yük olarak kullanılan küçük bir kod parçası.
- Worms: Saldırganın, diğer bilgisayarlara yayılmak için kendini kopyaladığı bir saldırı çeşididir. Genellikle, saldırgan kendisini yaymak için bir bilgisayar ağını kullanır ve bu ağa erişmek için hedef bilgisayardaki güvenlik hatalarına güvenir.

UNSW-NB15 veri kümesi gerçek dünya trafik ağ paketlerinden elde edilen kayıt başına 49 özellik içerir, özelliklerin 47 tanesi trafik girdisi ve son ikisi etiketler (saldırı türü) ve bir kaydı normal veya anormal olarak sınıflandıracak (0 veya 1) şekildedir. Bu veri kümesinde de NSL-KDD de olduğu gibi bir dağılım problemi bulunmaktadır. Kayıtların yarısından çoğu normal ağ davranışlarıdır ve bazı durumlarla ilgili veriler neredeyse yok denecek kadar azdır. NSL-KDD ve UNSW-NB15 veri kümeleri hakkında daha detaylı bilgiler için [29] ve [30] çalışmalarına bakınız. Bunlara ek olarak [28] nolu çalışmada yazarlar tarafından kullanılan veri kümesi hakkında bir bilgi paylaşılmamıştır.

Şekil 1'de görüldüğü üzere incelenen çalışmalardan sekiz tanesinde oluşturulan veri kümeleri kullanılmıştır. [17] çalışmasında veri olarak genel konum bilgisi, işletim sistemi bilgisi, kaynak ve hedef tarafın IP adresleriyle birlikte ağ paketleri arasındaki gecikmeler kullanılmıştır. [18] çalışmasında NSL-KDD veri kümesi zararlı davranışları temsil ederken Network Simulator-3 (NS-3) aracılığıyla oluşturulan trafik iyi huylu davranışları temsil etmektedir ve birlikte kullanılmışlardır. NS-3 ile toplanan veriler ağdaki kullanılan protokol, başlangıç zamanı, kaynak baytları ve hedef baytları gibi verileri içermektedir. [19] çalışmasında iş parçacıklarının CPU kullanımları ve kaynak erişimlerinin izlenmesi ile tespit işlemlerinin gerçekleşebileceğinden bahsedilmektedir. Veri olarak iş parçacıklarının anlık yaptıkları CPU paylaşımı, iş parçacıklarının kümülatif CPU paylaşımları, kritik kaynak erişim türü ve kritik kaynak erişim süresi kullanılmaktadır. Bu verilerin ek olarak n-gram algoritması ile geliştirilmiş versiyonları da kullanılmıştır. Düzenli bir ağ trafiğinin oluşturulabilmesi için bir bulut uygulaması aracılığıyla bazı senaryolar ile saldırıların benzetimi yapılmıştır. [20] çalışmasında NASA web günlükleri temel alınarak zaman serisi şeklinde veriler oluşturulmuştur ve saldırı simülasyonları gerçekleştirilmiştir. [23] çalışmasında çalıştırılabilir her dosya için günlük izleme işlevi kullanılarak Elog adında bir veri kümesi oluşturulmuştur. Bu veri kümesi bulut bulutu şeklinde bir ortamda sanal makine içi ve sanal makine dışı sistem çağrıları kullanılarak oluşturulmaktadır. [24] çalışmasında veri olarak; saniyede birim zaman başına kaynak IP sayısı, gelen akış paketlerinin standart sapması, akış varlıklarının hızı, yineleme oranı ve toplam akış sayısı kullanılmıştır. [25] çalışmasında güvenli olduğu bilinen IP adresleri veri olarak kullanılmıştır. [27] çalışmasında ise kripto madencilik uygulamalarının belirli sistem çağrılarını kullandığı düşüncesi ile Linux sistem çağrıları veri olarak kullanılmıştır.

