All Sciences Proceedings

5th International Conference on Applied Engineering and Natural Sciences

July 10-12, 2023 : Konya, Turkey

All Sciences Proceedings
http://as-proceeding.com/

https://www.icaens.com/

Web Uygulama Zafiyetlerinin Keşfinde Açıklanabilir Yapay Zekânın Yeri

Erhan BAŞ[1], Ahmet Ali Süzen[2*]

[1]*Bilgisayar Mühendisliği Bölümü /Lisansüstü Eğitim Enstitüsü, Isparta Uygulamalı Bilimler Üniversitesi, Türkiye*
[2*]*Bilgisayar Mühendisliği Bölümü /Teknoloji Fakültesi, Isparta Uygulamalı Bilimler Üniversitesi, Türkiye*

[*]*(ahmetsuzen@isparta.edu.tr)*

Özet **–** Günümüzün dijital çağında, siber güvenlik, toplumlar ve kuruluşlar için önemli bir endişe kaynağıdır. Yapay Zeka ve Makine Öğrenmesi teknolojileri, karmaşık ve sürekli değişen tehdit ortamında güvenliği sağlama çabalarını hızlandırırken, onların karar verme süreçlerini anlama yeteneği, bu teknolojilerin genel kabulünü ve etkinliğini sınırlar. Bu anlaşılmazlık, Açıklanabilir Yapay Zeka ihtiyacını doğurur. Açıklanabilir Yapay Zekâ, modelin kararlarının insanlar tarafından anlaşılabilir olmasını sağlayarak, karmaşık veri setleri üzerinde derinlemesine analiz yapma yeteneğine sahiptir. Bu sayede, güvenlik uzmanları, Açıklanabilir Yapay Zekâ ile keşfedilen zafiyetlerin nedenini ve çözüm yollarını daha kolay anlayabilir. Bu çalışma, web uygulamalarının güvenlik zafiyetlerini tespit etmede açıklanabilir yapay zeka teknolojilerinin önemini ve nasıl bir katkı sağlayabileceğini ele alır. Ayrıca açıklanabilir yapay zekanın siber güvenlikteki uygulamaları ve önemini detaylıca inceleyen bu çalışma, geliştirme ve uygulama süreçlerinin etkinliğini artırmak için açıklanabilir yapay zekanın nasıl kullanılabileceğini tartışacaktır.

Anahtar Kelimeler – Açıklanabilir Yapay Zeka, Siber Saldırı, Web Uygulama Zafiyetleri, Veri Sızıntısı

I. GİRİŞ

Bilişim teknolojileri, modern dünyada hayatımızın her alanını etkileyen kritik bir rol oynamaktadır. İletişim, eğitim, eğlence, iş ve daha birçok alanda bu teknolojilerin varlığı göz ardı edilemez. Gelişen teknoloji ve artan dijitalleşme, modern dünyayı şekillendirmekte ve hemen her alanda değişiklikler getirmektedir. Bu hızlı değişimin sonucu olarak siber saldırılar, hızla artan ve yaygınlaşan bir tehdit haline gelmiştir. İster bireyler, isterse de büyük ölçekli organizasyonlar olsun, siber saldırılar herkes için ciddi bir sorundur.

Siber saldırılar, bir bilgisayar sistemi, ağı veya teknolojik altyapıya yönelik kasıtlı ve genellikle kötü niyetli hareketleri ifade eder. (Öztürk & Zaim, 2019). Bu saldırılar, genellikle verileri çalmak, sistemleri devre dışı bırakmak veya manipüle etmek, işlemleri durdurmak veya hizmetleri kesmek amacıyla yapılır. Bu saldırılar, hem bireysel kullanıcıları hem de devlet kurumları, hastaneler, bankalar ve büyük şirketler gibi büyük kuruluşları hedef alabilir.

Siber saldırılar, zararlı yazılımların kullanılması, kimlik hırsızlığı, DDoS saldırıları, phishing ve sosyal mühendislik teknikleri gibi bir dizi yöntemle gerçekleştirilebilir [11]. İnternet ve dijital teknolojilerin yaygınlaşması, bilgiye erişimi ve iletişimi kolaylaştırmıştır. Ancak, bu süreç aynı zamanda daha geniş ve karmaşık bir siber alan yaratmıştır. Bu durum, güvenlik açıklarının ve savunmasızlıkların ortaya çıkması için bir alan yaratmıştır. Günümüzde, hemen hemen her işlem, veri alışverişi veya iletişim dijital kanallar aracılığıyla gerçekleştirilir. Bu, bireysel düzeyde olduğu gibi kurumsal ve devlet düzeyinde de geçerlidir. Bu durum, siber saldırıların etkisinin ve etki alanının genişlemesine yol açmıştır. COVID-19 pandemisi, dijitalleşmenin hızını daha da artırmıştır. İş yerlerinden eğitime kadar birçok alan, sosyal mesafe kurallarını uygulayabilmek için dijital platformlara geçmiştir. Bu, siber saldırıların olası hedeflerinin sayısını daha da artırmıştır [15,18]

Veri sızıntısı, hassas veya korunan verinin yetkisiz bireyler, uygulamalar, işletmeler veya sistemlere kasıtlı veya kazara aktarılması durumudur [3]. Bu,

bir dizi farklı yöntemle gerçekleştirilebilir, örneğin bir veri tabanındaki güvenlik açıklarının kötüye kullanılması, hedefli bir siber saldırı veya bir çalışanın bilerek veya yanlışlıkla hassas verilere erişim sağlaması. Web uygulamalarında veri sızıntıları, özellikle çevrimiçi işletmeler ve kuruluşlar için ciddi bir sorundur. Bunun nedeni, bu uygulamaların genellikle kullanıcıların kişisel ve finansal bilgilerini içeren büyük veri setlerine sahip olmasıdır (Erdinç & Çetin, 2020). Örneğin, bir e-ticaret sitesi, kullanıcının adı, adresi, kredi kartı bilgileri ve alışveriş alışkanlıkları dahil olmak üzere çok sayıda hassas bilgiyi saklayabilir. Bu bilgiler, kötü niyetli bir aktör tarafından erişildiğinde, kullanıcılara ve işletmelere zarar verebilir (Erdinç & Çetin, 2020).

Web uygulamalarındaki veri sızıntıları, genellikle web uygulamalarının güvenlik zafiyetlerinden kaynaklanır. Bu zafiyetler, SQL Enjeksiyonu, Cross-Site Scripting (XSS) veya Cross-Site Request Forgery (CSRF) gibi bilinen bir dizi farklı saldırı tipini içerebilir [10]. Özellikle, SQL Enjeksiyonu saldırıları, bir saldırganın bir web uygulamasının veritabanına zararlı SQL komutları enjekte etmesine ve genellikle hassas kullanıcı verilerine erişmesine olanak sağlar [10]. Bunun yanında, XSS ve CSRF saldırıları, bir saldırganın bir kullanıcının web tarayıcısına zararlı kod enjekte etmesine ve genellikle kullanıcının oturum bilgilerini çalmasına olanak sağlar [10].

Yapay Zeka (AI), bilgisayarların ve bilgisayar destekli sistemlerin insanların karar verme, problem çözme ve öğrenme yeteneklerini taklit etmeyi amaçlayan bir bilgisayar bilimleri dalıdır [2]. İnsanlar gibi düşünebilen ve öğrenme yeteneğine sahip makineler yaratma çabası, yapay zekanın genel amacını oluşturur.

