Elektrikli Araç Şarj İstasyonlarının Siber Güvenliği: CIC-EVSE 2024 Veri Kümesi ile IDS Sistemlerinin Performans Analizi

Havvanur Bozömeroğlu
Bilgisayar Mühendisliği
İstanbul Üniversitesi-Cerrahpaşa
İstanbul, Türkiye
orcid.org/0009-0007-1162-5443

Zeynep Gürkaş-Aydın
Bilgisayar Mühendisliği
İstanbul Üniversitesi-Cerrahpaşa
İstanbul, Türkiye
orcid.org/0000-0002-4125-0589

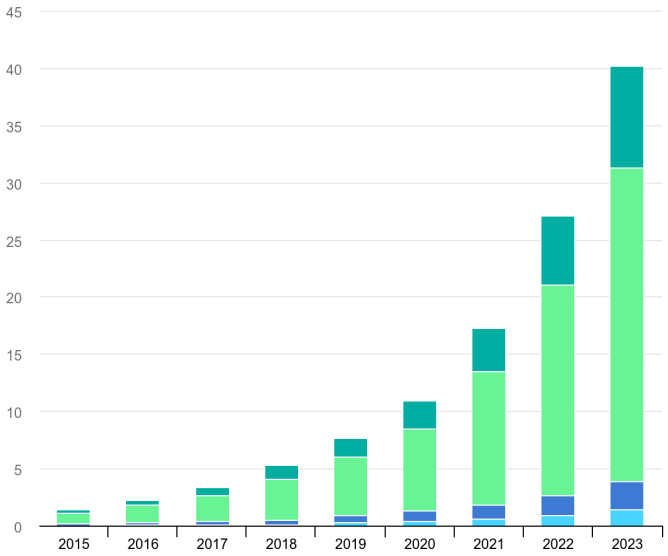
***Özet*—Elektrikli araçlar (EV'ler), fosil yakıtlara olan bağımlılığı azaltma ve çevresel etkileri minimize etme potansiyeliyle giderek daha fazla önem kazanmaktadır. Bu çalışmada, elektrikli araç şarj istasyonlarının siber güvenliğini değerlendirmek amacıyla Saldırı Tespit Sistemi (STS) sistemlerinin etkinliği incelenmiştir. Artan EV sayısı ile birlikte, şarj istasyonlarının siber tehditlere karşı korunması kritik hale gelmiştir. Bu bağlamda, CIC-EVSE 2024 veri kümesi kullanılarak çeşitli makine öğrenimi modelleri (SVM, Logistic Regression, Decision Tree, KNN, Adaboost, MLP, Naive Bayes, Random Forest) üzerinde deneyler yapılmıştır. Bu çalışmada, veri temizleme, aşırı örnekleme, normalizasyon, özellik seçimi ve eğitim-test ayrımı gibi ön işleme adımlarını içermektedir. Elde edilen sonuçlar, STS sistemlerinin siber tehditleri tespit etme ve önleme konusundaki başarısını göstermektedir. Modellerin doğruluk, F1 skoru, hassasiyet ve geri çağrım gibi metriklerle değerlendirilmesi, hangi modelin en iyi performansı sağladığını belirlemekte önemli bir rol oynamıştır. Bu araştırma, elektrikli araç şarj altyapısının güvenliğini artırmaya yönelik önemli bir adım olarak görülmekte olup, sürdürülebilir ulaşım çözümlerinin yaygınlaşmasına katkıda bulunmayı amaçlamaktadır.**

***Anahtar Kelimeler*—Elektrikli Araçlar (EV), Şarj İstasyonları, Siber Güvenlik, Saldırı Tespit Sistemi, CIC-EVSE 2024, Makine Öğrenimi**

I. Giriş

Elektrikli araçlar (EV'ler), fosil yakıtlara olan bağımlılığı azaltma ve çevresel etkileri minimize etme potansiyeliyle gün geçtikçe daha fazla önem kazanıyor[1]. Dünya genelinde artan çevresel kaygılar ve enerji sürdürülebilirliği gereksinimleri, EV'leri geleneksel içten yanmalı motorlu araçlara kıyasla daha cazip hale getiriyor. EV'ler, daha az karbon emisyonu salarak çevre dostu bir ulaşım çözümü sunuyor ve bu nedenle hükümetler tarafından çeşitli teşviklerle destekleniyor. Ancak, EV'lerin geniş çapta benimsenmesi, sağlam ve yaygın bir şarj altyapısının varlığına bağlı. Elektrikli araç kullanıcılarının günlük ihtiyaçlarını karşılayacak hızlı ve güvenilir bir şarj ağı, EV'lerin yaygın kullanımının anahtarıdır[2]. Şekil. 1'de gösterilen IEA tarafından hazırlanan grafikte[3] de 2015 ile 2023 yılları arasındaki şarj noktalarının gittikçe artmakta olduğu görülmektedir.

EV sayısındaki artış, şarj istasyonlarının siber tehditlere maruz kalma riskini de beraberinde getiriyor[4]. Şarj istasyonları, internete bağlı cihazlar olarak siber saldırıların hedefi olabilir. Bilgisayar korsanları, bu istasyonlara sızarak işlevselliğini bozabilir ve kullanıcıların güvenliğini riske atabilir. Özellikle, şarj istasyonlarına yönelik ağ ve ana makine tabanlı saldırılar, EV kullanıcılarının güvenliğini ciddi şekilde tehdit edebilir ve elektrikli araçların geniş çapta benimsenmesini olumsuz etkileyebilir[5].



Şekil 1. Güç değeri ve türe göre kamu ve özel kurulu hafif ticari araç şarj noktaları, 2015-2023[3].

