GAZİOSMANPAŞA BİLİMSEL ARAŞTIRMA DERGİSİ (GBAD)
Gaziosmanpasa Journal of Scientific Research
ISSN: 2146-8168
http://dergipark.gov.tr/gbad
Araştırma Makalesi (Research Article)

Cilt/Volume : 11
Sayı/Number: 3
Yıl/Year: 2022
Sayfa/Pages: 203-220

***Alınış tarihi (Received)*:** 07.11.2022
***Kabul tarihi (Accepted)*:** 06.12.2022

Bilgisayar Ağı Güvenliği için Hibrit Öznitelik Azaltma ile Makine Öğrenmesine Dayalı Bir Saldırı Tespit Sistemi Tasarımı

Muhammed Safa BIÇAKCI[1*], Sinan TOKLU[1]

[1] *Gazi Üniversitesi, Teknoloji Fakültesi, ANKARA*
**Sorumlu yazar: msafa.bicakci@gazi.edu.tr*

ÖZET: Günümüzde teknolojinin ve internetin hızla gelişiminden dolayı ciddi güvenlik tehditleri meydana gelmektedir. Bu gelişim tehditlerinde sürekli değişmesine, gelişmesine ve çeşitlerine neden olmaktadır. Günümüzde teknolojinin ve tehditlerin bu hızla ilerlemesi giderek artan ağ trafiğimizin kontrol ve analiz edilme ihtiyacını gün yüzüne çıkartmaktadır. Analiz sonucu tehditlerin sınıflandırılması için otomatize edilmiş bir saldırı tespit sistemine ihtiyaç duyulmaktadır. Bu ihtiyaç saldırı tespit sistemi ile karşılanabilir. Saldırı tespit sistemi bir tespit sistemi olarak kullanılmaktadır ve ağ güvenliği alanında da kullanılmaktadır. Bu çalışmada makine öğrenmesine dayalı bir saldırı tespit sistemi önerilmektedir. Çalışmada NSL-KDD veri kümesi kullanılarak hem öznitelik çıkartma hem de öznitelik seçme yöntemleri bir arada kullanılarak hibrit bir öznitelik azaltma yöntemi uygulanmıştır ve makine öğrenme modelleri ile sınıflandırma işlemi yapılmıştır. Çalışmanın amacı daha az öznitelik ile yüksek doğruluk oranı elde etmektir. Çalışmada öznitelik çıkartma yöntemi olarak Yığılmış Otomatik Kodlayıcı ve öznitelik seçme olarak SelectKBest yöntemleri uygulanmıştır. Rastgele Orman ve Destek Vektör Makineleri modelleri sınıflandırma için kullanılan makine öğrenme modelleridir. SAE-SKB-RF ve SAE-SKB-SVM önerilen modellerdir. Çalışma sonucunda önerilen modeller birbiri arasında ve literatürde var olan benzer çalışmalar ile karşılaştırılmıştır. Oluşturulan yapı ile saldırılar yüksek başarı oranı ile sınıflandırılmış ve SAE-SKB-RF sınıflandırma metodu kullanılarak %98,67 doğruluk oranı yakalanmıştır. Elde edilen bu oran kullanılan öznitelik azaltma yöntemi ile literatür taramasında yapılan çalışmalara göre en yüksek değeri elde etmiştir.

Anahtar Kelimeler- *Atak tespiti, Makine öğrenimi, Ağ güvenliği, Öznitelik azaltma*

Designing a Machine Learning Based Intrusion Detection System with Hybrid Feature Reduction for Network Security