Makine öğrenmesi (Machine Learning), AI'nın bir alt kümesi olarak, makinelerin deneyimden öğrenmesini ve belirli bir görevi gerçekleştirmek için gerekli becerileri 'öğrenirken' veri setlerinden öğrenmelerini sağlar. Derin öğrenme, makine öğrenmesinin bir alt kümesi olarak kabul edilir ve genellikle yapay sinir ağları olarak bilinen modelleri kullanır. Bu modeller, insan beyninin işleyişine benzer şekilde çalışır ve çok karmaşık ve soyut özellikleri öğrenebilirler [4].

Siber saldırılar ve siber güvenlik söz konusu olduğunda, AI, ML ve derin öğrenme bir dizi potansiyel uygulama sunar. ML algoritmaları, potansiyel tehditleri tespit etmek için ağ trafiğini analiz edebilir[3]. Derin öğrenme modelleri, karmaşık saldırı kalıplarını tespit etmek için kullanılabilir. Bu, güvenlik sistemlerinin sürekli değişen tehditlere karşı daha hızlı ve daha etkin bir şekilde yanıt vermesine yardımcı olabilir[5]. Ancak, AI'nin bu potansiyel faydalarına rağmen, bu teknolojilerin siber güvenlikteki kullanımı kendi zorluklarını da beraberinde getirir. Örneğin, kötü niyetli aktörler, AI teknolojilerini kullanarak daha karmaşık ve gelişmiş siber saldırıları gerçekleştirebilir[3].

Açıklanabilir Yapay Zeka (XAI), makine öğrenmesi ve derin öğrenme modellerinin karar süreçlerini şeffaf ve anlaşılabilir hale getirmeyi amaçlayan bir yapay zeka (AI) alt dalıdır [17]. Yapay zeka modelleri, çıktılarının nasıl ve neden üretildiğini ayrıntılı bir şekilde açıklamakta genellikle zorluk yaşar. Bunlar genellikle "kara kutu" olarak anılan modellerdir ve XAI, bu kara kutu modellerinin anlaşılırlığını artırmayı hedefler. Açıklanabilir Yapay Zeka'nın yeri ve önemi, özellikle karmaşık ve yüksek riskli alanlarda önem taşır. Örneğin, sağlık hizmetleri alanında bir AI modeli, bir hastanın teşhisini veya tedavi planını belirlemeye yardımcı olabilir. Ancak, bu kararların nasıl ve neden alındığının anlaşılabilir olması gereklidir [9]. Siber güvenlik alanında da XAI'nin önemi büyüktür. Bir AI modeli bir güvenlik tehdidini tespit edebilir, ancak tehdidin nasıl tespit edildiğini anlamak, güvenlik profesyonellerinin daha etkili yanıtlar geliştirmesine ve gelecekteki tehditleri daha iyi tahmin etmesine yardımcı olabilir [7]. XAI, yapay zeka teknolojilerine olan güveni artırmada da kritik bir rol oynar. Eğer insanlar bir AI modelinin nasıl çalıştığını ve belirli bir çıktıya nasıl ulaştığını anlarlarsa, bu modeli daha fazla güvenle kullanabilirler [8]. Sonuç olarak, XAI, AI'nın pratik uygulamalarını geliştirmede, hem de genel olarak AI teknolojilerine olan güveni artırmada önemli bir rol oynar. Bu çalışmada açıklanabilir yapay zekanın web uygulamalarının zafiyetlerinin keşfindeki yeri ve öneminden bahsedilmektedir.

II. WEB UYGULAMALARINDA TESPİT EDİLEN ZAFİYETLER

Bu bölümde, OWASP Top 10 zafiyet listesi, siber tehditlerin detaylı incelemesi için referans olarak kullanılmıştır. OWASP, Açık Web Uygulama Güvenliği Projesi anlamına gelir ve web uygulamalarındaki güvenlik açıklıklarının

giderilmesini amaçlayan özgür bir topluluktur. Topluluk, firmalardan ve web uygulama sızma testleri yapan kişilerden bilgi toplar, bu bilgileri analiz eder ve her yılın en riskli 10 güvenlik zafiyetini belirler ve bunu ücretsiz olarak sunar.

OWASP'ın zafiyet listesi, yazılım ve donanım zafiyetlerini listeleyen, sınıflandıran ve ölçen bir oluşum olan CWE (Common Weakness Enumeration) verilerini de içerir. CWE, yazılım güvenliğini artırmayı ve açıklıkları azaltmayı hedefler ve MITRE tarafından desteklenir. MITRE, ABD'deki çeşitli devlet kurumlarına destek sağlayan ve federal olarak finanse edilen araştırma ve geliştirme merkezlerini (FFRDC'ler) yöneten bir kurumdur.

Web uygulamalarında tespit edilen zafiyetlerin çoğu, OWASP (Open Web Application Security Project) tarafından belirlenen ve "OWASP Top 10" olarak adlandırılan listeye dahildir. 2021 itibarıyla OWASP Top 10'un güncel sürümü aşağıdaki gibidir:

A. Kırık Erişim Konrolü (Broken Access Control)

"Kırık Erişim Kontrolü" (Broken Access Control), kullanıcının bir web uygulamasının özelliklerine ve verilerine izin verilenden daha fazla erişim elde etmesine izin veren bir güvenlik zafiyetidir. OWASP (Open Web Application Security Project) bu durumu, 2021 yılında yayınladığı "OWASP Top 10" listesinde birinci sıraya koymuştur (OWASP, 2021). Erişim kontrol mekanizmalarının düzgün bir şekilde uygulanamaması sonucu oluşan bu zafiyet, yetkisiz kullanıcıların hassas verilere veya işlevlere erişimini, diğer kullanıcıların hesaplarına erişimini veya hatta kullanıcının rollerini değiştirmesini mümkün kılabilir. Bu durum genellikle, kullanıcı oturumları, kimlik doğrulama çerezleri, istek parametreleri veya API(Application Programming Interface) isteklerinin yeterince güvende olmaması sonucu oluşur. Kırık Erişim Kontrolünün önüne geçmek için güçlü erişim kontrol politikalarının uygulanması gerekmektedir. Özellikle rol tabanlı erişim kontrolü (RBAC) veya özellik tabanlı erişim kontrolü (ABAC) gibi sistemler, kullanıcı erişimlerini daha etkili bir şekilde yönetebilir. Ayrıca, uygulamanın yalnızca gerekli minimum yetkilere sahip olması da oldukça önemlidir. Bu, özellikle hassas bilgilerin korunması, işlevlere izinsiz erişimin engellenmesi ve kullanıcı rollerinin yanlışlıkla veya kötü niyetli bir şekilde değiştirilmesinin önlenmesi için gereklidir. Kullanıcıların ayrıca, oturumların ve çerezlerin güvende tutulması, istek parametrelerinin doğru bir şekilde doğrulanması ve API isteklerinin güvenlik kontrollerinden geçirilmesi de gerekmektedir. Bunun yanı sıra, kullanıcı kimlik doğrulaması ve yetkilendirme işlemlerinin düzenli olarak gözden geçirilmesi ve test edilmesi, olası güvenlik zafiyetlerinin erken tespit edilmesine yardımcı olabilir.