C. AS3: ÇALIŞMALARIN ODAKLANDIKLARI BİLGİ GÜVENLİĞİ KAVRAMLARI

Ulusal Güvenlik Sistemleri Komitesi (Committee on National Security Systems - CNSS), bilgi güvenliğini, bilgilerin ve bu bilgileri kullanan, depolayan, ileten sistemler, donanımlar dahil olmak üzere kritik unsurlarının korunması olarak tanımlamıştır [31]. CNSS bilgi güvenliği modeli, bilgisayar güvenliği endüstrisi tarafından geliştirilen CIA üçgeni (confidentiality (gizlilik) - integrity (bütünlük) - availability (erişilebilirlik)) özelliklerine dayanmaktadır. CIA üçgeni, ana çerçevenin geliştirilmesinden bu yana bilgisayar güvenliği için endüstri standardı olmuştur [32].

Gizlilik, bulut bilişim hizmeti kullanıcılarının bilgilerinin yetkisiz olarak ifşa edilmesini engellemeyi gerektirmektedir. Bulut servis sağlayıcıları, gizliliği garanti etmek için kullanıcılardan ücret talep etmektedir. Bulut bilişim sistemlerini kullanmanın amaçlarından bir tanesi, veri gizliliğini sağlayarak bulut üzerindeki çeşitli kaynakları kullanmaktır. Bilginin (veri) yetkisiz kişilere veya sistemlere maruz kalmasına karşı koruma gerçekleştirildiği zaman gizlilik sağlanmaktadır [32]. Gizlilik, yalnızca bilgiye erişim haklarına ve ayrıcalıklarına sahip olanların bunu yapabilmesini sağlar. Yetkisiz kişiler veya sistemler bilgileri görüntüleyebildiğinde, gizlilik ihlal edilmiş olur. Bu açıdan bakıldığında kimlik denetimi (erişim kontrolü) bulut bilişimde gizliliğin önemli bir parçasıdır.

Bulut bilişim sistemlerini kullanmanın amaçlarından bir tanesi, çeşitli kaynakları kullanmaktır. Bulut bilişim sistemlerinin tüm verileri desteklemesinin ve birçok kullanıcının aynı bulutlara bağlı kalmasının nedeni budur. Bilgi, eksiksiz ve bozulmamış olduğunda bütünlüğe sahiptir. Bilgi, bozulduğunda, hasara, tahribata maruz kaldığında veya özgün durumu herhangi bir şekilde değiştiğinde bilginin bütünlüğü tehdit altına girmektedir [32]. Kullanıcılar mevcut verileri değiştirme, güncelleme veya buluta yeni veriler eklemek isteyebilirler. Bu nedenle, veri bütünlüğünü sağlamak için kimlik denetimi işlemi yapılması gerekmektedir.

Erişilebilirlik, yetkili kullanıcıların (kişiler veya bilgisayar sistemleri) bilgiye müdahale veya engelleme olmaksızın erişmesini ve gerekli formatta almasını sağlar. Erişilebilirlik, tüketicinin sistemi beklendiği gibi kullanma yeteneğidir [31]. Bir bulut bilişim sisteminin önemli avantajlarından biri veri kullanılabilirliğidir. Erişilebilirlik bulut sistemlerinin ana bir parçası olduğundan, ortamın artan kullanımı, dışarıdan veya içeriden gerçekleşecek saldırılar erişilebilirlik problemi yaşanması olasılığını artıracak ve bu nedenle bulut sisteminin performansını düşürecektir. Bulut uygulamaları, kimlik denetimi yoluyla erişilebilirliği artırır.