ABSTRACT: Today, serious security threats occur due to the rapid development of technology and the internet. This causes a constant change, development and variety in development threats. Today, the rapid progress of technology and threats reveals the need to control and analyze our increasing network traffic. An automated intrusion detection system is needed for the classification of threats as a result of the analysis. This need can be met with an intrusion detection system. The intrusion detection system is used as a detection system and is also used in the field of network security. In this study, an intrusion detection system based on machine learning is proposed. In the study, a hybrid feature reduction method was applied by using both feature extraction and feature selection methods using the NSL-KDD dataset, and classification was performed with machine learning models. The aim of the study is to obtain a high accuracy rate with fewer features. In the study, Stacked Autoencoder (SAE) as feature extraction method and SelectKBest method as feature selection were applied. Random Forest and Support Vector Machine models are machine learning models used for classification. SAE-SKB-RF and SAE-SKB-SVM are recommended models. As a result of the study, the proposed models were compared with each other and with similar studies in the literature. With the structure created, attacks were classified with a high success rate and 98.67% accuracy was achieved by using the SAE-SKB-RF classification method. This ratio obtained the highest value compared to the studies made in the literature review with the feature reduction method used.

Keywords- *Intrusion detection, Machine learning, Network security, Attribute reduction*

1. Giriş

Günümüzde teknolojinin gelişmesi, çeşitli cihazların gelişmesi ve iletişim protokollerinin gelişmesi veri boyutunun muazzam bir dereceye ulaşmasına sebep olmaktadır. Analiz edilmeyen bu verilerin ağlarda dolaşması ciddi güvenlik endişelerini gündeme taşımaktadır. Son yıllarda dünya, akıllı şebekeler, nesnelerin interneti, 5G iletişimi gibi bağlantılı teknolojilerin farklı alanlarında önemli bir evrime tanık olunmuştur. Cisco tarafından yapılan bir çalışmada dünya çapında 2022 yılına kadar networke bağlı cihazların sayısının üç katına çıkması ve IP trafiğinin yıllık 4.8 ZB trafik üretmesini beklemektedir (Cisco, 2019). Bu hızlı büyüme iletişim protokollerin ve var olan teknolojilerin kullanımı ile güvenilmeyen olarak adlandırılan "Internet" üzerinde büyük miktarda veri iletimini gerçekleştirmektedir. Bu veri iletimi büyük güvenlik endişelerini beraberinde getirmektedir. Güvenli bir siber alanı oluşturabilmek için, veri ile etkileşime geçmeden önce güvenlik kontrollerinin yapılması gerekmektedir (Fujita, Gaeta, Loia, & Orciuoli, 2019). Bu uygulanması gereken kontroller saldırıları tespit edilmesinden ve bunlara yanıt verilmesinden sorumludur. Saldırı tespit sistemi (Intrusion Detection System - IDS), olası izinsiz girişleri ve şüpheli faaliyetleri gösteren anormallikleri ayrıca bir sistemi hedefleyen dahili ve harici izinsiz girişleri tespit etmek için yaygın olarak kullanılan bir tekniktir. Bir IDS, ek olarak bilgisayar sistemini ve ağ trafiğini izlemek için bir dizi araç ve mekanizma içermektedir (Scarfone & Mell, 2007). Bir IDS sistemini tespit açısından üç gruba ayırabiliriz. Bunlar; İmza Tabanlı Saldırı Tespit Sistemleri (Hubballia & Suryanarayanan, 2014), Anomali Tabanlı Saldırı Tespit Sistemleri (Jyothsna & Vaddella, 2011), (Aldweesh, Derhab, & Emam, 2020), Hibrit Sistemler (Cahyo, Sari, & Riasetiawan, 2020) olarak adlandırılabilirler. İmza tabanlı IDS'de, önceden tanımlanmış saldırıların imzasının karşılaştırılması yapılarak saldırı tespit edilmektedir. Anormallik tabanlı IDS ise davranış analizi ile tespit etmeye çalışır (Liao, Lin, Lin, & Tung, 2013). Hem imza tabanlı hem de anomali tabanlı saldırı tespit sistemlerinin bir arada kullanılması ise Hibrit saldırı tespit sistemi olarak bilinmektedir. Günümüze kadar anormallik tabanlı tespit, imza tabanlı çalışan tespit sisteminden çok geride kalmıştır ve bu nedenle anormallik tabanlı tespit, hala önemli bir araştırma alanı olmaya devam etmektedir (Tavallaee, Stakhanova, & Ghorbani, Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods, 2010). Anormalliği tanımlamak için sistemin daha önce bilgi sahibi olmadığı saldırıları belirlemesi gerekmektedir ve bu da anormallik tabanlı sistemlerin zorluğu olarak bilinmektedir. Sistemin bir şekilde trafiğin analizini yaparak zararlı ve zararsız olarak ayırabilmesi gerekmektedir. Bu ayrımın yapılması için son zamanlarda araştırmacılar tarafından makine öğrenme tekniklerine dayalı çalışmalar yapılmaktadır (Zamani & Movahedi, 2013). Bu konuyla alakalı birçok çalışma yapılmıştır, ama doğruluk en belirgin problemlerden biridir. Doğruluk oranını tespit ederken yanlış alarm oranı ve tespit oranı da göz önüne alınmaktadır. Doğruluk oranı yüksek tespit oranı ve düşük yanlış alarm oranı ile yükseltilebilmektedir. Böylece Destek Vektör Makinesi (Support Vector Machine - SVM) ve Rastgele Orman (Random Forest - RF) uygulanabilmektedir. Sınıflandırma bu algoritmalar tarafından ele alınabilir. Bunun dışında karşılaştırmalı bir analiz yapmak için Normalleştirme ve Özellik Azaltma da uygulanmaktadır.