Elektrikli araç şarj altyapısının güvenliğini sağlamak için Saldırı Tespit Sistemi(STS) gibi güvenlik önlemleri geliştiriliyor[6]. Saldırı tespit sistemleri, ağ trafiğini ve sistem aktivitelerini sürekli izleyerek potansiyel tehditleri tespit eder ve gerekli müdahaleyi yapar. Şarj istasyonlarında STS sistemlerinin kullanılması, siber tehditlere karşı daha güçlü bir savunma mekanizması sağlayarak bu altyapının güvenliğini artırabilir.

Elektrikli araçlar, çevre dostu ve sürdürülebilir ulaşım çözümleri olarak büyük bir potansiyele sahiptir[7]. Ancak,

bu potansiyelin gerçekleştirilmesi güçlü ve güvenli bir şarj altyapısına bağlıdır. Elektrikli araç şarj istasyonlarının siber güvenliğini sağlamak, bu altyapının güvenilir bir şekilde çalışması için kritik öneme sahiptir[8]. Bu çalışma, elektrikli araç şarj istasyonlarının siber güvenliğini değerlendirmek ve STS sistemlerinin etkinliğini incelemeyi hedeflemektedir. CIC-EVSE 2024 veri kümesi kullanılarak gerçekleştirilecek STS çalışması, şarj istasyonlarının güvenliğini artırmayı ve siber tehditlere karşı etkili bir savunma sağlamayı amaçlamaktadır. Böylece, sürdürülebilir ulaşım çözümlerinin benimsenmesi teşvik edilecektir.

Bu çalışmada, elektrikli araç şarj istasyonlarının siber güvenliği üzerine kapsamlı bir araştırma sunulmaktadır. İlgili çalışmaların incelenmesi, kullanılan veri kümesinin tanıtımı, ön işleme adımları ve Destek Vektör Makineleri (SVM), Lojistik Regresyon, Karar Ağacı, K-En Yakın Komşu (KNN), Adaboost, Çok Katmanlı Algılayıcı (MLP), Naive Bayes ve Rastgele Orman ile yapılan eğitimlerin sonuçları analiz edilecektir.

II. Literatür Taraması

Elektrikli Araçların artan benimsenmesi ve buna karşılık gelen sağlam Elektrikli Araç Şarj İstasyonlarına duyulan ihtiyaç, bu sistemlerin güvenliği üzerine önemli araştırmaların yapılmasını teşvik etmiştir. Nasr ve diğerleri, tersine mühendislik ve sızma testi tekniklerini kullanarak yaygın olarak kullanılan EV Şarj İstasyonu Yönetim Sistemlerinin kapsamlı bir güvenlik ve zafiyet analizini gerçekleştirmiş ve uzaktan siber saldırıları kolaylaştırabilecek ve şebeke kesintilerine neden olabilecek çeşitli zafiyetleri ortaya çıkarmıştır[9]. Benzer şekilde, ElKashlan ve diğerleri, EVCS ağlarının kararlılığını ve güvenliğini artırmak için gerçek bir IoT veri kümesi kullanarak Dağıtılmış Hizmet Engelleme (DDoS) saldırılarını tespit etmeye odaklanan makine öğrenimi tabanlı bir Saldırı Tespit Sistemi önermiştir[10]. Nesnelerin İnterneti'nin yaygınlaşması özellikle IoT tabanlı botnet saldırıları ile ilgili önemli güvenlik zorluklarını da beraberinde getirmiştir. Meidan ve diğerleri, derin otoenkoderler kullanarak ele geçirilmiş IoT cihazlarından anormal ağ trafiğini tespit eden ağ tabanlı bir anomali tespit yöntemi olan N-BaIoT'u geliştirmiştir[11]. Bezerra ve diğerleri ile Stoian, sırasıyla gerçek IoT cihaz verileri ve kötü amaçlı yazılım analizi üzerine odaklanarak saldırı tespiti araştırmaları için veri kümeleri oluşturarak bu alana katkıda bulunmuşlardır[12][13]. IoT güvenliği için makine öğrenimi uygulamaları alanında, Haji ve Ameen, IoT ağlarında saldırı ve anomali tespiti için çeşitli makine öğrenimi tekniklerini inceleyerek, Rastgele Orman ve Destek Vektör Makineleri gibi algoritmaların tehditleri belirlemedeki etkinliğini vurgulamıştır[14]. Chung ve diğerleri tarafından önerilen çerçeve, çok değişkenli zaman serisi segmentasyonu ve en yakın komşu sınıflayıcılarını kullanarak EV şarj ağlarında anomali tespiti sağlar[15]. Gerçekçi botnet veri kümelerinin geliştirilmesi de hayati önem taşımıştır. Koroniotis ve diğerleri, meşru ve simüle edilmiş IoT ağ trafiğini yakalayan ve ağ adli analiz sistemlerinin tasarımı ve doğrulamasına yardımcı olan Bot-IoT veri kümesini tanıtmıştır[16]. Guerra-Manzanares ve diğerleri ise botnet kötü amaçlı yazılım aşamalarından veri içeren MedBIoT veri kümesini oluşturarak makine öğrenimi tabanlı saldırı tespit sistemlerinin geliştirilmesine destek sağlamıştır[17]. Ayrıca, Basnet ve Ali'nin çalışması, EVCS için derin öğrenme tabanlı STS üzerine odaklanmış, DoS saldırılarını tespit etmede yüksek doğruluk sağlamak için derin sinir ağları ve uzun kısa vadeli hafıza algoritmalarını kullanmıştır[18]. Vinod Miskin ve diğerleri, CICIDS 2017 veri kümesini kullanarak EV altyapısı için güvenilir bir STS geliştirmek amacıyla çeşitli makine öğrenimi ve derin öğrenme modellerini karşılaştırarak, saldırı tespitinde çeşitli algoritmaların performansını vurgulamıştır[19].