Gizlilik ve bütünlük açısından bakıldığında kimlik denetimi bu özelliklerin sağlanması açısından son derecede önem taşımaktadır. Erişim kontrolü, yalnızca kullanıcı kimliğine göre belirlenebilir, ancak çoğu durumda, kullanıcı hakkında rolleri veya unvanları gibi ek nitelikler gerektirir [33]. Bulut için uygun erişim kontrol modellerinin uygulanması, mevcut erişim kontrol modelleri bulut sistemlerinin gereksinimlerini karşılamak için özel olarak tasarlanmadığından kritik olarak değerlendirilen alanlardan biridir [33]. Bu yüzden bu çalışmada üç ana bilgi güvenliği kavramına ek olarak dördüncü bir kavram kimlik denetimi dahil edilerek inceleme genişletilmiştir.

İncelenen çalışmalar genellikle IDS kavramı üzerinde durduklarından, bulut üzerindeki anormal davranış veya saldırıları tespit etmek üzerine yapılan çalışmalardır. İlk bakışta bu sistemlerin kendilerinin özellikle gizlilik ve bütünlük düşünüldüğünde bilgi güvenliği kavramlarını sağlamadıkları düşünülebilir. Ancak herhangi bir saldırıyı durdurmanın en iyi yollarından birisinin bu saldırıyı tespit etmek olduğu düşünüldüğünde bu çalışmaların ilgili bilgi güvenliği kavramlarını kısmen de olsa sağladıkları görülmektedir. Çalışmaların kullandıkları hazır veri kümeleri (örneğin NSL-KDD ve UNSW) ilgilenilen dört bilgi güvenliği kavramı için tespit edilecek saldırı türlerini barındırmaktadır. Bu yüzden bu kavramları karşılayan saldırıları tespit eden çalışmaların bilgi güvenliği kavramlarını sağladıkları düşünülmüştür. Ek olarak incelenen sistemlerde genel olarak görülen yapı, tespit mekanizması yanında bu saldırıların azaltılmasını veya sanal makine seçimi, işlem (process) durdurma ve belirlenen IP üzerinden gelen trafiği durdurma gibi yöntemlerle engellenmesini de içermektedir. Ancak, bulut bilişimde özellikle erişim kontrolü bilgi güvenliği kavramlarını sağlamak açısından çok

büyük önem taşımaktadır. Kimlik denetimi mekanizması kullanmayan çalışmaların veya kimlik denetimi ile ilgili bir tespit yapmayan çalışmaların gizlilik ve bütünlük kavramlarını sağlayamayacağı durumlar ortaya çıkabilecektir. Bu bilgiler doğrultusunda incelenen çalışmaların bilgi güvenliği kavramları açısından incelemeleri Tablo 7'de verilmiştir.

Tablo 7. *Çalışmaların sağladıkları bilgi güvenliği kavramları açısından değerlendirmeleri*

Çalışma	Gizlilik	Bütünlük	Erişilebilirlik	Kimlik Denetimi (Erişim Kontrolü)
[14]	✓	✓	✓	✓
[15]	✓	✓	✓	✓
[16]	✓	✓	✓	
[17]	✓	✓	✓	
[18]	✓	✓	✓	✓
[19]		✓	✓	✓
[20]			✓	
[21]	✓	✓	✓	✓
[22]	✓	✓	✓	✓
[23]			✓	✓
[24]			✓	
[25]	✓	✓		
[26]	✓	✓	✓	✓
[27]			✓	
[28]		✓		✓

D. DEĞERLENDİRME

Bu bölümde, araştırma soruları yanıtlandıktan sonra oluşturulan analiz ve bazı makalelerin gelecek çalışmalarının analizi sunulmaktadır. Bu açıdan paylaşılan Tablo 8 incelenen çalışmaların genel bir özetini elde ettikleri en iyi sonuçlarla birlikte sunmaktadır. Devamında incelenen çalışmalar IDS, atak tespiti, zararlı yazılım tespiti, DDoS ve anormallik tespiti şeklinde beş güvenlik yönü açısından sınıflandırılmıştır. Bu sınıflandırma ile ilgili bilgiler Tablo 9'da verilmiştir.

