YAPAY ZEKA YÖNTEMLERİYLE DESTEKLENEN TRAFO MERKEZİ EMÜLATÖRÜNÜN, SİBER GÜVENLİK ÇALIŞMALARINDA ETKİN BİR ARAÇ OLARAK KULLANILMASI

USING THE SUBSTATION EMULATOR SUPPORTED BY AI METHODS AS AN EFFECTIVE TOOL IN CYBER SECURITY WORKS

Sude Kozalıoğlu[1], Necati Keskin[2], Oğuzhan Elbil[3]
[1,2,3]ADM Elektrik Dağıtım A.Ş., Ar-Ge Müdürlüğü
sude.kozalioglu@admelektrik.com.tr, necati.keskin@admelektrik.com.tr, oguzhan.elbil@admelektrik.com.tr

ÖZET

Yeni nesil güvenlik ürünlerinde ve AR-GE çalışmalarında yapay zekâ teknikleri kullanılarak, SCADA sistemlerine yönelik yapılan siber saldırılara karşı koruma sağlanmaktadır. Ancak, saldırıların evrimi ve saldırı stratejilerindeki değişiklikler nedeniyle halka açık veri setleri ile yapay zekâ modellerinin eğitilmesi yetersiz kalmaktadır. Bu nedenle, bu çalışmada, bir trafo merkezi emülatörü geliştirilerek, normal ve anormal veri setleri toplanarak veri temizleme, veri entegrasyonu ve ayrıklaştırma işlemleri yapılmıştır. Öznitelik belirlenmesi çalışmaları yapılarak NSL-KDD veri seti kullanılarak yapay zeka modelleri eğitilmiştir. Modeller üzerinde cross validation yapılarak başarı değerleri karşılaştırılmıştır. Model AWS Sagemaker platformunda deploy edilerek sürekli iyileştirilmesi için contunious learning döngüsü içinde kullanılabilir hale getirilmiştir. Geliştirilen öğrenen uygulama ile kullanıcılar SCADA networkünden toplanan paketleri normal-anormal olarak sınıflandırabilmektedir.

Anahtar Kelimeler: *SCADA Sistemi, Yapay Zeka Teknikleri, Veri Seti, Trafo Merkezi Emülatörü, Anomali Tespiti, Model Eğitimi, NSL-KDD Veri Seti, AWS Sagemaker, Normal-Anormal Sınıflandırma.*

ABSTRACT

By using artificial intelligence techniques in new generation security products and R&D studies, protection against cyber attacks against SCADA systems is provided. However, due to the evolution of attacks and changes in attack strategies, training AI models with publicly available datasets is insufficient. Therefore, in this study, a substation emulator was developed and normal and abnormal data sets were collected and data cleaning, data integration and discretization processes were performed. Artificial intelligence models were trained using the NSL-KDD data set by performing feature determination studies. Success values were compared by cross validation on the models. The model was deployed on the AWS Sagemaker platform and made available within the continuous learning cycle for continuous improvement. With the developed learning application, users can classify the packets collected from the SCADA network as normal-abnormal.

Keywords: *SCADA System, Artificial Intelligence Techniques, Data Set, Substation Emulator, Anomaly Detection, Model Training, NSL-KDD Data Set, AWS Sagemaker, Normal-Anomalous Classification.*

1. GİRİŞ

Su dağıtım şebekelerinden elektrik üretim tesislerine kadar birçok kritik altyapı tesisinin kontrolü SCADA sistemleri ile sağlanmaktadır. SCADA sistemleri; güvenliği değil, yüksek kullanılabilirliği sağlayan tasarım mimarileri, kullandıkları haberleşme protokolleri ve uzaktan erişime ihtiyaç duymalarından dolayı siber saldırılara karşı korumasızdır.

SCADA sistemleri siber saldırganların stratejik hedefleri arasında başı çekerken ulusal savunma örgütlerinin en büyük endişelerinden birisidir. Çünkü bir petrol hattını yöneten SCADA sistemine