Saldırı Tespit Sistemleri hem akademik çalışmalar hem de siber güvenlik çalışmaları için ilgi çekici bir alan haline gelmiştir. Geçmiş yıllarda, bu konu hakkında birçok makale yayınlanmıştır. Bu bölümde, geçmiş yıllarda yapılan çalışmalara değinilmiştir.

Bu çalışmada NSL-KDD veri kümesi kullanarak saldırı tespit sistemi geliştirilmiştir. Çalışmada denetimli makine öğrenme yöntemlerine dayanan makine öğrenme algoritmaları ve öznitelik seçme yöntemini beraber kullanılmıştır. Öznitelik seçme yöntemi olarak korelasyon ve chi-square temelli yöntemleri kullanılmıştır ve sınıflandırma için SVM ve Yapay Sinir Ağı (Artificial Neoural Network - ANN) kullanılmıştır (Taher, Jisan, & Rahman, 2019). Bu çalışmada IoT ekosistemine yapılan saldırıları tespit etmek için bir derin öğrenme yöntemi olan Derin Sinir Ağı (Deep Neural Network - DNN) kullanılmıştır. Çalışma üç farklı veri kümesi üzerinde yapılmıştır (KDD-Cup'99, NSL-KDD ve UNSW-NB15). Genel olarak her veri kümesi için %90 üzeri doğruluk oranı elde edilmiştir (Choudharya & Kesswani, 2020). Bu çalışmada saldırı tespit sistemi için derin öğrenme yöntemi olan otomatik kodlayıcının yöntemi önerilmiştir. Otomatik kodlayıcının (Autoencoder - AE) kodlayıcısı, daha az önemli özellikleri sıkıştırmak ve kod çözücü olmadan temel özellikleri çıkarmak için kullanılmıştır. Çalışmada NSL-KDD veri kümesi kullanılmıştır (Zhang, ve diğerleri, 2019). Bu çalışmada ise saldırı tespit sistemi için öznitelik azaltma için seyrek otomatik kodlayıcı kullanılmıştır. Sınıflandırma işlemi ise Lojistik Regresyon (Logistic Regression - LR) ile yapılmıştır. Çalışmada veri kümesi olarak NSL-KDD veri kümesi kullanılmıştır (Gurung, Ghose, & Subedi, 2019). Genel doğruluk oranı %87,2 olarak bulunmuştur. Bu çalışmada ağ yapısında yer alan IoT cihazlarının sensör düğümlerini etkileyen saldırıları NSL-KDD veri kümesi kullanarak ele almışlardır. Çalışmada saldırıların tespiti için 11 adet makine öğrenme algoritması kullanılmıştır ve ağaç tabanlı yöntemlerin ve topluluk yöntemlerinin, incelenen diğer makine öğrenimi yöntemlerinden daha iyi performans gösterdiğini tespit edilmiştir (Liu, Kantarci, & Adams, 2020). En iyi sonucu veren XGBoost