B. Kriptografik Hatalar (Cryptographic Failures)

Cryptographic Failures, veya kriptografik hatalar, hassas verilerin korunmasında önemli bir rol oynar. Yetersiz veya hatalı uygulamalar, saldırganların bu verilere erişimini sağlayabilir, dolayısıyla kriptografik işlemlerin doğru bir şekilde gerçekleştirilmesi hayati önem taşır. Kriptografik hatalar genellikle üç ana başlık altında incelenebilir: hassas verilerin şifrelenmemesi, zayıf veya güncel olmayan şifreleme algoritmalarının kullanılması ve şifreleme anahtarlarının yanlış yönetimi. Hassas verilerin şifrelenmemesi, bilgilerin açık metin halinde saklandığı veya aktarıldığı durumlardır. Bu, bir saldırganın verilere doğrudan erişimini sağlar ve önemli bir kriptografik hatadır. Bilgilerin düz metin halinde aktarılması, verilerin ele geçirilmesini veya değiştirilmesini kolaylaştırır. Zayıf veya güncel olmayan şifreleme algoritmalarının kullanılması da önemli bir zafiyettir. Bu durum, eski algoritmaların kullanılmasından veya algoritmaların yanlış bir şekilde uygulanmasından kaynaklanabilir. Örneğin, MD5 veya SHA1 gibi eski hash fonksiyonları güvenli olmayan seçeneklerdir, çünkü çeşitli saldırılara karşı zayıf oldukları bilinmektedir. Şifreleme anahtarlarının yanlış yönetimi de önemli bir kriptografik hatadır. Anahtarların güvende tutulması, rotasyonu ve gerektiğinde yenilenmesi önemlidir. Anahtarlar ayrıca rastgele bir şekilde oluşturulmalı ve hafızada byte dizileri olarak tutulmalıdır. Bu zafiyetlere karşı korunma yöntemleri, hassas verilerin detaylı bir şekilde belirlenmesi ve şifrelenmesi, güncel ve işlevine uygun algoritmaların kullanılması ve şifreleme anahtarlarının yönetiminin ve kontrolünün doğru bir şekilde yapılmasını içerir. Kriptografik hataların doğru bir şekilde ele alınması, hem finansal kayıpların önlenmesine yardımcı olabilir, hem de kurumların itibarlarını koruyabilir.

C. *Enjeksiyon Saldırıları (Injection)*

Enjeksiyon saldırıları, bir saldırganın bir yorumlayıcıya kötü amaçlı veri gönderdiği güvenlik açığı türleridir. Bu saldırılar genellikle, saldırganın yorumlayıcıya kötü amaçlı komutları vermesine izin veren kullanıcı girişinin yetersiz doğrulaması veya temizlenmesi sonucu ortaya çıkar [13]. Bu, genellikle saldırganın kontrol ettiği verinin, bir sorgu veya komutun bir parçası olarak yorumlanmasını sağlar. SQL enjeksiyonu, en yaygın ve en tehlikeli enjeksiyon saldırısı türlerinden biridir. Bu tür bir saldırıda, saldırgan, bir SQL sorgusuna kötü amaçlı bir parça ekler, bu da saldırganın veri tabanını kontrol etmesine, hassas verilere erişmesine veya veri tabanının yapısını değiştirmesine izin verebilir. Enjeksiyon saldırıları, diğer birçok yorumlayıcı için de geçerlidir, örneğin OS komutları, XML işlemciler veya LDAP sorguları. Bu tür bir saldırı, genellikle saldırganın belirli bir işlemci üzerinde kontrol sahibi olmasına izin verir, genellikle bu, verinin saldırgan tarafından kontrol edilen bir alanın bir parçası olarak yorumlanması anlamına gelir. Enjeksiyon saldırılarının başarılı olmasını önlemek için bir dizi strateji kullanılabilir. İlk olarak, kullanıcı girişlerinin her zaman doğru bir şekilde doğrulanması ve temizlenmesi önemlidir. İkinci olarak, parametreli sorgular veya hazırlanmış ifadeler gibi teknikler, yorumlayıcıyı kullanıcı verilerinin doğru bir şekilde izole edilmesini sağlar. Son olarak, en iyi uygulama genellikle en az ayrıcalık ilkesini uygulamaktır, bu da yorumlayıcının yapabileceği işlemleri sınırlar ve böylece potansiyel zararı azaltır [13].

D. *Güvensiz Tasarım (Insecure Design)*

Güvensiz tasarım, bir web uygulaması veya herhangi bir yazılımın geliştirme sürecinde güvenlik önlemlerinin yetersiz olarak veya hiç düşünülmediği durumları ifade eder. Bir uygulamanın güvenliği, tasarım aşamasında düşünülerek doğru bir şekilde uygulandığında, çok daha etkili ve etkin bir hale gelir. Bu nedenle, güvenlik konularını sonradan eklemek yerine, yazılımın tasarım aşamasından itibaren entegre etmek önemlidir. Güvensiz tasarım genellikle, uygulamanın mimarisi, altyapısı veya uygulamanın temel bileşenlerindeki güvenlik zafiyetlerini içerir. Bu tür hatalar genellikle saldırı yüzeyini genişletir ve bu da saldırganların yazılımın özelliklerini ve bileşenlerini kötüye kullanma imkanı sağlar.

Güvensiz tasarımın örneklerinden bazıları şunlardır:

- Yazılım mimarisinde veya altyapısında güvenlik önlemlerinin hiç düşünülmemesi,
- Güvenlik kontrollerinin yanlış veya eksik uygulanması,
- Güvenlikle ilgili teknolojilerin yanlış veya uygunsuz bir şekilde kullanılması.

Bu tür zafiyetlerle mücadele etmek için, yazılımın tasarım aşamasından itibaren güvenlikle ilgili konuların düşünülmesi ve güvenlik önlemlerinin uygulanması gerekmektedir. Kullanıcı kimlik doğrulama, oturum yönetimi, veri doğrulama ve günlükleme gibi uygulamanın temel bileşenleri, bir yazılımın güvenliği için hayati önem taşır. Bu bileşenlerin tasarımı ve uygulaması, yazılımın genel güvenlik durumunu belirler. Insecure Design, genellikle daha karmaşık ve kapsamlı saldırılara karşı savunmasızlık yaratır. Bu tür zafiyetler genellikle daha karmaşık ve zaman alıcı olabilir ve genellikle yüzey saldırısını genişletir.