düzenlenecek siber saldırı patlamaya ve can kaybına sebep olabilecekken; elektrik santrallerine yönelik yapılacak siber saldırı ülke genelinde elektrik kesintisine sebep olmaktadır. Uluslararası güvenlik raporları incelendiğinde gün geçtikçe SCADA sistemlerine yapılan saldırıların sayısının arttığı ve saldırı yöntemlerinin değişkenlik gösterdiği görülmektedir. Bu yüzden SCADA sistemlerinin güvenliği için sunulan çözüm; hem iç networkte oluşabilecek operasyonel tehditlere hem dış networkten gelebilecek saldırılara karşı koruma sağlamalıdır. Buna ek olarak bazı kurumlar saldırıya uğrayıp uğramadığının bile farkında değildir. Kullanılan geleneksel güvenlik ürünleri yetkisiz erişimi engelleme yeteneğine sahiptir ancak saldırı analizine ya da saldırgan profili hakkında yorum yapmaya imkan sağlamazlar. Günümüzde SCADA sistemleri siber savaşın en kritik hedefleri olarak belirlendiğinden SCADA sistemlerine yapılacak saldırının tipi, saldırgan profili hakkında yorum yapabilmek kritik tesisler için önem arz etmektedir. SCADA sistemlerine yapılacak siber saldırıların arkasında; meraklı bir siber saldırgan çocuk olabileceği gibi devlet destekli ve savaş sebebi sayılabilecek büyük organize çalışmalar da olabilmektedir. Bu yüzden bu sistemlerin siber saldırılara karşı korunması kadar önemli olan diğer konu da saldırı ve saldırgan profili hakkında yorum yapabilmektir [1].

1.1 SCADA VE ENDÜSTRİYEL KONTROL SİSTEMLERİNİ HEDEF ALAN SALDIRILARA GENEL BAKIŞ

Uluslararası Otomasyon Topluluğu (ISA), otomasyon ve kontrol sistemleri tanımını "endüstriyel bir sürecin güvenli ve güvenilir çalışmasını etkileyebilecek veya etkileyebilecek personel, donanım ve yazılım koleksiyonu" olarak yapmaktadır [2]. Son yıllarda, çeşitli kötü amaçlı yazılım kampanyaları endüstriyel kontrol sistemlerini kesintiye uğratmayı hedef almıştır. Bu kampanyalar siber casusluk olarak kategorize edilebilir, çok hedeflidir ve bilgi çalmaya yöneliktir. Siber suç gibi diğer siber saldırı türleri de öncelikle hedeften veri toplanmasını gerektirebilecek veya gerektirmeyecek finansal kazançla ilgilenmektedir. Siber casusluk saldırılarında kullanılan kötü amaçlı yazılımlar, Gelişmiş Sürekli Tehdit (APT) yöntemlerini de kullanmaktadır. Yani bu saldırılarda kötü amaçlı yazılım keşfedilmeden önce uzun bir süre boyunca bilgi toplamak için kurbanın sisteminde yerleşik olacak şekilde tasarlanmaktadır [3]. Otomatik endüstriyel işlemlere karşı yürütülen APT, hem bilgi teknolojisi hem de kullanılan belirli endüstriyel sistemler hakkında uzman bilgisi gerektirmektedir.

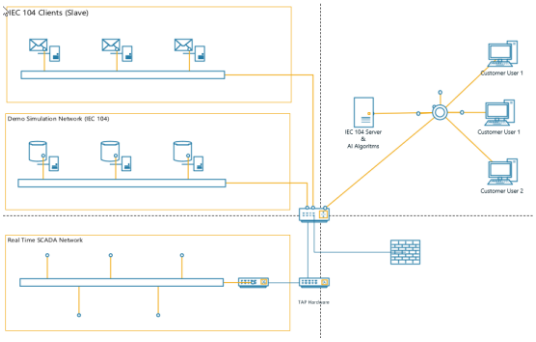
1.2 SCADA GÜVENLİĞİ ALANINDA YAPILAN AR-GE ÇALIŞMALARI

Mevcut tekniklerde farklı siber güvenlik firmalarının sunduğu geleneksel çözümler incelendiğinde; genel anlamda kural ve log tabanlı bir çözüm sundukları ancak SCADA ağını saldırılardan korumak için yeterli çözümü sağlamadıkları gözlemlenmiştir. geleneksel güvenlik çözümleri yetersiz kalmaktadır. Teknikte iç ağdaki ajan yazılımlar ile sistemin performansını ve güvenliğini sağlamayı amaçlayan, belirlenen değerlere göre sistemin alarm vermesini sağlayan bir çözümden bahsetmektedir. Teknikte genel ağ takibi, belli anomalilerde ve birtakım güvenlik ihlallerinde kullanıcıları alarm sistemi ile haberdar eden altyapı kullanılmaktadır [4]. IP, port gibi standart belli konfigürasyonlar ile belli erişimleri engelleme yöntemi kullanılmaktadır [5].