modelinin doğruluk oranı %97 olarak bulunmuştur. Ayrıca başka bir çalışmada, evrişimli sinir ağlarını (Convolutional Neural Network - CNN'ler) kullanan yeni bir ağ saldırı tespit modeli önerilmiştir. Ham veri setinden trafik özelliklerini otomatik olarak seçmek için CNN kullanılmıştır ve dengesiz veri seti problemini çözmek için sayılarına göre her sınıfın maliyet fonksiyonu ağırlık katsayısı belirlenmiştir. Önerilen CNN modelinin performansını değerlendirmek için standart NSL-KDD veri seti kullanılmıştır. Önerilen modelin doğruluk değeri %79,48 olarak elde edilmiştir (Wu, Chen, & Li, 2018). Bu çalışmada saldırı tespit sistemi için NSL-KDD veri kümesi kullanılmıştır. Sınıflandırma işlemi için SVM ve Naive Bayes yöntemi kullanılmıştır. Ayrıca veri kümesinin performansını iyileştirmek için öznitelik azaltma yöntemi olan CfsSubsetEval kullanılmıştır (Halimaa & Sundarakantham, 2019). Farklı bir çalışmada ise, Ana Bileşen Analizi (Principal Component Analysis - PCA) adı verilen boyut azaltma tekniği ile anomali tabanlı bir saldırı tespit sistemi modeli önerilmiştir. PCA, daha düşük boyutlu bir biçimde temsil etmek için giriş özellikleri arasındaki bağımlılıkları kullanarak yüksek boyutlu verileri azaltır. Çalışmada NSL-KDD veri seti kullanılmıştır. Sınıflandırma için SVM, Çok Katmanlı Algılayıcı (Multilayer Perceptron – MLP), C4.5 ve Naive Bayes modelleri kullanılmıştır. Çok sınıflı sınıflandırma için en iyi sonuç SVM modeli ile doğruluk oranı %97,61 olarak elde edilmiştir (Subba, Biswas, & Karmakar, 2016).

Bu çalışmada NSL-KDD veri kümesi kullanılmıştır. NSL-KDD Cup 99 veri kümesi, KDD Cup 99 veri seti kümesinin yeni versiyonudur. NSL-KDD Cup 99 veri seti, KDD Cup 99 veri kümesinin bazı sınırlamalarını çözmektedir (Tavallaee, Bagheri, Lu, & Ghorbani, 2009). Örneğin, tekrarlayan örnekler sorunu gibi.

Literatürde yapılan çalışmalar ve önerilen yöntemlerin karşılaştırması Tablo 6'te 4. bölümde sunulmuştur.

Bu çalışmada spesifik olarak;

- Öznitelik azaltma yöntemleri için hibrit bir yöntem kullanılmıştır.
- En iyi SAE modeli bulunması amaçlanmıştır.
- Son olarak önerilen modellerin kendi arasında ve var olan çalışmalar ile karşılaştırması yapılmıştır.