Genel olarak, bu çalışmalar, EVCS ve IoT ağlarını gelişen siber tehditlerden korumak için ileri güvenlik önlemleri ve sağlam veri kümelerinin kritik önemini vurgulamakta, makine öğrenimi ve derin öğrenme tekniklerinden yararlanarak tespit ve yanıt yeteneklerini artırmayı hedeflemektedir.

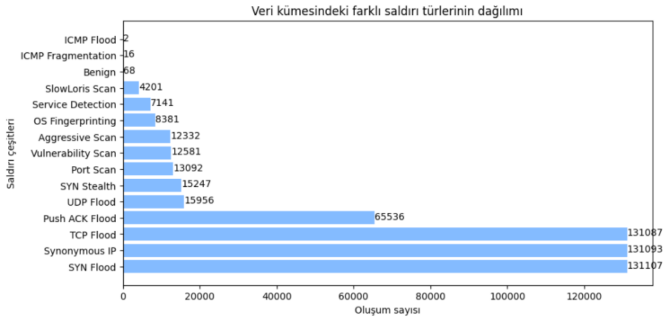
III. Veri Kümesi

Bu tür çalışmalarda veri kümesi, araştırmanın doğruluğu ve güvenilirliği için kritiktir. Elektrikli araç şarj istasyonlarının güvenliği gibi konularda, kapsamlı ve çeşitli veri kümelerine ihtiyaç vardır. Bu veri kümeleri, gerçek dünya senaryolarını simüle etmeye ve farklı saldırı ve normal durumları analiz etmeye olanak tanır. Geniş gözlem sayısı ve çok boyutlu yapı, modellerin eğitim ve test sürecinde yüksek doğruluk sağlar. Sonuç olarak, veri kümeleri bu tür çalışmaların başarısı için vazgeçilmezdir.

Bu çalışmada, elektrikli araç şarj istasyonlarının güvenliğini artırmak amacıyla oluşturulmuş olan CIC EV Charger Attack Dataset 2024 (CICEVSE2024) kullanılmıştır. Bu çok boyutlu veri kümesi, elektrikli araç şarj istasyonlarının (EVSE) normal işleyişi ve saldırı koşullarındaki davranışlarını anlamak için tasarlanmıştır[20]. Veri kümesi, güç tüketimi, ağ trafiği ve HPC ve çekirdek olayları gibi çeşitli veri kaynaklarını içermektedir. Çalışmada özellikle EVSE-A ve EVSE-B'nin Ağ Trafiği bölümü kullanılmıştır. Bu bölüm, hem benign hem de saldırı senaryolarını içeren ağ trafiği verilerini .pcap dosyaları olarak sunmaktadır. Veriler, NFStream Python kütüphanesi kullanılarak .pcap dosyalarından çıkarılmış ve belirli trafik akış örnekleri elde edilmiştir. Ağ trafiği verileri, güvenlik tehditlerinin tespit edilmesi, analiz edilmesi ve sınıflandırılması için makine öğrenimi algoritmaları ile incelenmiştir.

Çalışmamızda özellikle EVSE-A ve EVSE-B'nin ağ trafiği verileri kullanılmıştır. Veri kümesi toplam 547,840 gözlem ve 87 değişken içermektedir. Veri kümelerindeki gözlemler, çeşitli saldırı türlerini ve normal durumları kapsamaktadır. Aşağıda, veri kümelerindeki bazı önemli saldırı türleri ve bunların gözlem sayıları Şekil 2'de gösterilmiştir.

Bu çalışmada CICEVSE2024 veri kümesini kullanmamızın birkaç önemli nedeni vardır. Öncelikle, veri kümesinin yüksek gözlem sayısı (547,840 gözlem) geniş kapsamlı analizler yapmamıza olanak tanır ve istatistiksel anlamlılık sağlar. Ayrıca, veri kümesi, çeşitli saldırı türlerinin farklı

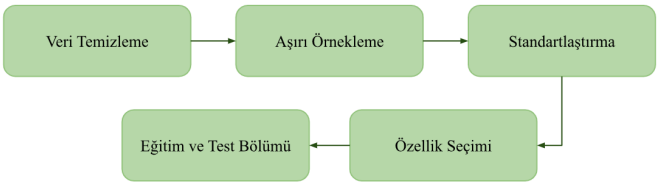


Şekil 2. CICEVSE2024 saldırı çeşitleri ve gözlem sayırları.

dağılımlarını içermektedir. Bu çeşitlilik, güvenlik tehditlerinin farklı yönlerini anlamak ve etkili savunma stratejileri geliştirmek için önemli bir kaynak sağlar. Bu nedenlerle, CICEVSE2024 veri kümesi, elektrikli araç şarj istasyonlarının güvenliği üzerine yaptığımız çalışmada ideal bir veri kaynağıdır.

A. *Veri Ön İşleme*

Veri işleme adımı, modelin eğitim ve test süreçlerinin sağlıklı bir şekilde gerçekleşebilmesi için büyük önem taşır. Bu çalışmada kullanılan veri kümesi, çeşitli ön işleme adımlarından geçirilmiştir. Bu ön işleme adımları Şekil 3'de gösterilmiştir.



Şekil 3. Ön işleme adımları.