Bu yöntem geleneksel güvenlik duvarları ile benzerlik göstermektedir. SCADA networkünde ilk gözlenen değerlere göre anomali tespiti yapan yöntem de kullanılmaktadır[6]. Endüstriyel kontrol sistemleri için durum bilincini ve siber güvenlik yönetimini geliştirmek amacıyla SCADA sistemlerinin siber güvenlik yönetimi için bir usul ve sistem temin edilmiştir. Endüstriyel kontrol sistemi için güvenlikle ilgili verileri toplamak üzere bir SCADA' ya bir merkezi Sistem Güvenlik Yöneticisi entegre edilmiştir. Bir entegre kumanda ve kontrol kullanıcı ara yüzü, güvenlikle ilgili verileri, bir sistem güvenlik seviyesini görüntülemektedir. Kullanıcının, endüstriyel kontrol sisteminin sistem güvenlik ayarlarını, toplanmış güvenlikle ilgili veriler bazında değiştirmesine olanak sağlamak için bir ara yüz sunar. Güvenlik seviyesi bazında sistem ara yüzlerinin ve sistem erişim yollarının kullanımını sınırlamak üzere SCADA' nın işletim durumundaki değişiklikleri yönetir [4]. SCADA networkünde izleme ile tüm cihazların whitelistlerini (Device Name, MAC Address, IP Address, Port Number, Device State) oluşturularak whitelistleri database üzerinde saklama ve sürekli olarak SCADA konfigürasyonuna göre harici saldırı tespit ve engelleme sisteminin konfigürasyonunu güncelleyen saldırı tespit yöntemi mevcutta kullanılmaktadır [7].

2. MATERYAL METOT

Yapılan çalışmada öncelikle yapay zeka modelinin anomal trafiği ayırt edebilmesi için trafo merkezi simülatör ortamından normal çalışma alanındaki network paketleri toplanması sonrası anormal veri paketi oluşturulması için çalışmalar yapılmıştır. Sanallaştırma altyapısı Şekil 1'deki gibi oluşturulmuştur.