Tablo 9. *Çalışmaların bazı güvenlik yönleri açısından sınıflandırılmaları ve güvenlik yönü açısından en iyi sonuçları elde eden çalışmalar*

Güvenlik Yönü	Çalışma Referansları	En Yüksek Başarı Elde Eden Çalışmalar
İzinsiz Giriş Tespiti	[14], [18], [22], [26], [27]	[18] ve [22] çalışmaları %99 üzerinde doğruluk sonucu elde etmiştir
Atak Tespiti	[15], [16], [20], [21], [25], [26]	[21] çalışması bazı veri kümeleri için %99 üzerinde doğruluk sonucu elde etmiştir. [20] çalışması ise %99 doğruluk sonucu elde etmiştir.
Zararlı Yazılım Tespiti	[19], [23]	[19] çalışması %98.9 F1-Puanı sonucu elde etmiştir
DDoS	[17], [24]	[24] çalışması %98.9 F1-Puanı sonucu elde etmiştir
Anormallik Tespiti	[20], [27]	[20] çalışması %99 doğruluk, kesinlik ve duyarlılık sonucu elde etmiştir.

Tablo 8. *İncelenen çalışmaların elde ettikleri en iyi sonuçlar. Test ve eğitim veri sayısı paylaşılmayan çalışmalarda sadece veri sayıları paylaşılmıştır. (Ç: Çalışma, VK: Veri Kümesi, VM: Veri Miktarı, EV: Eğitim Verisi Miktarı, TVM: Test Verisi Miktarı, Z: Zararlı, N: Normal, OV: Oluşturulmuş Veri, D: Doğruluk, TP: Gerçek Pozitif Oranı, FN: Yanlış Negatif Oranı, TN: Gerçek Negatif Oranı, FP: Yanlış Pozitif Oranı, K: Kesinlik, Du: Duyarlılık, AO: Algılama Oranı, YAO: Yanlış Alarm Oranı, MKK: Mathew Korelasyon Katsayısı, NTO: Negatif Tahmin Oranı, PTO: Pozitif Tahmin Oranı, CK: Cohen's Kappa, TSH: Test Sınıflandırma Hatası, YSD: Yanlış Sınıflandırma Doğruluğu)*