E. *Yanlış Güvenlik Yapılandırmaları (Security Misconfiguration)*

Security Misconfiguration, bir web uygulamasının güvenlik ayarlarının hatalı veya yanlış bir şekilde yapılandırılması sonucu ortaya çıkan bir zafiyettir. Bu hatalı yapılandırmalar genellikle saldırganlara hassas bilgilere erişim imkanı sağlar veya uygulamanın düzgün bir şekilde çalışmasını engeller [13]. Security Misconfiguration genellikle aşağıdaki durumlarda meydana gelebilir:

- Hatalı veya yanlış uygulama sunucusu, veritabanı sunucusu, platform, çerçeve veya tüm sistem yapılandırmaları
- Default (varsayılan) hesapları ve şifrelerin devre dışı bırakılmaması
- Hatalı dosya ve kaynak izinleri
- Hatalı güvenlik başlık ayarları ve hatalı hata mesajları
- Güncel olmayan veya güvenli olmayan yazılım sürümlerinin kullanımı

Security Misconfiguration zafiyetinin etkisi genellikle uygulamanın veya sistemlerin doğru yapılandırılmama seviyesine bağlıdır. Bu tür bir zafiyet, saldırganın uygulamanın işleyişi hakkında bilgi edinmesine, yetkisiz olarak sistemlere erişmesine veya hassas kullanıcı verilerini ele geçirmesine olanak sağlar [13]. Security Misconfiguration'ı önlemek ve yönetmek için bir dizi yöntem kullanılabilir. İlk olarak, tüm yazılım ve

sistemlerin en güncel ve güvenli sürümleri kullanılmalıdır. İkinci olarak, gereksiz hizmetler ve özellikler devre dışı bırakılmalı, default hesaplar ve şifreler değiştirilmeli, hatalı dosya ve kaynak izinleri düzeltilmeli ve güvenlik başlıkları doğru bir şekilde yapılandırılmalıdır. Ayrıca, hata mesajlarının saldırganın işine yarayabilecek herhangi bir bilgi vermemesi gerektiğinden, bu mesajlar dikkatli bir şekilde yapılandırılmalıdır [20] Düzenli güvenlik taramaları ve otomatik yapılandırma kontrolleri yapılmalıdır. Bu, olası Security Misconfiguration zafiyetlerini tespit etmek ve hızlı bir şekilde düzeltmek için en etkili yöntemlerden biridir [20]

F. *Savunmasız ve Eski Bileşenler (Vulnerable and Outdated Components)*

Vulnerable and Outdated Components zafiyeti, bir web uygulamasının bileşenlerinin (örneğin, kütüphaneler, çerçeveler, diğer yazılım modülleri) güncel olmaması veya bilinen güvenlik açıkları içermesi durumunda ortaya çıkar. Bu durum saldırganların bu açıklıkları kullanarak sisteme zarar vermesine veya verilere erişmesine neden olabilir [13]. Bu zafiyetin oluşması genellikle, yazılım geliştiricilerin kullandıkları üçüncü taraf bileşenleri düzenli olarak güncelleme ihtiyacının göz ardı edilmesinden kaynaklanır. Birçok durumda, bu bileşenlerin son sürümlerinin kullanılmaması veya bileşenlerin içerdiği güvenlik açıklıklarının farkında olunmaması, saldırganların sistemi istismar etmesine ve veri ihlallerine yol açabilir [13]. Önlem almak için, geliştiricilerin bir bileşen güncelleme stratejisi uygulaması önemlidir. Bu, düzenli olarak kullanılan tüm bileşenlerin ve bağımlılıkların bir envanterinin çıkarılmasını ve güncellemelerin ve güvenlik düzeltmelerinin düzenli olarak takip edilmesini içerir [21] Ayrıca, kullanılan bileşenlerin güvenli yapılandırmaları hakkında bir anlayış geliştirmek, geliştiricilere bileşenlerin doğru bir şekilde nasıl kullanılacağı ve güvenlik açıklıklarının nasıl önleneceği konusunda rehberlik edebilir [13]. Yazılım geliştiricileri, bileşenlerin ve bağımlılıkların güncel versiyonlarını kullanmak için bir strateji uygulamalı ve bu bileşenlerin güvenliğini düzenli olarak izlemeli, değerlendirmeli ve test etmelidir. Bu, otomatik güvenlik taramaları ve bağımlılık kontrol araçları kullanılarak yapılabilir [21]

G. *Tanımlama ve Kimlik Doğrulama Hataları (Identification and Authentication Failures)*

"Identification and Authentication Failures" zafiyeti, kullanıcıların sisteme kimliklerini doğrulama ve erişim sağlama mekanizmalarının yetersiz veya hatalı olması durumunda ortaya çıkar. Bu tür bir güvenlik zafiyeti, saldırganların yetkisiz kullanıcılar gibi hareket etmelerine ve sistem üzerinde kontrol sağlamalarına veya hassas verilere erişmelerine izin verebilir [13]. Bu tür zafiyetler genellikle, kimlik doğrulama mekanizmalarında zayıf veya tahmin edilebilir kimlik bilgilerinin kullanılması, çok sayıda başarısız oturum açma denemesine izin verilmesi veya oturum bilgilerinin hatalı yönetimi gibi nedenlerle ortaya çıkar. Bu tür zafiyetlerden korunmak için, güçlü kimlik doğrulama ve oturum yönetimi uygulamaları kullanılması önemlidir. Bunlar arasında çok faktörlü kimlik doğrulama, karmaşık ve benzersiz parolaların kullanılması, oturum bilgilerinin güvenli bir şekilde saklanması ve geçersiz kılınması ve oturum süre aşımlarının uygulanması bulunur [13]. Ayrıca, oturum bilgilerinin hatalı yönetimi ve yetkilendirme hataları, saldırganların erişim kontrolünü atlatmasına ve yetkisiz erişim sağlamasına neden olabilir. Bu nedenle, etkili bir erişim kontrol politikasının uygulanması ve düzenli olarak güncellenmesi de bu tür zafiyetleri önlemede önemlidir. Bu tür zafiyetlerin önlenmesi, güvenlik protokollerinin ve uygulamalarının düzgün bir şekilde uygulanmasını, sistem ve verilerin güvenliğinin sürekli olarak izlenmesini ve değerlendirilmesini gerektirir.

H. *Yazılım ve Veri Bütünlüğü Hataları (Software and Data Integrity Failures)*

"Software and Data Integrity Failures" zafiyeti, yazılım ve veri bütünlüğünün korunamaması durumunda ortaya çıkar. Yazılım ve veri bütünlüğü, sistemlerin ve verinin doğru, tam ve tutarlı kalmasını sağlamak için gereklidir. Yazılım ve veri bütünlüğü ihlalleri, genellikle bilgisayar virüsleri, kötü amaçlı yazılım, insan hataları veya sistem hataları gibi nedenlerle oluşur [13]. Yazılım ve veri bütünlüğünün başarısız olmasının sebeplerinden biri, doğrulama ve kontrol süreçlerinin zayıf veya eksik olmasıdır. Örneğin, bir yazılım güncellemesi yapıldığında, güncellemenin doğruluğunun ve bütünlüğünün doğrulanması gereklidir. Aksi takdirde bu, kötü amaçlı bir aktörün zararlı kodu sisteme yerleştirmesine veya verileri değiştirmesine olanak sağlar. Bunun yanı sıra, veri bütünlüğü hataları da veri tabanı bozulmaları, hatalı veri girişi

veya veri tabanı sistemlerinin hatalı konfigürasyonları nedeniyle ortaya çıkabilir. Bu tür hatalar genellikle veri kaybına veya hatalı verilere neden olur, bu da yanıltıcı veya yanlış kararların alınmasına yol açabilir. Software and Data Integrity Failures'ı önlemek için, güçlü veri doğrulama ve kontrol mekanizmalarının uygulanması gereklidir. Bunun yanı sıra, yazılım ve veri tabanı sistemlerinin düzenli olarak bakımının yapılması ve güncellemelerin doğru bir şekilde uygulanması önemlidir. Ayrıca, yetkisiz erişimin önlenmesi ve sistemlerin düzenli olarak izlenmesi ve denetlenmesi, bu tür zafiyetlerin ortaya çıkmasını önleyebilir [13].