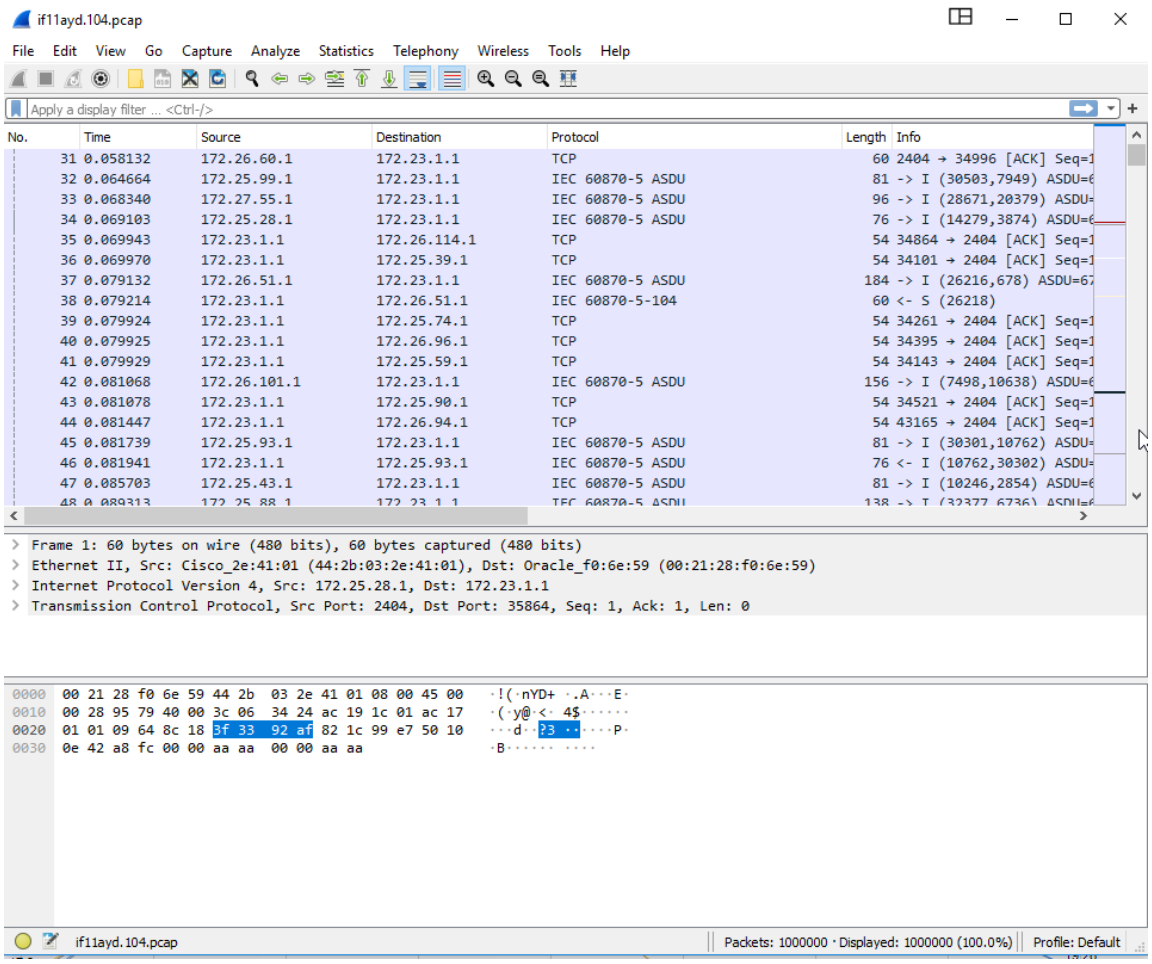


Şekil 1. Proje Sanallaştırma Mimari Tasarımı

Mevcut hiyerarşik SCADA sistemleri, dahili güvenlik mekanizmasına sahip olmayan iletişim protokollerini kullanır. Çalışmada kullanılan haberleşme protokolleri, EC 60870-5-101 ve IEC 60870-5-104 protokolleri için zayıf noktalar; IEC 60870-5-101 protokolü ve IEC 60870-5-104 protokolünde tamamen bağımlı bir veri bütünlüğü için Checksum, veri iletimi için IEC 60870-5- 101 ve IEC 60870-5-104 protokollerinin kullanıldığı MTU ve RTU arasındaki iletişim olarak tespit edilmiştir. IEC 60870-5-104 protokolu sayesinde, trafo merkezleri, üretim santralleri ve diğer merkezlerden çeşitli veriler toplanır. Bu mimaride, sahadaki diğer cihazlar, farklı protokoller kullanarak verileri okurlar. Bu veriler daha sonra IEC 60870-5-104 protokolüne dönüştürülür. Hatalı sonuçlar oluşturmaması için sadece IEC 104 paketleri gönderilmesi için filtreleme yapılmıştır. Gerçek zamanlı toplanan verileri TCP edit uygulaması ile destination adresleri yeniden düzenlenerek emülatör ortamında verilerin doğru oluşturulması sağlanmıştır. Bu amaçla if11ayd.104.pcap dosyası

oluşturulmuştur. Şekil 2'de filtrelenmiş IEC104 paketleri yer almaktadır.