1) Veri Temizleme: Veri temizleme işlemi, veri kümesindeki gereksiz veya eksik bilgilerin kaldırılmasını içerir. İlk olarak, modelin eğitimi ve performansı için gereksiz olan `"requested_server_name"`, `"client_fingerprint"`, `"server_fingerprint"`, `"user_agent"`, ve `"content_type"` sütunları veri kümesinden çıkarılmıştır. Bu sütunlar, veri analizinde ve model oluşturma sürecinde anlamlı bir katkı sağlamadığı için kaldırılmıştır. Ardından, çeşitli sütunlardaki kayan nokta değerleri yuvarlanarak tamsayıya dönüştürülmüştür. Bu işlem, verilerin daha düzgün bir şekilde işlenmesi ve analiz edilmesi için gereklidir. Özellikle `"bidirectional_mean_ps"`, `"bidirectional_stddev_ps"`, `"src2dst_mean_ps"`, `"src2dst_stddev_ps"`, `"dst2src_mean_ps"`, `"dst2src_stddev_ps"`, `"bidirectional_mean_piat_ms"`, `"bidirectional_stddev_piat_ms"`, `"src2dst_mean_piat_ms"`, `"src2dst_stddev_piat_ms"`, `"dst2src_mean_piat_ms"`, ve `"dst2src_stddev_piat_ms"` sütunları için uygulanmıştır.

Saldırı türlerini ifade eden `"Attack Type"` sütunundaki kategorik değerler, aşağıdaki şekilde numerik değerlere dönüştürülüp `"Attack_Num"` olarak adlandırılmıştır. Bu işlem, modelin saldırı türlerini sayısal olarak işlemesini kolaylaştırmak amacıyla yapılmıştır. Daha sonra, veri kümesinde yeterli gözlem sayısına sahip olmayan `"ICMP Fragmentation"` ve `"ICMP Flood"` saldırı türleri çıkarılarak veri kümesi yeniden düzenlenmiştir. Bu adım, modelin eğitiminde az sayıda gözlemle temsil edilen sınıfların performansı olumsuz etkilemesini önlemek için gereklidir. Son olarak, `"Attack Type"` sütunu veri kümesinden kaldırılmıştır. Belirli kimlik bilgileri, yani `'id'`, `'expiration_id'`, `'src_port'`, `'dst_port'`, `'src_ip'`, `'dst_ip'`, `'src_mac'`, `'dst_mac'`, `'src_oui'`, `'dst_oui'`, `'application_category_name'` ve `'application_name'` sütunları, anonimlik ve veri güvenliği sebepleriyle veri kümesinden çıkarılmıştır. Bu sütunlar, veri gizliliğini korumak ve veri kümesinin anonimleştirilmesini sağlamak için kaldırılmıştır.

2) Aşırı örnekleme: Saldırı ve normal durumlar arasındaki dengesizliği gidermek için SMOTE yöntemi kullanılmıştır. Bu teknik, azınlık sınıflarındaki örnekleri artırarak sınıf dengesini sağlar ve modelin dengesiz veri kümeleri üzerinde daha iyi performans göstermesine yardımcı olur. SMOTE, azınlık sınıflarına ait yeni örnekler oluşturmak için mevcut örnekler arasındaki farkları kullanarak, modelin bu sınıfları daha iyi öğrenmesini sağlar. Bu, modelin dengesiz veri kümeleri üzerinde daha dengeli ve adil bir şekilde öğrenmesini sağlar.

3) Standartlaştırma: Veri kümesindeki özelliklerin farklı ölçeklerde olması, modelin eğitim sürecini olumsuz etkileyebilir. Bu nedenle, tüm özellikler Min-Max ölçeklendirme yöntemi ile [0, 1] aralığına normalleştirilmiştir. Min-Max ölçeklendirme, her bir özelliğin minimum ve maksimum değerlerini kullanarak değerleri belirlenen aralığa çeker. Bu adım, modelin daha hızlı ve etkili bir şekilde öğrenmesini sağlar. Ölçeklendirme, özellikle farklı ölçeklerdeki verilerin aynı seviyeye getirilerek modelin daha tutarlı sonuçlar üretmesini sağlar.

4) Özellik Seçimi: Veri kümelerindeki özelliklerin varyansları hesaplanarak, varyansı belirli bir eşik değerinin (0.01) altında olan özellikler çıkarılmıştır. Varyansı düşük olan özellikler, genellikle modelin öğrenme sürecine katkıda bulunmayan sabit veya az değişen değerlerdir. Bu adım, modelin yalnızca anlamlı ve değişkenlik gösteren özelliklerle eğitilmesini sağlayarak, modelin genel performansını artırmayı hedefler. Seçilen özellikler, modelin daha iyi performans göstermesini sağlamak için dikkatle seçilmiştir. Varyans analizi, modelin sadece önemli özellikleri kullanarak daha etkin bir öğrenme süreci geçirmesini sağlar.

5) Eğitim ve Test Bölümü: Veri kümesi, modelin performansını değerlendirmek için %80 eğitim ve %20 test setlerine bölünmüştür. Eğitim kümesi modelin öğrenmesi, test kümesi ise modelin genelleme yeteneğini değerlendirmek için

kullanılmıştır. Bu adım, overfitting riskini azaltarak modelin yeni verilere karşı performansını test etmeye yardımcı olur. Ayrıca, sınıf etiketleri kategorik hale getirilmiş ve modelin çok sınıflı sınıflandırma problemlerinde kullanılabilmesi sağlanmıştır. Eğitim ve test kümelerinin ayrılması, modelin performansını doğru değerlendirmek için kritiktir.

IV. Deneysel Kurulumlar

A. Donanım ve Ortam Ayarı

Bu çalışmadaki deneyler, yüksek performanslı donanım bileşenleri sağlayan bulut tabanlı bir platform olan Google Colab üzerinde yürütülmüştür. Kullanılan donanım konfigürasyonu; Google Colab tarafından sağlanan Intel Xeon işlemciler, 12GB bellek (RAM), Google Drive entegrasyonu ile sağlanan bulut depolama ve NVIDIA Tesla T4 grafik işlem birimini (GPU) içermektedir. Bu konfigürasyon, deneylerin yüksek hesaplama gücü gerektiren işlemleri etkin bir şekilde gerçekleştirmesini sağlamıştır.