				Performans Kriterleri																
Ç	**VK**	**VM**		**D** *(%)*	**TP** *(%)*	**FN** *(%)*	**TN** *(%)*	**FP** *(%)*	**K** *(%)*	**Du** *(%)*	**F1** *(%)*	**ROC-AUC** *(%)*	**AO** *(%)*	**YAO** *(%)*	**MKK** *(%)*	**NTO** *(%)*	**PTO** *(%)*	**CK** *(%)*	**TSH** *(%)*	**YSD** *(%)*
		EVM	**TVM**																	
[14]	KDD Cup '99	396,743 Z 97,278 N	231,647 Z 60,593 N	92.5	-	-	-	-	-	-	-	-	-	-	-	-	-	7.2	-
[15]	UNSW	120,000 Z 60,000 N	43,000 Z 40,000 N	96.45	98.6	1.4	98.3	≈0	-	-	-	-	-	-	-	-	-	-	-
	ISOT	56,000 Z 170,000 N		98.6	63.9	36.1	99.87	≈0	-	-	-	-	-	-	-	-	-	-	-
[16]	UNSW (Çok sınıflı)	4500 Z	900 Z	-	-	-	-	3.6	96.4		97.5								
	UNSW (Tek sınıflı)	2000 Z 2000 N	1000 Z 1000 N	-	98	-	-	2											
[18]	NSL-KDD (sağdaki veriler NS – 3 (Normal veriler, sayısı verilmemiştir)	58,630 Z 67,343 N	12,833 Z 9.771 N	99.43	-	1.53	-	0.96	99.92	-	-	-	-	-	-	-	-	-	-	-
[19]	OV	1000 Z 99,000 N		-	98.99	0.016	0.989	0.006	99.4	98.4	98.9	-	-	-	-	-	-	-	-	2.2
[20]	NASA Web Sunucu Günlükleri	İlk 820 saat verileri	821 – 1369 saatleri arası verileri	99	-	-	-	-	99	99	-	-	-	-	-	-	-	-	-	-
[21]	NSL-KDD	148,517 örnek		96.06	-	-	-	-	≈95	≈93	-	95	-	0.076	-	-	-	-	-	-
	Honeypot	250,092 örnek		99.93	-	-	-	-	≈95	≈94	-	98	-	0.001	-	-	-	-	-	-
	Kyoto	257,673 örnek		99.93	-	-	-	-	≈94	≈2	-	96	-	0.001	-	-	-	-	-	-
[22]	CICIDS-2017	20,000 Z 20,000 N	20,000 Z 20,000 N	99.93	-	0.08	99.95	0.05	99.95	-	99	99.93	99.92	-	-	-	-	-	-	-
	NSL-KDD	70,373 Z 80,792 N	12,833 Z 9,711 N	99.86	-	0.17	99.91	0.09	99.93	-	99	99.87	99.83	-	-	-	-	-	-	-
	CIDDS-001	39,392 Z 571,979 N	26,260 Z 407,579 N	99.92	-	0.14	99.92	0.08	99.87	-	99.89	99.89	99.86	-	-	-	-	-	-	-
[23]	UNM (Sonuçlar her sınıf için ayrı verildiğinden aralık paylaşılmıştır)	2,612 Z 32,104 N		81.25 - 99.92	86.7 - 99	0 - 4	73.6 - 100	0.7 - 16.8	-	-	-	-	-	-	-	-	-	-	-	-
	Elog (Sonuçlar her sınıf için ayrı verildiğinden aralık paylaşılmıştır)	-		95.43 - 97.81	89.1 - 100	0 - 10.9	95.8 - 98.3	0.7 - 4.2	-	-	-	-	-	-	-	-	-	-	-	-
[24]	OV	-	4000 Z 2000 N	-	98.48	-	98.52	-	98.33	98.48	98.9	-	-	-	-	-	-	-	-	-
[25]	OV	100 ile 500 arasında değişen paket sayılarında veriler kullanılmıştır		96.25	96.12	-	94.32	-	95.65	93.45	92.11	-	-	-	89.9	94.47	85.78	-	-	-
[26]	NSL-KDD	262,283 Z 812,814 N	29,378 Z 47,911 N	98	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	UNSW	321,283 Z 2,218,761 N		91	-	-	-	4	-	-	-	94	-	-	-	-	-	-	-	-
[27]	OV	176,811 Z 135,686 N		97.1	-	-	-	-	97	97	97	97.01	-	-	-	-	-	94.03	-	-

Tablo 9'da paylaşılan çalışmalardan görüldüğü üzere bu beş güvenlik yönünden bakıldığında, modellerin %99 üzeri başarılara ulaştıkları ve bu güvenlik sorunlarının neredeyse çözüldüğü düşünülebilir. Ancak, Tablo 8'de elde edilen sonuçlar kullanılan veriye oldukça bağımlıdır. Örneğin, [21] çalışmasını ele alırsak üç farklı veri kümesi için iki veri kümesinde nispeten yakın sonuçlar alınmış olsa da, diğer veri kümesinde (NSL-KDD) daha düşük başarı elde edilmiştir. Burada kullanılan verinin büyüklüğü de önemli olmaktadır. İncelenen çalışmalarda görüldüğü üzere genellikle dengesiz veri kümelerinin (sınıfların eşit sayıda örnek içermediği durum) kullanıldığı görülmektedir. Örneğin AS2'de bahsedilen veri kümeleri düşünüldüğünde farklı atak tipleri bulunmaktadır ancak her atak tipi için eşit sayıda veri bulunmamaktadır. Bu durum yeterince güvenilir performans metrikleri ile değerlendirme yapılmadığında, sınıflandırıcının performansı konusunda şüphelere sebep olabilir. Dengesiz veri kümeleri sıklıkla kullanılmaktadır ve bu tarz durumlarda değerlendirme yapmak için en çok tercih edilen yöntem F1-Puanı incelemesidir. Ancak incelemelerde görülmektedir ki sadece altı çalışma bu metriği incelemesine dahil etmiştir. Dengesiz veri kümelerinin kullanımı yapıldığında güvenilir performans kriterlerinin kullanımı önemli ve kesinlikle incelenmesi gereken bir durumdur.