kaldırılarak modele öğretme işleminin sağlanmasıdır. Temizleme işlemi



Şekil 2. Wireshark ile Filtrelenmiş IEC104 Paketlerinin Ekran Görüntüsü

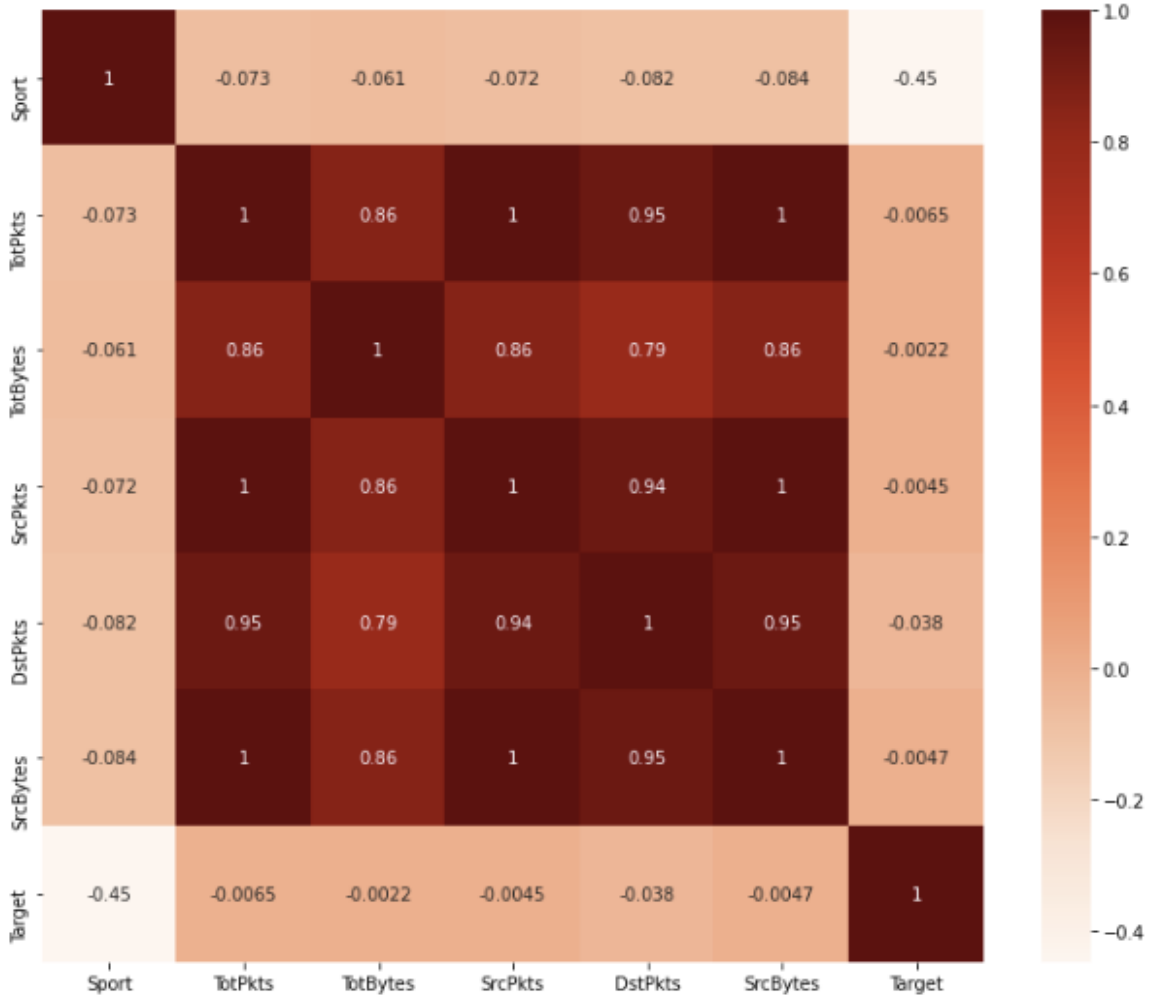
Özellik çıkarımı, ön işleme ve veri hazırlama fonksiyonları hazır hale getirilmiştir. Burada anomali verileri iletilmediği için örnek veri seti üzerinde çalışmalar gerçekleştirilmiştir. Veri TCP protokolüne göre bir ağdan toplanan verileri temsil etmektedir. Çalışılacak test ortamından toplanan 25 öz niteliğe sahip veri, yapılan çalışmalar kapsamında 7 öz niteliğe indirgenmiş temiz bir veri haline getirilmiştir. Bu temizleme işlemlerinde; kayıp verilerin temizlenmesi, bozuk verilerin temizlenmesi ve aykırı değerlerin temizlenmesi işlemleri kullanılmıştır. Bunun sebebi ise öğrenilen verinin öncelikle bu bozucu etkilerinin sonucunda ortaya çıkan öznitelikler içinde bulunan atak sınıfları Tablo 1'de yer almaktadır.

Tablo 1. Öznitelikler İçindeki Atak Sınıfları

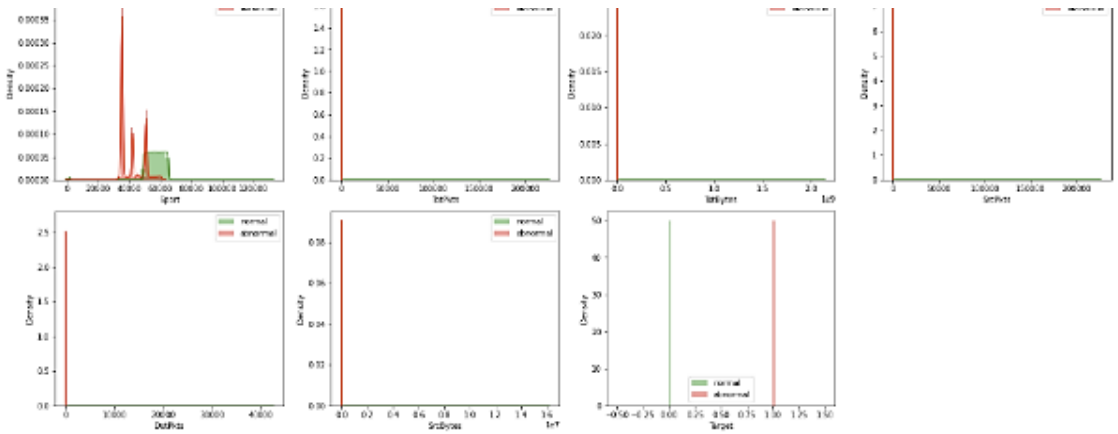
Port Scanner Attack:	%0.0003
Adress Scan Attack:	%0.0075
Device Idendification Attack:	%0.0001
Exploit Attack:	%1.1312