B. Değerlendirme Metrikleri

Çalışmamızda, modellerin performansını değerlendirmek için doğruluk, hassasiyet, geri çağrım ve F1 skoru gibi metrikler kullanılmıştır. Doğruluk, modelin doğru tahmin yüzdesini gösterirken, hassasiyet yanlış pozitif tahminlerin oranını minimize eder. Geri çağrım, modelin pozitif sınıfı ne kadar iyi yakaladığını ölçerken, F1 skoru hassasiyet ve geri çağrımın dengelenmiş bir ortalamasıdır. Bu metriklerin birlikte kullanılması, modelin hem dengeli hem de dengesiz veri kümelerinde ne kadar iyi performans gösterdiğini anlamamıza yardımcı olur. Sonuç olarak, modellerimizin gerçek dünya senaryolarında başarılı olma potansiyelini kapsamlı bir şekilde değerlendirebiliriz.

C. Model Parametreleri

Çeşitli makine öğrenimi modelleri, belirli hiperparametrelerle birlikte kullanılarak farklı veri kümeleri üzerinde performans gösterebilir. SVM modelinde, `'linear'` kernel hiperparametresi kullanılarak verilerin doğrusal olarak ayrılabilir olduğu varsayılır. Karar Ağacı modelinde, `'gini'` kriteri ve `'best'` splitter kullanılarak ağaç oluşturma sürecinde en iyi bölme noktaları belirlenir. Rastgele Orman modelinde, ağaç sayısı `'n_estimators'` 100 olarak belirlenirken, her ağacın oluşturulmasında `'gini'` kriteri kullanılır. KNN modelinde, komşu sayısı `'n_neighbors'` 5 ve algoritma `'auto'` olarak ayarlanır. Lojistik Regresyon modelinde, `'l2'` ceza türü ve `'lbfgs'` çözücü kullanılarak model düzenliliği sağlanır. Naive Bayes modelinde, Gaussian dağılımı varsayılarak herhangi bir özel hiperparametre belirlenmez. MLP modelinde, gizli katmanlar `'hidden_layer_sizes'` 100 nöron içerir, aktivasyon fonksiyonu `'relu'` ve çözücü `'adam'` olarak ayarlanır. AdaBoost modelinde ise, temel öğreniciler `'n_estimators'` 50 ve algoritma `'SAMME.R'` olarak belirlenir. Bu hiperparametreler, modellerin performansını optimize etmek için kritik öneme sahiptir. Bu hiperparametreler Tablo I'de gösterilmiştir.

Tablo I. Modellerin Hiperparametreleri

Model	Hyperparameters
SVM	kernel: 'linear'
Karar Ağacı	criterion: 'gini', splitter: 'best'
Rastgele Orman	n_estimators: 100, criterion: 'gini'
KNN	n_neighbors: 5, algorithm: 'auto'
Lojistik Regresyon	penalty: 'l2', solver: 'lbfgs'
Naive Bayes	None (uses Gaussian distribution)
MLP	hidden_layer_sizes: (100,), activation: 'relu', solver: 'adam'
AdaBoost	n_estimators: 50, algorithm: 'SAMME.R'

V. Bulgular ve Tartışma

Bu çalışmada, elektrikli araç şarj istasyonlarının siber güvenliğini sağlamak için çeşitli makine öğrenimi modelleri kullanılmış ve performansları değerlendirilmiştir. Modellerin doğruluk, F1 skoru, hassasiyet ve geri çağırım metrikleri kullanılarak performansları karşılaştırılmıştır. Tablo II'de her bir modelin performans sonuçları yer almaktadır.

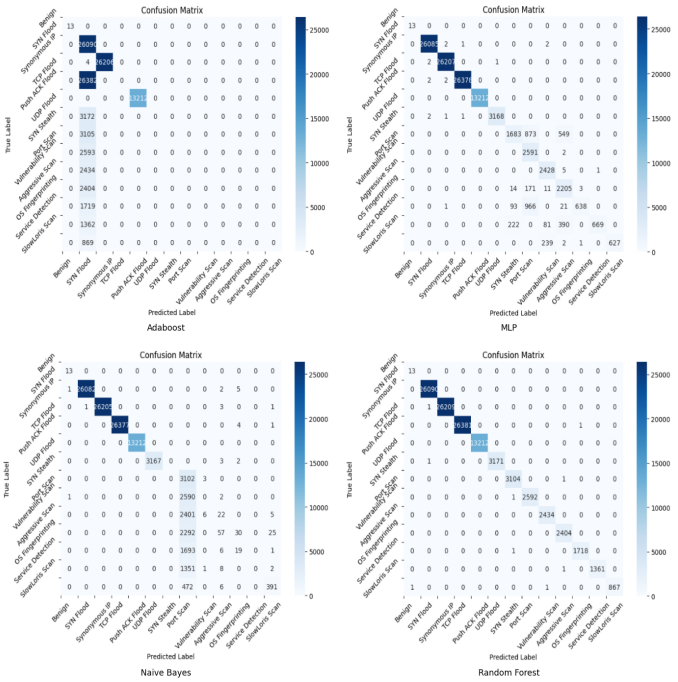
Tablo II. Eğitimlerin Sonuçları

Model	Doğruluk	F1-Skoru	Hassasiyet	Geri Çağırım
SVM	0.9103	0.9011	0.9182	0.9103
Lojistik Regresyon	0.9092	0.8990	0.9101	0.9092
Karar Ağacı	0.9999	0.9999	0.9999	0.9999
KNN	0.9992	0.9992	0.9992	0.9992
Adaboost	0.5980	0.4890	0.4485	0.5980
MLP	0.9665	0.9647	0.9753	0.9665
Naive Bayes	0.8955	0.8813	0.9092	0.8955
Rastgele Orman	0.9999	0.9999	0.9999	0.9999