İ. Güvenlik Kayıtları ve İzleme Hataları (Security Logging and Monitoring Failures)

"Security Logging and Monitoring Failures", bir sistemin güvenlik olaylarını doğru bir şekilde kaydetme ve izleme yeteneğinin yetersiz olması durumunda ortaya çıkan bir web uygulama zafiyetidir. Etkin bir güvenlik izleme ve kayıt sistemine sahip olmamak, saldırganların bir sistemdeki zayıf noktaları keşfetmesine ve istismar etmesine olanak sağlar [13]. Güvenlik kayıtları, bir sistemdeki güvenlikle ilgili olayları izlemenin ve belgelendirmenin bir yolu olduğu için kritiktir. Bunlar, bir saldırı girişimini, başarısız veya başarılı bir oturum açma denemesini, veri erişimini veya herhangi bir anormal etkinliği belgelendirebilir. Etkili bir güvenlik kayıt sistemi, bir organizasyonun potansiyel güvenlik tehditlerini belirlemesine, yanıtlamasına ve önlemesine yardımcı olabilir. Bunun yanı sıra, sürekli güvenlik izleme, bir sistemin durumunu ve performansını anlamak için gereklidir. Bu, ağ trafiğini, sistem kaynaklarını, kullanıcıları ve uygulamaları içerir. Güvenlik izleme, bir sistemde neyin normal ve neyin anormal olduğunu belirleme yeteneği sağlar, bu da potansiyel güvenlik tehditlerini daha hızlı ve daha etkin bir şekilde belirlemeye yardımcı olur.

Security Logging and Monitoring Failures'ı önlemek için, organizasyonların etkin bir güvenlik kayıt ve izleme politikası uygulaması gereklidir. Bu politika, neyin kaydedileceğini, neyin izleneceğini, ne zaman ve nasıl analiz edileceğini ve kayıtların nasıl saklanacağını ve korunacağını belirlemelidir. Ayrıca, sistem ve ağ güvenliği uzmanlarının düzenli olarak eğitilmesi ve güncellenmesi de önemlidir, böylece yeni ve gelişen tehditleri tanıyabilir ve buna göre yanıt verebilirler [13].

J. Sunucu Taraflı İstek Sahteciliği (Server-Side Request Forgery)

Server-Side Request Forgery (SSRF) bir web uygulaması zafiyetidir. Bu zafiyet, saldırganın bir sunucunun dahili ağ kaynaklarına erişimini sağlar veya sunucunun işlemlerini tetiklemesine izin verir. SSRF saldırıları, genellikle bir saldırganın ağ isteklerini yanıtlamak için tasarlanmış bir web uygulamasını veya API'yi istismar etmek yoluyla gerçekleşir [13]. SSRF saldırıları öncelikle iki şekilde çalışır: ilk yol, saldırganın kendi sunucusuna bir istek göndererek sunucudan gelen yanıtları almasıdır. Bu, saldırganın sunucunun belirli özelliklerini, işlemlerini veya verilerini ifşa etmesine yol açabilir. İkinci yol, saldırganın sunucu üzerinden bir hedefe istek göndermesi ve hedefin yanıtlarını almasıdır. Bu, bir saldırganın aksi takdirde ulaşılamayan bir hedefe erişmesini sağlayabilir. SSRF saldırılarının tehlikesi, genellikle saldırganların bir sunucunun dahili ağ kaynaklarına erişimini sağlaması ve sunucunun işlemlerini tetiklemesine izin vermesidir. Bu, potansiyel olarak hassas bilgilere erişimi, yanıtları değiştirme veya hatta sunucunun tamamen ele geçirilmesini içerebilir. SSRF saldırıları genellikle dikkatli ağ ayarları ve düzenlemeleri ile önlenir, ancak bu her zaman pratik veya mümkün olmayabilir. Bu nedenle, web uygulamalarının ve API'lerin SSRF saldırılarına karşı korunabilmesi için güvenli kodlama uygulamaları ve dikkatli yapılandırma önemlidir [13]. Bu tür saldırıları önlemek için, özellikle dış hizmetlere istek gönderme yeteneğine sahip olan uygulamaların geliştiricileri, uygun giriş doğrulama ve sınırlama mekanizmaları kullanmalıdır. Bu, URL'leri beyaz listeye almak veya belirli ağ bölgelerine erişim taleplerini sınırlamak gibi eylemleri içerebilir.

Tablo 1. Zafiyetlerin web uygulamaları üzerinde yaratabileceği potansiyel etkiler

OWASP 2021 Top 10 Zafiyeti	Oturum Çalma	Veri Tabanına Erişim	Yetki Yükseltme	Kaynak Kodlarına Erişim	Veri Etkileme	Servis Dışı Bırakma	Kişisel Verilere Erişim
Broken Access Control			✓	✓	✓		✓
Cryptographic Failures	✓	✓					✓
Injection		✓			✓		
Insecure Design	✓	✓	✓	✓	✓	✓	✓
Security Misconfiguration		✓	✓	✓		✓	
Vulnerable and Outdated Components		✓	✓	✓	✓	✓	✓
Identification and Authentication Failures	✓		✓				✓
Software and Data Integrity Failures		✓			✓		
Security Logging and Monitoring Failures						✓	
Server-Side Request Forgery (SSRF)		✓		✓	✓		

III. AÇIKLANABİLİR YAPAY ZEKA

Yapay zekâ (YZ), makinelerin insan zekâsının belirli yönlerini simüle etme yetenekleri anlamına gelir. Bu, özellikle makine öğrenmesi ve derin öğrenme kavramlarında belirgin hale gelmiştir. Makine öğrenmesi, algoritmaların ve istatistiksel modellerin verilerden bağımsızca öğrenme ve tahminler yapma yetenekleri üzerine kurulmuştur. Derin öğrenme ise, makine öğrenmesinin bir alt kümesi olarak kabul edilir ve esas olarak yapay sinir ağlarının karmaşık veri kümelerinden öğrenme yeteneklerine odaklanır. Derin öğrenme algoritmaları, katmanlı sinir ağları (yüzlerce veya hatta binlerce katmanlı olabilen) kullanır ve bu ağlar milyonlarca veya hatta milyarlarca parametre içerebilir. Bu, bu tür bir YZ'nin "kara kutu" olarak görülmesine yol açar. Yani, bu sistemler bir girdi aldığında ve bir çıktı verdiğinde, aradaki süreç genellikle insan tarafından anlaşılmaz. Açıklanabilir Yapay Zekâ (AYZ), bu durumu çözme girişimidir. AYZ, algoritmaların karar
verme süreçlerinin daha açık ve anlaşılır olmasını sağlar [26].