Atak sınıflarının toplam veri içindeki dağılımı %6.07'dir.
Verilerin öz nitelikler arasındaki ve target arasındaki korelasyonu Şekil 4.1'de gösterilmektedir.

Her bir özniteliğin(feature) kendi içindeki dağılımlarının gözlemlenmesinin sebebi özel, rastsal dağılımlı verileri iyi analiz ederek özniteliklere müdahale etmektir.



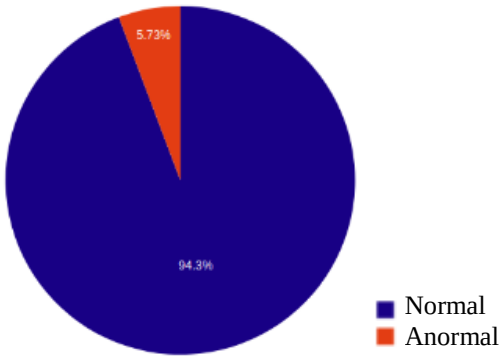
Şekil 4.1. Verilerin Öz Nitelikler Arası ve Target Arasındaki Korelasyonunun Gözlemlenmesi



Şekil 4.2. Veri İçerisindeki Normal Ve Anormal Verilerin Dağılımı

Her bir öz nitelik değerinin "Target" parametresine göre kendi içindeki dağımları Şekil 4.2'de gösterilmektedir.

Target yani verilerin normal ve anormal olduğunu temsil eden etiket verisi içerisinde veri normal ve anormal olarak iki ayrı sınıf taşımaktadır. Normal ve anormal verilerin dağılımları Şekil 4.3'de gösterilmiştir.

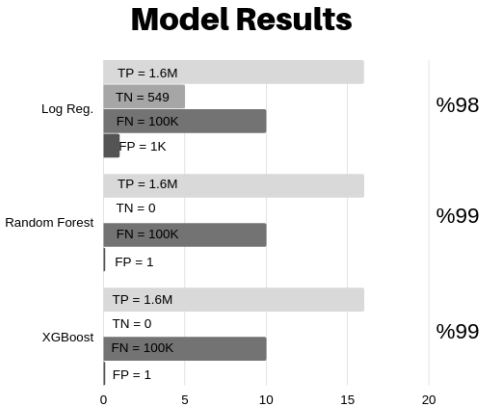


Şekil 4.3. Tek Bir Özniteliğin İçindeki Verilerin Normal ve Anormal Durumlara Göre Dağılımı

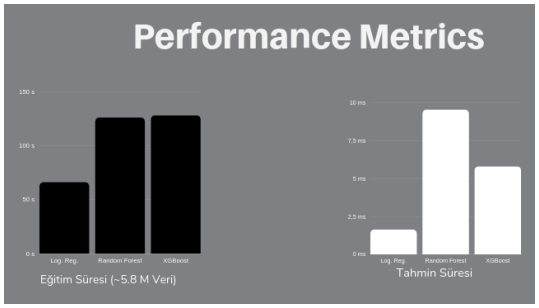
Özniteliklerin belirlenmesi çalışması yapılırken ilk önce NSL-KDD public veri seti kullanılmıştır. Çalışma kapsamında toplamda 22540 farklı train datası kullanılarak model eğitilmiştir.

NSL-KDD veri setinde "normal" ve "anormal" olarak etiketlenen verilerden anomali tespiti yapılması için gereken model oluşturulmuştur. Anormal veri setine dahil edilen veriler 4 farklı kategoride "DoS", "Probe", "R2L", ve "U2R" etiketlenerek alt kategoriler oluşturulmuştur.