Bu sonuçlar, Karar Ağacı ve Rastgele Orman modellerinin en yüksek performansı gösterdiğini ortaya koymaktadır. Her iki model de neredeyse %100 doğruluk, F1 skoru, hassasiyet ve geri çağırım oranlarına sahiptir. KNN modeli de oldukça yüksek performans göstermiştir. Buna karşılık, Adaboost modeli, diğer modellere kıyasla düşük performans sergilemiştir. Şekil 4'teki karmaşıklık matrisine göre, SVM modeli genel olarak yüksek doğruluk ve performans göstermiştir. Karmaşıklık matrisleri, modelin gerçek sınıflarla tahmin ettiği sınıfları karşılaştırarak doğru ve yanlış sınıflandırmaları göstermektedir. Her satır, gerçek sınıfı, her sütun ise modelin tahmin ettiği sınıfı temsil eder. Ancak, bazı saldırı türleri arasında, özellikle SYN Flood ve TCP Flood gibi benzer saldırılar arasında, karışıklıklar bulunmaktadır. Bu durum, SVM'nin doğrusal olmayan verileri ayırt etme kabiliyetinin sınırlı olabileceğini göstermektedir. Lojistik Regresyon modeli, SVM ile benzer bir performans sergilemiştir. SYN Flood ve TCP Flood gibi saldırılar arasında karışıklık gözlemlenmiştir. Modelin doğrusal ilişkilere duyarlılığı, bu tür karışıklıklara neden olabilir.

Karar Ağacı modeli, neredeyse mükemmel bir performans sergilemiştir. Karmaşıklık matrisi, modelin tüm sınıfları doğru bir şekilde sınıflandırdığını göstermektedir. Bu matris, her sınıf için doğru tahmin sayısını ve yanlış tahmin sayısını göstererek, modelin hangi sınıfları doğru tahmin ettiğini ve hangi sınıflar arasında karışıklık yaşadığını anlamamıza yardımcı olur. Aynı şekilde KNN modeli de yüksek performans göstermiştir. Ancak, SYN Flood ve TCP Flood gibi saldırılar arasında bazı karışıklıklar bulunmaktadır. KNN modelinin performansı, komşu sayısının optimize edilmesi ile artırılabilir. Bu model,

veri kümesinin yapısına duyarlıdır ve yüksek boyutlu veri kümelerinde performans düşebilir.

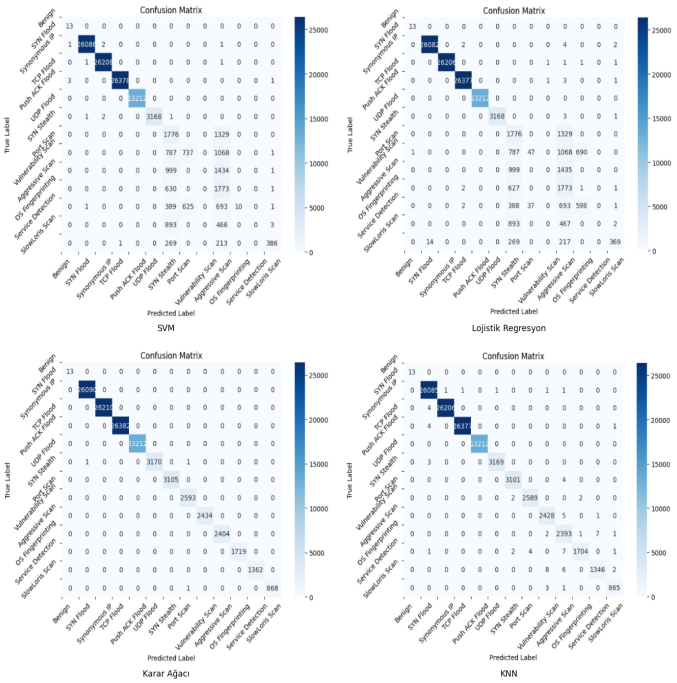


Şekil 4. SVM, Lojistik Regresyon, Karar Ağacı ve KNN modellerinin karmaşıklık matrisleri.

Adaboost modeli, diğer modellere kıyasla daha düşük performans sergilemiştir. SYN Flood, TCP Flood ve Push ACK Flood gibi saldırılar arasında karışıklıklar gözlemlenmiştir. Adaboost, her iterasyonda hataları düzelterek modelin doğruluğunu artıran bir topluluk öğrenme yöntemidir. MLP modeli, yüksek doğruluk oranları ile dikkat çekmektedir. Ancak, SYN Flood ve TCP Flood gibi bazı saldırı türlerinde karışıklıklar gözlemlenmiştir. Modelin performansı, daha fazla veri ile eğitilmesi ve hiperparametre optimizasyonu ile iyileştirilebilir. MLP, karmaşık ve doğrusal olmayan ilişkileri öğrenme yeteneği sayesinde iyi sonuçlar verir.

Naive Bayes modeli, genel olarak yüksek doğruluk oranlarına sahiptir. Ancak, SYN Flood ve TCP Flood gibi sınıflar arasında karışıklık olabilir. Rastgele Ağaç modeli, neredeyse mükemmel bir performans sergilemiştir. Model, tüm sınıfları doğru bir şekilde sınıflandırmış ve yanlış pozitif/negatif oranları neredeyse yok denecek kadar azdır. Topluluk yöntemi sayesinde veri kümesindeki gürültü ve varyansı azaltarak yüksek doğruluk oranlarına ulaşmıştır. Rastgele Ağaç, çok sayıda karar ağacının birleşiminden oluşur ve genelleme yeteneği yüksektir.