Açıklanabilir Yapay Zekâ (XAI), yapay zekâ ve makine öğrenmesi uygulamalarının işleyişini ve sonuçlarını anlama ve yorumlama kapasitesini artıran teknikler ve yöntemler kümesidir. XAI'nin amacı, model çıktılarını ve kararlarını anlamlı ve anlaşılır bir biçimde sunmaktır. Bu, modelin iç çalışma mekanizmalarının ve işleyişinin daha fazla şeffaflık ve anlaşılabilirlik ile sunulmasını gerektirir. XAI'ın önemi, modelin uygulandığı alan ve beklenen sonuçlara bağlı olarak değişir. Örneğin, bir görüntü sınıflandırma modelinde, bir görselin kedi mi yoksa köpek mi olduğunu belirlemek genellikle düşük açıklanabilirlik ihtiyacı gerektirir. Bu durumda, modelin doğru tahmin yapmasına

odaklanabiliriz ve modelin iç işleyişini anlamak genellikle ikincil bir öneme sahiptir. Öte yandan, medikal teşhis modelleri gibi daha kritik uygulamalarda, yüksek düzeyde açıklanabilirlik ihtiyacı olabilir. Örneğin, bir model bir hastanın kanser olup olmadığını tahmin ediyorsa, bu tahminin dayandığı belirli özellikler veya desenlerin anlaşılması çok önemlidir. Bu tür bir durumda, modelin bir hastanın kanser olup olmadığını belirlemek için hangi özellikleri veya desenleri kullandığını anlamak, teşhisin güvenilirliğini ve doğruluğunu belirlemek için kritik öneme sahip olabilir. Ayrıca, bir modelin açıklanabilirliği, modeli kullanan kişilerin, modelin neden belirli bir sonuç verdiğini anlamasına ve böylece modelin güvenilirliği ve doğruluğuna olan güvenini artırmasına yardımcı olabilir. Bu, yapay zekâ ve makine öğrenmesi modellerinin daha geniş bir kabul ve uygulama bulmasına yardımcı olabilir. XAI, bu nedenle, yapay zekâ ve makine öğrenmesinin etik ve sorumlu kullanımını teşvik eden önemli bir araçtır [27].

A. *Açıklanabilir Yapay Zekânın Avantajları*

Karar Verme Sürecinin Şeffaflığı: XAI, makine öğrenmesi ve yapay zekâ modelinin tahminlerini ve kararlarını şeffaflaştırır [16]. Bu, modelin iç işleyişini anlama ve böylece kullanıcıların ve paydaşların modelin sonuçlarına daha fazla güven duymalarını sağlar.

Risk Azaltma: XAI, modelin yanıltıcı veya hatalı sonuçlar verme riskini azaltabilir [6]. XAI ile, modelin hangi özellikleri ve desenleri kullandığı açıklanabilir, bu da modelin hatalı tahminler yapmasına veya yanıltıcı sonuçlar vermesine yol açabilecek hataların belirlenmesine ve düzeltilmesine yardımcı olabilir.

Daha İyi Model Geliştirme: XAI, model geliştiricilerinin bir modelin ne kadar iyi performans gösterdiğini ve hangi özelliklerin en etkili olduğunu anlamalarını sağlar [12]. Bu, modeli iyileştirmek ve optimize etmek için bilgili kararlar almayı mümkün kılar.

Daha Fazla Uygulama Alanı: XAI, makine öğrenmesi ve yapay zekânın daha geniş uygulamalar için kabulünü artırır [1]. XAI'nin sağladığı şeffaflık ve anlaşılabilirlik, kullanıcıların ve paydaşların bu teknolojilere daha fazla güven duymasını ve daha geniş uygulamalar için bu teknolojileri kabul etmelerini sağlar.

Etik ve Yasal Uygunluk: Birçok durumda, XAI, kullanıcıların ve düzenleyicilerin, modelin adaletsiz veya ayrımcı sonuçlar üretip üretmediğini belirlemesine yardımcı olabilir [19]. Bu, hem etik hem de yasal gerekliliklere uyumu sağlamak için önemli olabilir, özellikle GDPR gibi düzenlemeler kapsamında "otomatik karar alma" hakları olduğunda.

Kullanıcı Güvenini Artırma: XAI, son kullanıcılarda yapay zekâya güveni artırır [15]. Bu, kullanıcıların yapay zekâ tabanlı sistemlere daha fazla güven duymalarını ve bu sistemleri daha geniş bir şekilde benimsemelerini sağlar.

Sorumluluk ve Sorumluluk Tespiti: XAI, bir modelin bir hata yaptığı durumlarda, bu hatanın kaynağının belirlenmesine yardımcı olabilir . Bu, modelin geliştiricilerinin ve kullanıcılarının, modelin hatalarından ve sonuçlarından sorumluluk almasını ve bu hataları düzeltmek için gerekli adımları atmasını sağlar.

Model Validasyonu: XAI, modelin doğru bir şekilde çalışıp çalışmadığını doğrulamada bir araç olarak kullanılabilir . Bu, bir modelin doğruluğunu ve güvenilirliğini artırabilir.

XAI'nin makine öğrenmesi ve yapay zekâ tabanlı sistemlerin daha güvenli, adil, güvenilir ve etkili kullanımını desteklemek için nasıl bir araç olarak kullanılabileceğini göstermektedir. Ancak, XAI'nin kullanımı, modelin uygulandığı alana ve kullanıcıların ve paydaşların gereksinimlerine bağlı olarak değişebilir. XAI, ayrıca bazı durumlarda modelin karmaşıklığını artırabilir ve performansı etkileyebilir, bu nedenle her durumda dikkatli bir değerlendirme yapılmalıdır.

Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence - XAI), bir yapay zeka modelinin kararlarını ve tahminlerini anlaşılabilir ve denetlenebilir bir biçimde sunabilme yeteneği demektir. XAI, hem işletmelerin algoritmanın nasıl çalıştığını daha iyi anlamasına yardımcı olur, hem de hukuki düzenlemelere uygunluk ve hesap verebilirliği artırır.

B. *Açıklanabilir Yapay Zekânın Kullanım Alanları*

Sağlık Hizmetleri: XAI, hastaların durumlarına dair karmaşık analizleri anlamada ve hatta hastalıkların teşhisinde yardımcı olabilir. Yapay zeka algoritmaları, genellikle doktorlar tarafından anlaşılmayan birçok özelliği analiz eder. Ancak XAI sayesinde, doktorlar bir yapay zeka modelinin

bir hastalığı teşhis etmek için hangi belirtileri, hangi laboratuvar sonuçlarını kullandığını anlayabilir.
Finans: XAI, kredi risk analizleri ve yatırım tavsiyeleri gibi birçok finansal işlemde kullanılır. Bu alanlarda, XAI'nin kararları ve tahminleri genellikle düzenleyiciler ve müşteriler tarafından denetlenir. Bu denetimler, bankaların ve finans kuruluşlarının hukuki düzenlemelere uymasını sağlar ve müşterilerin XAI'nin tavsiyelerini daha iyi anlamasına yardımcı olur.
Otomotiv Sektörü: Özellikle otonom araçlarda, XAI'nin güvenliği artırma ve sürüş kararlarını açıklama potansiyeli vardır. XAI sayesinde, otonom bir aracın belirli bir durumda neden belirli bir manevra yaptığını veya bir kaza durumunda neyin yanlış gittiğini anlamak mümkün olabilir.
E-ticaret: XAI, kullanıcılarına kişiselleştirilmiş ürün önerileri sunan öneri sistemlerinde kullanılabilir. Müşteriler, bir ürünün neden önerildiğini anladıklarında, bu öneriye daha fazla güven duyarlar. Bu, müşteri memnuniyetini ve satışları artırabilir.
İnsan Kaynakları: XAI, işe alım süreçlerinde ve performans değerlendirmelerinde kullanılabilir. İşe alım sürecinde, bir XAI modeli, bir adayın neden seçildiğini veya reddedildiğini açıklayabilir. Bu, işe alım sürecinin adil ve tarafsız olduğunu gösterir ve ayrımcılık iddialarını önler.
Enerji ve Çevre: XAI, enerji kullanımını optimize etmek ve çevresel etkileri azaltmak için kullanılabilir. Örneğin, bir XAI modeli, bir bina veya fabrikanın enerji kullanımını nasıl en aza indirebileceğini açıklayabilir.
Bu kullanım alanları, XAI'nin nasıl daha etkili ve şeffaf bir şekilde yapay zeka kullanılabileceğini göstermektedir. Ancak, XAI'nin tam potansiyelini kullanabilmek için hâlâ daha fazla araştırma ve geliştirme gerekmektedir.