Modeller üzerinde K-fold cross validation yapılarak, NSL-KDD veri seti üzerinde Logistic Regression, Support Vector Machine(SVM), KNN algoritmalarının anomali tespitindeki başarı değerleri hesaplanmış aşağıdaki gibi sonuçlar elde edilmiştir. Yapay zeka çalışmalarında Python dili tercih edilmiştir. Bu kapsamda çalışılan makine öğrenmesi modelleri: K-Means, KNN, Logistic Regression, Mean Shift, One Class SVM, Random Forest, SVM. Bu model parametre optimizasyonu ve model çalışmalarının sonucunda en iyi 3 modelin başarı sonuçlarına Şekil 5.1 ve Şekil 5.2'de yer verilmiştir.



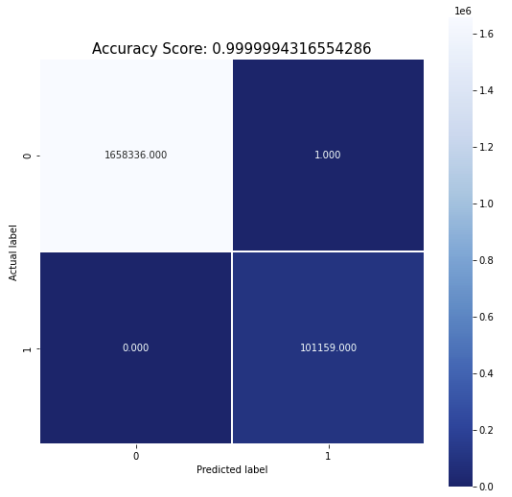
Şekil 5.1. Modellerin Başarı Sonuçları



Şekil 5.2. Modellerin Performans Metrikleri

Performans ve doğruluk oranı olarak en başarılı XGBoost modelinin sınıflandırma raporu ve confusion matrix sonuçlarına Şekil 5.3'de yer verilmiştir.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1658337
1	1.00	1.00	1.00	101159
accuracy			1.00	1759496
macro avg	1.00	1.00	1.00	1759496
weighted avg	1.00	1.00	1.00	1759496



Şekil 5.3. XGBoost Modelinin Matrix Sonuçları

Yapay zekanın eğitiminde unsupervised learning metodu kullanılmaktadır. Bu yöntem ile sisteme ek bir sınıflandırıcı eklenmeden eğitim yapılabilmektedir. Yapay zekanın docker üzerinde kurulumu yapılırken sağlanan bir dosya ile eğitimin ilk aşaması başlamaktadır. Sisteme gönderilen her paket incelendikten sonra sonucu kullanıcıya bildirilir ve eğer sistem ilk defa bu içeriğe sahip bir paket ile karşılaşırsa sonucu değerlendirilmek üzere sisteme kaydeder. Bu işlemleri kullanıcıya açan wrapper yapısıyla kullanıcı yapay zekaya sorgulama istekleri atabilir ve yeni bilgileri sisteme gönderebilir. Aynı zamanda yapay zekanın sağlık durumuna ait bilgilere erişebilir.

3. DEĞERLENDİRME, ANALİZ VE SONUÇLAR

Bu çalışmada, elektrik üretim, iletim ve dağıtım tesisleri, iletişim ve haberleşme, ulaşım, gaz üretim gibi kritik altyapı olarak nitelendirilebilecek farklı tesislerde denetim amaçlı kullanılan SCADA sistemleri ve bu sistemlerin güvenliğini artırmaya yönelik bir yazılım geliştirilmiştir. SCADA sistemlerinin haberleşmesinde en sık kullanılan haberleşme protokollerinden IEC-104 protokolünün kimlik doğrulama zafiyeti ve veri iletimi sırasında haberleşme olmaması zafiyeti incelenmiş, bu zafiyetler istismar edilerek veriler manipüle edilmiştir. Gerçekleştirilen saldırıda bir Linux işletim sistemiyle önemli işletmelerin otomasyon sistemlerini kontrol edebilecek bir SCADA Simülasyonu Test Panosu cihazının üzerindeki iki yönlü switch ile kontrol edilerek simülatördeki adreslerdeki değerlerin değiştirilebileceği ve basit bir sistemle karmaşık sistemlere zarar verilebileceği gösterilmiştir. Bu sebeple endüstriyel otomasyon sistemlerine yönelik güvenlik çözümleri üretilirken basit bir saldırı senaryosundan karmaşık bir saldırı senaryosuna kadar çok aşamalı saldırı tekniklerinin dikkate alınması gerektiği ortaya konulmuştur.
IEC-104 protokolünde tespit edilen güvenlik zafiyeti sonrasında, bu açığa yönelik saldırıları hafifletmek veya engellemek için ara kontrol katmanı üzerinde bir Python kodu geliştirilmiş ve bir kontrol mekanizması oluşturulmuştur. Geliştirilen bu güvenlik kontrol mekanizması ile EKS(Extended Karplus-Strong) algoritmasında en sık kullanılan haberleşme protokollerinden birisi olan IEC-104 protokolüne yönelik dışarıdan yetkisiz bir kullanıcı tarafından müdahalede bulunulması engellenmiş ve aynı zamanda kod içerisindeki kontrol fonksiyonuyla EKS ağı içerisinde kötü niyetli bir kullanıcı tarafından veya yetkili personel tarafından sisteme girilen değerler denetim altına alınmıştır. Sunulan çalışma gerçek sistemlerle entegre çalıştırılması hem iç ağdan hem de dış ağdan gelebilecek siber saldırıların etkilerini hafifletmektedir.