Çalışmalarda bahsedilen açık durumlar incelendiğinde en çok öne çıkanlar, derin öğrenme modellerinin kullanılmak istenmesi ve farklı veriler üzerinde yapılmak istenen deneylerdir. İncelenen çalışmalarda görüldüğü üzere, sürekli değişen saldırı yapılarından dolayı geleneksel yöntemlerin artık geleneksel güvenlik sistemleri için yetersiz kalması ve artan veri sayısı ile birlikte derin öğrenme modellerinin kullanımının ön plana çıktığı görülmektedir. Derin öğrenme modellerinin büyük sayıdaki veriler ile geleneksel yöntemlere kıyasla çok daha verimli çalıştığı bilinen bir durumdur. Diğer bir öne çıkan durum ise gerçek dünya verileri ile oluşturulan bu modeller üzerinde deneylerin yapılmasıdır. Çoğu çalışmada hazır verilerin kullanıldığı, bu verilerin büyük çoğunluğunun güncel olmaktan uzak olduğu ve her saldırı türü için yeterli örnek olmadığı düşünüldüğünde, önerilen sistemlerin gerçek performanslarının anlaşılabilmesi açısından gerçek dünya verilerinin kullanımı büyük önem arz etmektedir. Aksi takdirde önerilen sistemler sadece veri kümelerinde bulunan belirli durumlara karşı savunma sağlayabilecek olsa bile gerçekleşebilecek farklı bir saldırıyı tespit etmekte veya durdurmakta başarısız olacaklardır.

IV. SONUÇ

Bu çalışmada bulut bilişim güvenliği alanında kullanılan ML yöntemlerini incelemek için bir sistematik literatür taraması çalışması yapılmıştır. İlgili çalışmalar seçildikten sonra belirlenen üç tane araştırma sorusunun cevaplanması amaçlanmıştır. Bu üç soru; kullanılan ML yöntemleri ve performans kriterleri, kullanılan veri kümeleri ve çalışmaların odaklandıkları bilgi güvenliği kavramlarıdır. Seçme ve eleme kriterleri doğrultusunda toplamda 15 tane çalışma bu incelemeye dahil edilmiştir. İnceleme sonuçları aşağıda özetlenmiştir:

- AS1 sonucu olarak toplamda 23 tane ML yönteminin ve 17 tane performans ölçütünün kullanıldığı görülmüştür. En çok kullanılan ML yöntemleri; altı tane ile karar ağaçları, beş tane ile sinir ağları (yapay sinir ağları ve derin sinir ağları) ve dört tane ile destek vektör makinesi yöntemleridir. Performans ölçütü olarak da 11 çalışmada doğruluk ve dokuz çalışmada kesinlik ölçütlerinin kullanıldığı görülmüştür.
- AS2 sonucu olarak toplamda 11 tane hazır veri kümesinin kullanıldığı görülmüştür. Bu veri kümeleri içerisinden dört çalışma ile NSL-KDD ve üç çalışma ile UNSW veri kümeleri en çok kullanılanlar olarak öne çıkmaktadır. İncelenen çalışmalardan sekiz tanesinde ise yazarların kendi oluşturdukları veri kümelerini kullandıkları görülmüştür. Çalışmalarda genellikle dengesiz dağılımlara sahip veri kümeleri kullanılmıştır. Önerilen modellerin performanslarını daha iyi inceleyebilmek için özellikle dengesiz dağılımlar üzerinde sonuçları daha düzgün bir şekilde ifade eden F1-Puanı gibi daha farklı performans kriterinin incelenmesi gerekliliği görülmektedir. Çalışmaların gerçek dünya verilerini daha çok kullanması veya bu verileri çok daha iyi ifade edebilecek simülasyonlar yapmalarının gerektiği anlaşılmaktadır.