Genel olarak, Karar Ağacı ve Rastgele Ağaçlar modelleri en yüksek performansı göstermiştir. Bu modeller, doğru sınıflandırma oranları ile elektrikli araç şarj istasyonlarına yönelik siber saldırıların tespitinde etkili olduğunu kanıtlamıştır. Diğer modellerin performansı da tatmin edici düzeyde olup, özellikle SYN Flood ve TCP Flood gibi benzer saldırılar arasında karışıklıklar gözlemlenmiştir.



Şekil 5. Adaboost, MLP, Naive Bayes, ve Rastgele Orman modellerinin karmaşıklık matrisleri..

Tablo III. Sonuçların Karşılaştırılması

Model	**Doğruluk (HPC ve Çekirdek Olayları) [20]**	**Doğruluk (Ağ Trafiği) (Bu Çalışma)**
Karar Ağacı	0.9373	0.9999
KNN	0.9368	0.9992
Adaboost	0.9376	0.5980
MLP	0.9056	0.9665
Naive Bayes	0.9213	0.8955
Lojistik Regresyon	0.8262	0.9092
Rastgele Orman	0.9374	0.9999
SVM	0.9413	0.9103

Tablo III'de farklı makine öğrenimi modellerinin CICEVSE2024 veri kümesi üzerindeki doğruluk değerlerini karşılaştırmaktadır. Veri kümesi içinde iki ayrı veri kümesi bulunur: HPC ve Çekirdek Olayları[20] ve Ağ Trafiği (bu çalışma). Her iki veri seti üzerinde eğitim alan modellerin performansları karşılaştırılarak, hangi modellerin hangi veri setinde daha iyi performans gösterdiği analiz edilmiştir. Genel olarak, Karar Ağacı, KNN ve Rastgele Orman modelleri ağ trafiği veri setinde çok yüksek doğruluk sağlamıştır. Bu, bu modellerin veri setinin yapısına ve özelliklerine iyi uyum sağladığını gösterir. Adaboost modeli ise bu veri setinde düşük performans göstermiştir, bu da modelin zayıf öğrenicilerinin ve gürültüye duyarlılığının etkisi olabilir. Diğer modeller, veri setine bağlı olarak değişen performanslar göstermiştir, ancak genellikle ağ trafiği veri setinde daha yüksek doğruluk elde edilmiştir. Bu sonuçlar, model seçiminin veri setinin yapısına ve özelliklerine bağlı olarak değişebileceğini ve her modelin her veri setinde en iyi performansı göstermeyebileceğini vurgular.

VI. SONUÇ

Bu çalışma, elektrikli araç şarj istasyonlarının siber güvenliğini artırmayı amaçlayarak, çeşitli makine öğrenimi modellerinin siber saldırıları tespit etme performansını değerlendirmiştir. Hem normal hem de saldırı koşullarında ağ trafiği verilerini içeren kapsamlı CIC-EVSE 2024 veri kümesi kullanılarak, çeşitli saldırı tespit sistemleri modelleri geliştirilmiş ve test edilmiştir.

Deneysel sonuçlarımız, Karar Ağaçları ve Rastgele Orman modellerinin EV şarj istasyonlarına yönelik çeşitli siber saldırıları doğru bir şekilde tespit etmede mükemmel performans sergilediğini göstermektedir. Bu modeller, yüksek doğruluk, hassasiyet, geri çağırma ve F1 skorları ile sağlamlıklarını ve güvenilirliklerini ortaya koymuştur. Bununla birlikte, SVM ve Lojistik Regresyon gibi modeller de iyi performans göstermiş olsalar da, özellikle SYN Flood ve TCP Flood gibi benzer saldırılar arasında farklılıkları bulmakta bazı sınırlamalar gözlemlenmiştir.

Karmaşıklı matrisi analizleri, Karar Ağaçları ve Rastgele Orman modellerinin tüm saldırı türlerini doğru bir şekilde sınıflandırarak neredeyse mükemmel sonuçlar verdiğini göstermektedir. Bu durum, topluluk yöntemlerinin CIC-EVSE 2024 gibi karmaşık ve çeşitli veri kümelerini işlemede ne kadar etkili olduğunu ortaya koymaktadır. Öte yandan, Adaboost gibi modellerin daha düşük performans gösterdiği ve daha fazla optimizasyon gerektirdiği belirlenmiştir. AdaBoost modeli, hatalı sınıflandırılmış örneklere daha fazla ağırlık vererek öğrenme sürecini iteratif olarak güçlendirir. Ancak, veri setinde yüksek düzeyde gürültü varsa AdaBoost bu hatalı örneklere aşırı uyum sağlayarak modelin genelleme kabiliyetini zayıflatır. Bu durum, özellikle veri seti dengeli değilse veya özellikler arasında belirgin bir ayrım yoksa, AdaBoost'un düşük performans göstermesine yol açabilir.

Bu çalışmanın bulguları, gelişmiş makine öğrenimi tekniklerinin EV şarj istasyonlarını siber tehditlere karşı koruma konusundaki kritik önemini vurgulamaktadır. Başarılı STS modellerinin uygulanması, siber saldırılarla ilişkili riskleri önemli ölçüde azaltabilir ve şarj altyapısının güvenilirliğini sağlayabilir. Gelecekte, hiperparametre optimizasyonu ve veri artırma teknikleri kullanarak model performansı artırılabilir veya topluluk yöntemlerinin uygulanması tahmin doğruluğunu yükseltebilir. Ayrıca farklı coğrafi bölgelerden ve çeşitli üreticilerden elde edilen daha geniş veri kümeleri kullanılarak modelin genelleme yeteneği artırılabilir ve daha karmaşık derin öğrenme tekniklerinin uygulanması değerlendirilebilir. Son olarak, gerçek zamanlı saldırı tespit sistemlerinin ve donanım tabanlı güvenlik mekanizmalarının geliştirilmesiyle şarj istasyonlarının güvenliği daha da artırılabilir.