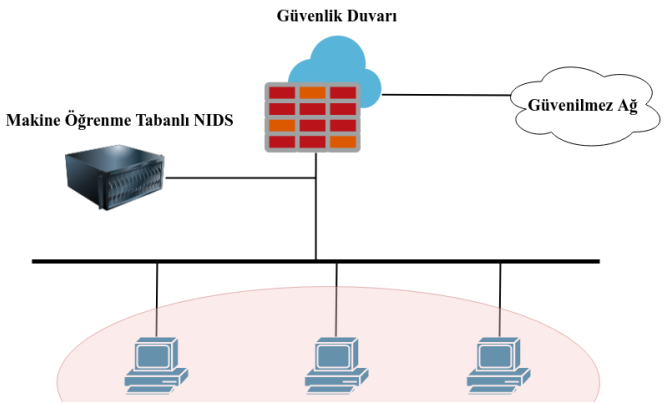
Bölüm 2'de, çalışmada kullanılan yöntemler ve önerilen modelin çalışması hakkında bilgi verilmiştir. Çalışma sonucunda elde edilen değerlerin tartışılması ve NSL-KDD veri setini kullanan önceki birkaç çalışma ile karşılaştırma işlemi ve sonuçları Bölüm 3'te sunulmuştur. Son olarak, sonuç Bölüm 4'te tartışılmaktadır.

2. Materyal ve Yöntem

Bu çalışmada sınıflandırma olarak makine öğrenme yöntemi olan Rastgele Orman ve Destek Vektör Makine yöntemleri kullanılarak bir saldırı tespit sistemi önerilmiştir. Ayrıca saldırı tespit sistemlerinin boyutluluk problemi göz önüne alınarak bu problem için öznitelik azaltma ve öznitelik seçme yöntemleri bir arada kullanılmıştır. Öznitelik azaltma yöntemlerinden elde edilen yeni veri seti çalışmada önerilen sınıflandırma modellerine girdi olarak verilmiştir ve çok sınıflı sınıflandırma yapılmıştır. Bu çalışmadaki amaç elde edilen düşük boyutlu veri seti ile yüksek doğruluk oranı elde etmektir.

Boyut azaltma yöntemleri, özellik çıkarma ve özellik seçimi olarak iki başlıkta kategorize edilebilmektedir. Öznitelik çıkarma ve öznitelik seçimi farklı yöntemlerdir. Öznitelik çıkarma algoritmaları, orijinal özniteliklerden, öznitelik ölçüm maliyetini düşürmek, sınıflandırıcı verimliliğini artırmak ve sınıflandırma doğruluğunu iyileştirmek için yeni öznitelikler türetmektedir (Chumerin & Hulle, 2006), diğer yandan öznitelik seçimi, bir boyut azaltma tekniği olarak, alakasız, fazlalık veya gürültülü özellikleri kaldırarak, orijinal özelliklerden ilgili özniteliklerin küçük bir alt kümesini seçmeyi amaçlar. Özellik azaltma genellikle daha iyi öğrenme performansına, yani daha yüksek öğrenme doğruluğuna, daha düşük hesaplama maliyetine ve daha iyi model tasarımına olanak sağlamaktadır (Miao & Niu, 2016). Amaçları, sınıflandırma görevi için kullanılan daha düşük boyutlu bir karakteristik özellik vektörü türetmek veya seçmektir.

Şekil 1, bağlantı noktası yansıtma teknolojisiyle yapılandırılmış bir ağ anahtarına bağlı olduğu yerde NIDS'nin pasif bir yapısını gösterir. Görev, izinsiz girişleri tespit etmek ve trafik izleme gerçekleştirmek için tüm gelen ve giden ağ trafiğini NIDS'ye yansıtmaktır.



Şekil 1. Saldırı Tespit Sistemi Modeli
Figure 1. Intrusion Detection System Model

2.1 Öznitelik Çıkarma

Öznitelik Çıkarma için kullanılan güncel yöntemler olan Otomatik Kodlayıcı ve Yığılmış Otomatik Kodlayıcı yöntemlerine aşağıda değinilmiştir.

a. Otomatik Kodlayıcı (Autoencoder - AE)