- AS3 sonucu olarak çalışmaların kullandıkları veri kümelerinin yapılarından dolayı gizlilik, bütünlük, erişilebilirlik ve kimlik denetimi kavramlarını genel olarak sağladıkları görülmüştür. Ancak bu kısımda bilgi güvenliği kavramlarını sağladığı söylenen çalışmalar, çoğunlukla bu kavramlar ile ilgili saldırıların tespitlerini içermektedir. Bulut bilişim açısından düşünüldüğünde kimlik denetimi gizlilik ve bütünlük kavramlarını karşılayabilmek açısından büyük önem arz etmektedir ve 15 çalışmadan dokuz tanesinin bu özelliği sağladığı görülmüştür.

V. KAYNAKLAR

[1] P. Mell and T. Grance. (2011, Sep). *The NIST definition of cloud computing* [Online]. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

[2] A. B. Nassif, M. A. Talib, Q. Nasir, H. Albadani, and F. M. Dakalbab, "Machine learning for cloud security: A systematic review," *IEEE Access*, vol. 9, pp. 20717–20735, 2021.

[3] L. Alhenaki, A. Alwatban, B. Alahmri, and N. Alarifi, "Security in cloud computing: A survey," *International Journal of Computer Science and Information Security*, vol. 17, pp. 67–90, 2019.

[4] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, "A survey on security issues and solutions at different layers of cloud computing," *J Supercomput*, vol. 63, no. 2, pp. 561–592, 2013.

[5] M. De Donno, A. Giaretta, N. Dragoni, A. Bucchiarone, and M. Mazzara, "Cyber-storms come from clouds: Security of cloud computing in the IoT era," *Future Internet*, vol. 11, no. 6, Jun. 2019, Art. no. 127.

[6] S. Singh, Y.-S. Jeong, and J. H. Park, "A survey on cloud computing security: Issues, threats, and solutions," *Journal of Network and Computer Applications*, vol. 75, pp. 200–222, 2016.

[7] U. A. Butt, M. Mehmood, S. B. H. Shah, R. Amin, M. W. Shaukat, S. M. Raza, D. Y. Suh, and M. J. Piran, "A review of machine learning algorithms for cloud computing security," *Electronics*, vol. 9, no. 9, Sep. 2020, Art. no. 1379.

[8] Md. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi, "A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing," *Future Generation Computer Systems*, vol. 28, no. 6, pp. 833–851, 2012.

[9] A. Patel, M. Taghavi, K. Bakhtiyari, and J. Celestino Júnior, "An intrusion detection and prevention system in cloud computing: A systematic review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 25–41, 2013.

[10] S. G. Kene and D. P. Theng, "A review on intrusion detection techniques for cloud computing and security challenges," in *2nd International Conference on Electronics and Communication Systems (ICECS)*, 2015, pp. 227–232.

[11] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in cloud," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 42–57, 2013.

[12] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," *Cluster Comput*, vol. 22, no. 1, pp. 949–961, 2019.

[13] S. Shamshirband, M. Fathi, A. T. Chronopoulos, A. Montieri, F. Palumbo, and A. Pescapè, "Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, and open research issues," *Journal of Information Security and Applications*, vol. 55, Dec. 2020, Art. no. 102582

[14] A. Abusitta, M. Bellaiche, M. Dagenais, and T. Halabi, "A deep learning approach for proactive multi-cloud cooperative intrusion detection system," *Future Generation Computer Systems*, vol. 98, pp. 308–318, 2019.