C. *Açıklanabilir Yapay Zeka Modelleri*

Açıklanabilir Yapay Zeka modelleri, makine öğrenimi modelinin tahminlerini ve kararlarını insanlar için anlaşılabilir kılmayı amaçlar. Bu, modelin iç çalışma mekanizmalarını ve özelliklerin tahminlere nasıl katkıda bulunduğunu anlamak için önemlidir. AYZ, bir dizi teknik ve yaklaşımla gerçekleştirilebilir. En yaygın AYZ modelleri şu şekildedir;
LIME (Local Interpretable Model-agnostic Explanations): LIME, bir makine öğrenimi modelinin karmaşık tahminlerini yerel olarak açıklamak için basit modeller kullanır. LIME, bir örneklem etrafında bir dizi bozulmuş veri noktası üretir ve bu verilere dayalı olarak basit bir model eğitir. Bu basit model, karmaşık modelin kararlarının anlaşılmasına yardımcı olur [22].
SHAP (SHapley Additive exPlanations): SHAP değerleri, bir oyun teorisi kavramı olan Shapley değerlerine dayanır ve her özelliğin modelin çıktısına katkısını açıklar. SHAP, her özelliğin katkısını adil bir şekilde tahmin eder ve modelin kararlarını daha şeffaf hale getirir [23].
Karar Ağaçları: Karar ağaçları, doğal olarak açıklayıcıdır çünkü veriyi basit karar kuralları kullanarak bölümlere ayırır. Karar ağaçları, modelin nasıl kararlar verdiğini ve hangi özelliklerin önemli olduğunu görsel bir şekilde gösterir [24].
Katsayı Tabanlı Modeller: Lineer regresyon ve lojistik regresyon gibi katsayı tabanlı modeller, özelliklerin modelin tahminlerine olan katkısını açıkça gösterir. Her özelliğin bir katsayısı vardır ve bu katsayı, özelliğin tahmin üzerindeki etkisini gösterir [25].
Karşılaştırmalı Açıklama Yöntemleri: Bu tür yöntemler, modelin tahminlerini, kullanıcının anlayabileceği referans noktalarına veya karşılaştırmalı örneklemelere dayalı olarak açıklar. Örnek olarak, "counterfactual explanations" (karşıt-gerçek açıklamalar) adlı yöntem, bir örneklem için modelin farklı bir sonuç üretmesi için hangi özelliklerin nasıl değiştirilmesi gerektiğini gösterir [22].
Özellik Önemi: Rastgele orman gibi modeller, özelliklerin önemini ölçer ve hangi özelliklerin modelin tahminlerini en çok etkilediğini gösterir. Özellik önemi, modelin kararlarını daha anlaşılabilir kılmaya yardımcı olabilir [23].

IV. SİBER GÜVENLİKTE AÇIKLANABİLİR YAPAY ZEKA

Siber güvenlikte Açıklanabilir Yapay Zeka'nın (AYZ) yeri, karmaşık ve dinamik tehditleri anlamak ve önlemek için şeffaf ve anlaşılabilir modeller gerektirmesinden kaynaklanmaktadır. AYZ'nin etkili bir şekilde uygulanması, tehdit tespiti ve müdahale süreçlerini güçlendirirken, yasal uyumu da sağlar. Bu, siber güvenlik çözümlerinin genel etkinliğini ve güvenilirliğini artırabilir. AYZ, siber güvenlikte birçok önemli rol oynar:
Güvenin Artırılması: Yapay Zeka modellerinin “siyah kutu” doğası, çoğu zaman nasıl karar verdiklerini anlamayı zorlaştırır. AYZ, güvenlik

uzmanlarının ve sistem yöneticilerinin, YZ modellerinin tahminlerini ve kararlarını daha iyi anlamasına yardımcı olur. Bu, kullanıcıların YZ modellerine daha fazla güven duymasına ve bu modelleri daha etkili bir şekilde kullanmasına olanak tanır.

Daha Hızlı ve Doğru Karar Alma: Siber güvenlikte, zaman kritik bir faktördür. AYZ, güvenlik uzmanlarının, YZ modellerinin tahminlerini daha hızlı bir şekilde anlamalarına ve bu bilgileri, potansiyel tehditlere hızlı bir şekilde yanıt vermek için kullanmalarına olanak tanır.

Düzenleyici Uyum: AYZ, siber güvenlik çözümlerinin, Genel Veri Koruma Yönetmeliği (GDPR) gibi düzenlemelere uyum sağlamasına yardımcı olabilir. GDPR, bireylerin otomatik karar alma süreçlerine dair açıklamalar talep etme hakkını tanır ve AYZ bu tür yasal gereklilikleri karşılamaya yardımcı olabilir.

Model Optimizasyonu: AYZ, güvenlik uzmanlarına, modelin nasıl çalıştığına dair ayrıntılı bilgiler sunar. Bu, modelin performansını artırmak ve yanlış pozitif veya yanlış negatif tahminleri azaltmak amacıyla modeli optimize etmek için kritik öneme sahiptir.

Olay İncelemesi ve Eğitim: Bir güvenlik olayı meydana geldiğinde, AYZ, olayın neden meydana geldiğini ve YZ sisteminin nasıl tepki verdiğini anlamak için değerli iç görüler sağlar. Bu, gelecekteki olaylara daha etkili yanıt vermek için stratejiler geliştirmeye ve güvenlik ekiplerini eğitmeye yardımcı olabilir.

V. SONUÇ

Açıklanabilir Yapay Zeka'nın siber güvenlikte kullanımı, daha etkili güvenlik önlemleri alınmasını sağlayarak, siber tehditlerle daha etkili bir şekilde mücadele etmeyi mümkün kılar. Ayrıca, siber güvenlik uzmanlarının, YZ modellerinin çalışma şeklini ve alınan kararları daha iyi anlamalarını sağlayarak, güven ve şeffaflığı artırır. Çalışmanın sonucunda açıklanabilir yapay zekanın web uygulamalarının zafiyetlerinin keşfindeki kullanımının aşağıdaki şekilde olabileceği kanaatine varılmıştır.

Anomali Tespiti: AYZ, ağ trafiğinde veya sistem davranışlarında normalden sapmaları (anomalileri) tespit etmekte kullanılabilir. AYZ modelleri, hangi özelliklerin anomali olarak kabul edilmesine yol açtığını açıklayarak, güvenlik uzmanlarına daha derinlemesine analiz ve müdahale olanağı sağlar. Örneğin, bir AYZ modeli ağ trafiğindeki anormal bir artışı tespit edip, bu artışın neden potansiyel bir DDoS saldırısına işaret ettiğini açıklayabilir.