Otomatik kodlayıcı sinir ağı, hedef değerleri girişlere eşit olacak şekilde ayarlayarak geri yayılım uygulayan denetimsiz bir öğrenme algoritmasıdır. Otomatik kodlayıcı, faklı alanlarda kullanılan bir derin öğrenme modelidir. Bu alanlara örnek olarak gürültü giderme ve boyut azaltma verilebilir. Bir AE mimarisi, bir kodlayıcı ve kod çözücü işleminden oluşur: ilk olarak, girdi veri vektörünü tipik olarak daha düşük bir temsile (kodlayıcı) dönüştürür; daha sonra, sıkıştırılmış vektörden (kod çözücü) orijinal girdiyi yeniden oluşturmaya çalışır. AE etiketlenmemiş verilerden önemli özellikleri yakalayabilir (HINTON & SALAKHUTDINOV, 2006).

Denklem 1 ve 2'de sırası ile kodlayıcı ve kod çözücü denklemleri verilmiştir.

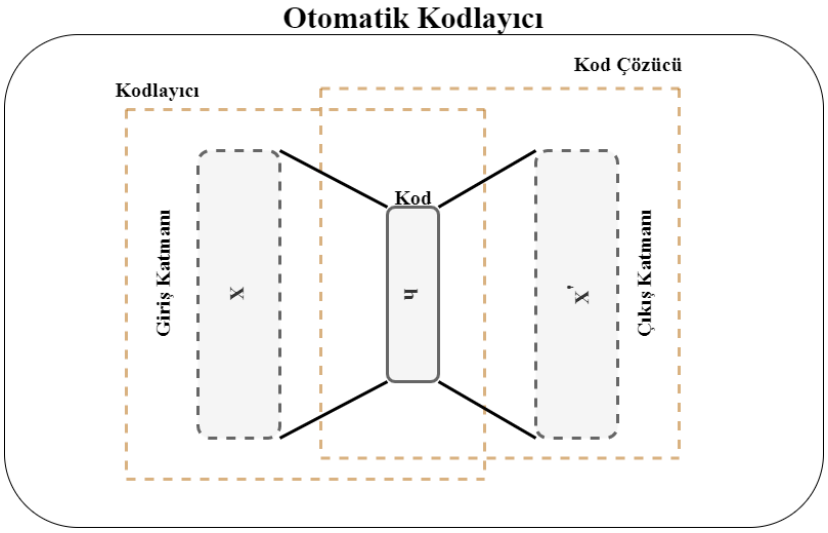
$$y = f(W_x + b_1)$$
Denklem 1

$$z = f_d(W'_y + b_2)$$
Denklem 2

Denklem 1 için; "x" girdi verisin, "y" kodlama işleminden sonra elde edilen değeri, "b_1" giriş sapmasını, "W" girişten gizli katmana katman ağırlığını," f()" doğrusal olmayan aktivasyon fonksiyonunu temsil etmektedir.

Denklem 2 için; "y" gizli katmanda saklanan kodlanmış değerini, "z" otomatik kodlayıcının çıktısını, "b_2" gizli katman sapmasını, "W'" girdi-gizli katman ağırlıklarının devriğini temsil etmektedir.

Şekil 2'de örnek bir otomatik kodlayıcı modeli verilmiştir.