Yapay zeka sistemini eğitmek için bu çalışmada kullanılan veri kümesinde, mevcut SCADA protokollerinden IEC-104 için farklı kategorilerde çeşitli ataklar gerçekleştirilmiştir. Bu veri kümesi kullanılarak SCADA veya EKS'lere yönelik atak yapılan ve atak yapılmayan durumlar farklı niteliklere göre değerlendirilip analiz edilmektedir. Bunun için Logistic Regression, Support Vector Machine(SVM), KNN, XGBoost algoritmaları veri madenciliği yapılarak kullanılmaktadır. En doğru sınıflandırma oranına XGBoost Algoritması sahip olduğu için bu algoritma sonuçları ayrıntılı olarak incelenmektedir. Siber terör ataklarının davranış analizleri ve atak tespiti üzerinde çalışılarak ilgili alana katkı sağlanması amaçlanmıştır. SCADA atak tespiti sistemi araştırmalarında kullanılmak üzere yeni veri kümeleri sağlanmalıdır. Bunun için de bu tür saldırıların SCADA sistemlerine karşı uygulanması gereklidir. Bu konuda yapılacak çalışmalar SCADA güvenliğinin sağlanması için gereklidir. Gelecekteki çalışmalar için farklı veri kümeleri elde edilerek siber terör atak davranış analizleri daha kapsamlı olarak gerçekleştirilecektir. Ayrıca elde edilen sonuçlar birbirlerine göre kıyaslanmış ve böylece çalışma alanı daha da genişletilmiştir.

Özellikle ülkemizde elektrik üretim, iletim ve dağıtımına ilişkin kontrol sistemlerinin haberleşme protokollerinin yeniden gözden geçirilmesi ve siber güvenlik açısından ele alınması gerekmektedir. Kritik altyapılar içerisinde en önemli sistemlerden birisi olan SCADA sistemlerinin güvenliği hayati derecede öneme sahiptir. Bu nedenle, sunulan çalışmanın kritik bir altyapı olan SCADA sistemlerinin güvenliğine katkı sağlayacağı değerlendirilmektedir.

4. TEŞEKKÜR

Bu çalışma EPDK tarafından 15.03.2019 tarihli 01/19/03-2 numaralı komisyon kararı ile desteklenen ve ADM EDAŞ ve GDZ EDAŞ firmalarının birlikte gerçekleştirdikleri "Siber Güvenlik Çalışmalarında Yapay Zeka Metodlarının Kullanılması için Trafo Merkezi Emülatörü Geliştirme" Ar-Ge projesi kapsamında desteklenmiştir.

5. KAYNAKLAR

[1]https://www.dell.com/downloads/global/solutions/2014_DSAT_Report_Final.pdf

[2] GICSP, E. H., Assante, M., & Conway, T. (2014). An abbreviated history of automation & industrial controls systems and cybersecurity. SANS Institute, Tech. Rep.

[3] Wangen, G. (2015). The role of malware in reported cyber espionage: a review of the impact and mechanism. Information, 6(2), 183-211.

[4] Skare P. M. (2009). Method and system for cyber security management of industrial control systems U.S. Patent No. 2007/294369

[5] Naedele M., Biderbost O. (2004). Network security system E. Patent No. 1544707

[6] Falavigna L., Bima C. (2007) Industrial plant security apparatus and monitoring method of security of an industrial plant E. Patent No. 1881388

[7] AROV M., OCHMAN R., Cohen M. (2016) System and method for detecting a cyber-attack at scada/ics managed plants W.O. Patent No. 2017/090045