Kimlik Doğrulama ve Erişim Kontrolü: AYZ, kullanıcı kimlik doğrulaması ve erişim kontrol sistemlerinde kullanılabilir. AYZ, bir kullanıcının erişim talebini reddediyorsa, reddetme nedenini açıklayarak, sistemin güvenilirliğini ve kullanıcıların sisteme olan güvenini artırır.

Saldırı Tespit Sistemleri (IDS): AYZ, siber saldırıların tespitinde kullanılabilir. Geleneksel YZ modelleri sadece bir saldırının varlığını tespit edebilirken, AYZ hangi veri noktalarının saldırıya işaret ettiğini açıklayabilir. Bu, güvenlik uzmanlarının hızlı ve etkili bir şekilde müdahale etmelerine yardımcı olur.

Zararlı Yazılım Tespiti: AYZ, zararlı yazılımları tespit etmek ve analiz etmek için kullanılabilir. AYZ, bir dosyanın zararlı yazılım olarak sınıflandırılmasına yol açan özellikleri açıklayabilir. Bu, güvenlik uzmanlarının zararlı yazılımın davranışını anlamalarına ve etkili koruma stratejileri geliştirmelerine yardımcı olur.

Güvenlik Olaylarının İncelenmesi ve Raporlanması: AYZ, güvenlik olaylarına dair daha kapsamlı raporlar oluşturmak için kullanılabilir. Bir güvenlik olayının ardından, AYZ kullanılarak, olayın neden meydana geldiği, hangi sistemlerin etkilendiği ve benzer olayların gelecekte nasıl önlenebileceği hakkında detaylı bilgiler sağlanabilir.

Veri Sızıntısı Önleme: AYZ, veri sızıntısı tespiti ve önlemede kullanılabilir. AYZ, sistemlerin hassas verileri nasıl işlediğini açıklayarak ve olası sızıntı risklerini belirleyerek güvenlik politikalarını geliştirmeye yardımcı olabilir.

Web uygulamalarının zafiyetlerinin keşfinde Açıklanabilir Yapay Zekanın kullanımı, güvenlik uzmanlarına, zafiyetlerin kökenini ve doğasını daha iyi anlama ve bu zafiyetlere etkili bir şekilde müdahale etme olanağı sağlar. Siber güvenlik tehditlerinin karmaşıklığı ve çeşitliliği göz önüne alındığında, açıklanabilirlik, güvenlik uzmanlarının karşılaştıkları zorlukları çözmelerine yardımcı olabilir. Ayrıca, organizasyonların güvenlik altyapılarını güçlendirerek, web uygulamalarının güvenilirliğini ve kullanıcı verilerinin gizliliğini korumalarına yardımcı olur. İleride, AYZ'nin web uygulama güvenliğini artırmada daha da etkili hale gelmesi ve bu alandaki inovasyonların hızlanması beklenmektedir. Ayrıca, siber güvenlik ekiplerinin, AYZ modellerini kendi güvenlik çözümlerine

entegre etme ve bu modellerden elde edilen içgörüleri kullanma yeteneklerini geliştirmeleri gerekmektedir.

KAYNAKLAR

[1] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160.

[2] Akıllı, N., & Güngör, V. C. (2019). Derin öğrenme. Bilişim Teknolojileri Dergisi, 12(1), 75-84.

[3] Bakırcı, T. (2020). Yapay öğrenme ve siber güvenlik. Bilgi Ekonomisi ve Yönetimi Dergisi, 15(1), 23-33.

[4] Bengio, Y. (2019). Derin öğrenme. Hacettepe Üniversitesi Mühendislik Fakültesi Dergisi, 30(1), 1-7.

[5] Ermiş, U., Gören Şahin, S., & Bulkan, S. (2018). Siber güvenlik ve yapay zeka. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 23(3), 751-763.

[6] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42.

[7] Gürbüz, M. Z., Ekşi, İ., & Aydın, M. A. (2021). Siber Güvenlikte Yapay Zekanın Rolü ve Etkinlik Analizi. Bilişim Teknolojileri Dergisi, 14(1), 1-14.

[8] İnan, A., Kocadag, M., & Şen, S. (2020). Türkiye'de yapay zeka uygulamaları: mevcut durum, fırsatlar ve tehditler. Teknolojik Araştırmalar: Yeni Teknoloji, Yeni Toplum, 1(1), 45-54.

[9] Kaya, B., & Alhajj, R. (2019). Açıklanabilir yapay zeka: Yeni bir yaklaşım ve bir uygulama. ACM Bilişim Dergisi, 30(1), 22-29.

[10] Keser, A. (2019). Web uygulamalarının güvenlik zafiyetleri ve saldırı teknikleri: SQL enjeksiyonu, XSS ve CSRF. Bilişim Teknolojileri Dergisi, 12(1), 85-94.

[11] Kirda, E., & Invernizzi, L. (2020). Understanding cybercrime from its roots to its current state. Computers & Security, 92, 101733.

[12] Molnar, C. (2020). Interpretable machine learning. Lulu. com.

[13] OWASP. (2021). OWASP Top Ten 2021. Open Web Application Security Project. https://owasp.org/Top10/

[14] Öztürk, O., & Zaim, S. (2019). Türkiye'de Siber Güvenlik ve Siber Saldırılar: Sorunlar ve Çözüm Önerileri. Sosyoekonomi, 27(40), 11-24.

[15] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

[16] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215.

[17] Sakar, B. E., & Kursun, O. (2020). Bir derin öğrenme yöntemi olan evrişimli sinir ağlarının tıbbi görüntü sınıflandırmadaki performansının incelenmesi. Bilgisayar Bilimleri ve Mühendisliği Dergisi, 6(2), 135-146.

[18] Singh, R., Singh, K., & Kim, T. H. (2020). Rise in cyber security concerns amid COVID-19 pandemic: an overview. Journal of Information Security and Applications, 54, 102581.

[19] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. Science Robotics, 2(6), eaan6080.

[20] Deepa, G., & Thilagam, P. S. (2016). Securing web applications from injection and logic vulnerabilities: Approaches and challenges. Information and Software Technology, 74, 160-180.

[21] Felmetsger, V., Cavedon, L., Kruegel, C., & Vigna, G. (2010). Toward automated detection of logic vulnerabilities in web applications. In 19th USENIX Security Symposium (USENIX Security 10).

[22] Nagaraj, P., Muneeswaran, V., Dharanidharan, A., Balananthanan, K., Arunkumar, M., & Rajkumar, C. (2022, April). A Prediction and Recommendation System for Diabetes Mellitus using XAI-based Lime Explainer. In 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (pp. 1472-1478). IEEE.

[23] Chromik, M. (2020). reshape: A framework for interactive explanations in xai based on shap.

[24] Mahbooba, B., Timilsina, M., Sahal, R., & Serrano, M. (2021). Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model. Complexity, 2021, 1-11.

[25] Weber, L., Lapuschkin, S., Binder, A., & Samek, W. (2022). Beyond explaining: Opportunities and challenges of XAI-based model improvement. Information Fusion.

[26] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. Science robotics, 4(37), eaay7120.

[27] Čyras, K., Rago, A., Albini, E., Baroni, P., & Toni, F. (2021). Argumentative XAI: a survey. arXiv preprint arXiv:2105.11266.