Sonuç olarak bu araştırma, EV şarj istasyonları için etkili STS modellerinin geliştirilmesi konusunda sağlam bir temel sağlamaktadır. Bu çalışmada elde edilen umut verici sonuçlar, bu kritik alanda daha ileri çalışmalar için yol açmakta ve elektrikli araçların daha geniş çapta benimsenmesini desteklemekte önemli bir rol oynamaktadır.

REFERANSLAR

[1] F.M. Eltoumi, M. Becherif, A. Djerdir, H.S. Ramadan, *The key issues of electric vehicle charging via hybrid power sources: Techno-economic viability, analysis, and recommendations*, Renew. Sustain. Energy Rev., vol. 138, p. 110534, 2021.

[2] R. Gottumukkala, R. Merchant, A. Tauzin, K. Leon, A. Roche, P. Darby, *Cyber-physical System Security of Vehicle Charging Stations*, 2019 IEEE Green Technologies Conference (GreenTech), Lafayette, LA, USA, 2019, pp. 1-5, doi: 10.1109/GreenTech.2019.8767141.

[3] International Energy Agency, *Trends in electric vehicle charging*, in *Global EV outlook 2024*, 2024. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024/trends-in-electric-vehicle-charging.

[4] M. ElKashlan, H. Aslan, M. S. Said Elsayed, A. D. Jurcut, M. A. Azer, *Intrusion Detection for Electric Vehicle Charging Systems (EVCS)*, Algorithms, vol. 16, no. 2, p. 75, 2023. [Online].

[5] A. Chandwani, S. Dey, A. Mallik, *Cybersecurity of Onboard Charging Systems for Electric Vehicles—Review, Challenges and Countermeasures*, IEEE Access, vol. 8, pp. 226982-226998, 2020.

[6] J. Johnson, T. Berg, B. Anderson, B. Wright, *Review of Electric Vehicle Charger Cybersecurity Vulnerabilities, Potential Impacts, and Defenses*, Energies, vol. 15, no. 11, p. 3931, 2022. [Online].

[7] Q. Feng, H. Li, Y. Zhou, D. Feng, Y. Wang, Y. Su, *Review of electric vehicles' charging data anomaly detection based on deep learning*, 2022 Power System and Green Energy Conference (PSGEC), Shanghai, China, 2022, pp. 337-341.

[8] M. Basnet, M. Hasan Ali, *Deep Learning-based Intrusion Detection System for Electric Vehicle Charging Station*, 2020 2nd International Conference on Smart Power Internet Energy Systems (SPIES), Bangkok, Thailand, 2020, pp. 408-413.

[9] T. Nasr, S. Torabi, E. Bou-Harb, C. Fachkha, C. Assi, *Power jacking your station: In-depth security analysis of electric vehicle charging station management systems*, vol. 112, 2022.

[10] M. ElKashlan, H. Aslan, M. S. Elsayed, A. D. Jurcut, M. A. Azer, *Intrusion Detection for Electric Vehicle Charging Systems (EVCS)*, Algorithms, vol. 16, no. 2, p. 75, 2023.

[11] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, D. Breitenbacher, A. Shabtai, *N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders*, IEEE Pervasive Computing, vol. 17, no. 3, pp. 12-22, 2018.

[12] V. Bezerra, V. Costa, R. Martins, S. Junior, R. Miani, B. Zarpelão, *Providing IoT host-based datasets for intrusion detection research*, Anais do XVIII Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais, pp. 15-28, 2018, Porto Alegre: SBC.

[13] N. A. Stoian, *Machine Learning for anomaly detection in IoT networks: Malware analysis on the IoT-23 dataset*, 2020.

[14] S. Haji, S. Ameen, *Attack and Anomaly Detection in IoT Networks using Machine Learning Techniques: A Review*, Asian Journal of Research in Computer Science, vol. 9, no. 2, pp. 30-46, 2021.

[15] Y.-W. Chung, et al., *The Framework of Invariant Electric Vehicle Charging Network for Anomaly Detection*, 2020 IEEE Transportation Electrification Conference Expo (ITEC), Chicago, IL, USA, 2020, pp. 631-636.

[16] N. Koroniotis, N. Moustafa, E. Sitnikova, B. Turnbull, *Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset*, Future Generation Computer Systems, vol. 100, pp. 779-796, 2019.

[17] A. Guerra-Manzanares, J. Medina-Galindo, H. Bahsi, S. Nomm, *MedBIoT: Generation of an IoT Botnet Dataset in a Medium-sized IoT Network*, 2020.

[18] M. Basnet, M. H. Ali, *Deep Learning-based Intrusion Detection System for Electric Vehicle Charging Station*, 2020 2nd International Conference on Smart Power Internet Energy Systems (SPIES), Bangkok, Thailand, 2020, pp. 408-413.

[19] V. Miskin, S. Chandaragi, U. V. Wali, *Intrusion Detection System for Electric Vehicle Charging Station*, 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, India, 2023, pp. 1-7.

[20] E. Dana Buedi, A. A. Ghorbani, S. Dadkhah, *Enhancing EV Charging Station Security Using A Multi-dimensional Dataset*, CICEVSE2024, 10 March 2024, PREPRINT (Version 1) available at Research Square. [Online].