[15] E. K. Subramanian and L. Tamilselvan, "A focus on future cloud: machine learning-based cloud security," *SOCA*, vol. 13, no. 3, pp. 237–249, 2019.

[16] M. Rabbani, Y. L. Wang, R. Khoshkangini, H. Jelodar, R. Zhao, and P. Hu, "A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing," *Journal of Network and Computer Applications*, vol. 151, Feb. 2020, Art. no. 102507.

[17] S. Dey, Q. Ye, and S. Sampalli, "A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks," *Information Fusion*, vol. 49, pp. 205–215, 2019.

[18] M. Aloqaily, S. Otoum, I. A. Ridhawi, and Y. Jararweh, "An intrusion detection system for connected vehicles in smart cities," *Ad Hoc Networks*, vol. 90, Jul. 2019, Art. no. 101842.

[19] M. T. Sandıkkaya, Y. Yaslan, and C. D. Özdemir, "DeMETER in clouds: detection of malicious external thread execution in runtime with machine learning in PaaS clouds," *Cluster Comput*, vol. 23, no. 4, pp. 2565–2578, 2020.

[20] A. Agarwal, A. Prasad, R. Rustogi, and S. Mishra, "Detection and mitigation of fraudulent resource consumption attacks in cloud using deep learning approach," *Journal of Information Security and Applications*, vol. 56, Feb. 2021, Art. no. 102672.

[21] S. Krishnaveni, S. Sivamohan, S. S. Sridhar, and S. Prabakaran, "Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing," *Cluster Comput*, vol. 24, no. 3, pp. 1761 – 1779, 2021.

[22] Z. Chiba, N. Abghour, K. Moussaid, A. El omri, and M. Rida, "Intelligent approach to build a deep neural network based IDS for cloud environment using combination of machine learning algorithms," *Computers & Security*, vol. 86, pp. 291–317, 2019.

[23] P. Mishra, I. Verma, and S. Gupta, "KVMInspector: KVM based introspection approach to detect malware in cloud environment," *Journal of Information Security and Applications*, vol. 51, Apr. 2020, Art. no. 102460.

[24] S. Mishra, S. Kumar Sharma, and M. A. Alowaidi, "Multilayer self-defense system to protect enterprise cloud," *Computers, Materials & Continua*, vol. 66, no. 1, pp. 71–85, 2020.

[25] D. R. Rani and G. Geethakumari, "Secure data transmission and detection of anti-forensic attacks in cloud environment using MECC and DLMNN," *Computer Communications*, vol. 150, pp. 799–810, 2020.

[26] Z. Chkirbene, A. Erbad, R. Hamila, A. Mohamed, M. Guizani, and M. Hamdi, "TIDCS: A dynamic intrusion detection and classification system based feature selection," *IEEE Access*, vol. 8, pp. 95864–95877, 2020.

[27] R. R. Karn, P. Kudva, H. Huang, S. Suneja, and I. M. Elfadel, "Cryptomining detection in container clouds using system calls and explainable machine learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 3, pp. 674–691, 2021.

[28] D. Praveena and P. Rangarajan, "A machine learning application for reducing the security risks in hybrid cloud networks," *Multimed Tools Appl*, vol. 79, no. 7–8, pp. 5161–5173, 2020.

[29] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009, pp. 1–6.

[30] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," *Military Communications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6.

[31] J. McConnell, "National training standard for information systems security (INFOSEC) professionals," National Security Agency/Central Security Service, Fort George, G Meade Md, Jun. 20, 1994.

[32] M. E. Whitman and H. J. Mattord, *Principles of Information Security*, 4th ed., Boston, MA, USA: Cengage Learning, 2012.

[33] S. Y. Lim, M. L. Mat Kiah and T. F. Ang, "Security issues and future challenges of cloud service authentication," *APH*, vol. 14, no. 2, pp. 69-89, 2017.