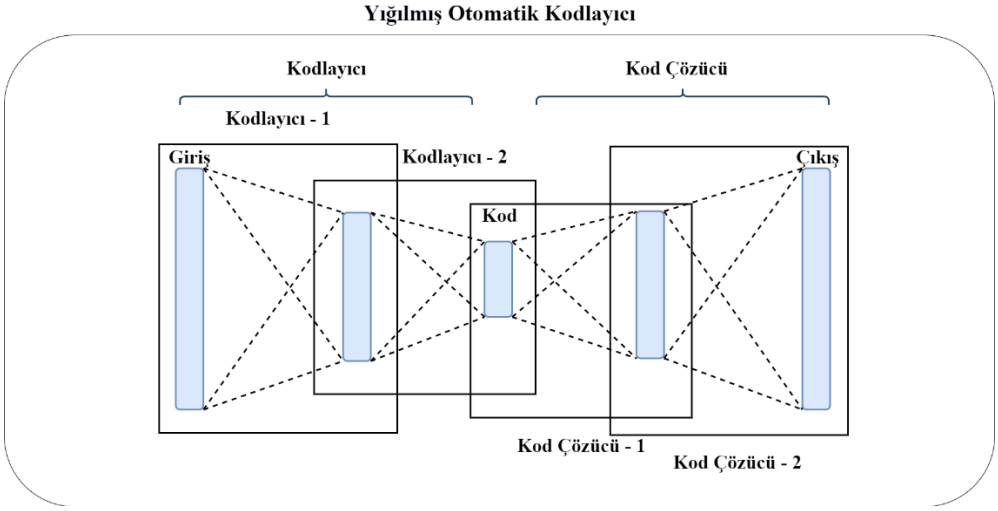
Şekil 2. Otomatik Kodlayıcı Modeli

Figure 2. Autoencoder Model

b. Yığılmış Otomatik Kodlayıcı (Stacked Autoencoder - SAE)

Yığılmış bir otomatik kodlayıcı, her bir gizli katmanın çıktısının ardışık gizli katmanın girişine bağlı olduğu birkaç otomatik kodlayıcı katmanından oluşan bir sinir ağıdır (Yan, Qi, Wang, Lin, & Chen, 2020). Bu yüzden SAE modeli, çok katmanlı bir AE olarak kabul edilebilir.

Şekil 3'te örnek bir otomatik kodlayıcı örneği verilmiştir.



Şekil 3. Yığılmış Otomatik Kodlayıcı Modeli

Figure 3. Stacked Autoencoder Model

2.2 Öznitelik Seçme

Öznitelik Seçme için kullanılan güncel yöntem olan SelectKBest yöntemine aşağıda değinilmiştir.

2.2.1. Select K-best features (SelectKBest)

Select K-Best (SelectKBest) özellik seçme tekniği, doğası gereği tek değişkenlidir. Farklı tek değişkenli istatistiksel testler kullanarak, özellik setinden K-en iyi özellikleri seçer (Pedregosa, ve diğerleri, 2011).
Bu çalışmada, K-en iyi öznitelikleri seçmek için chi-kare (chi-square - χ^2) testine dayalı yöntem kullanılmıştır. χ^2 testi yalnızca negatif olmayan özellikler üzerinde gerçekleştirilebilir. Her negatif olmayan özellik ve hedef özellik için χ^2 puanını hesaplar. Beklenen ve gözlenen frekansların n çifti için χ^2 puanı Denklem 3 ile türetilebilir.

$$\chi^2 = \sum_{i=1}^{n} \frac{(OF_i - EF_i)^2}{EF_i}$$

Denklem 3

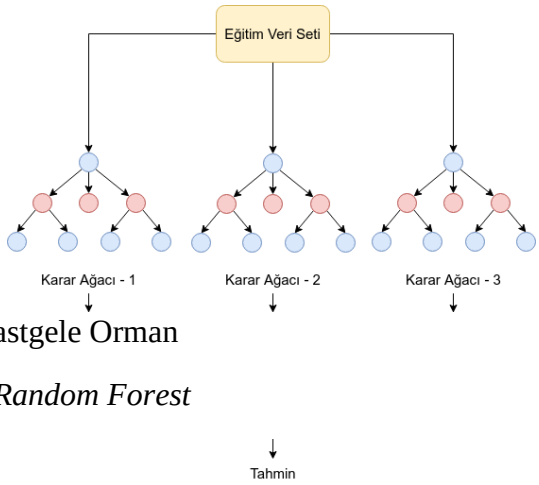
Burada OFi, F özelliğinin i-inci değeri için gözlenen frekans ve EFi, F özelliğinin i-inci değeri için beklenen frekanstır [24].

2.2.2. Destek Vektör Makineleri (Support Vector Machine - SVM)

Destek vektör makinesi (SVM), Vapnik tarafından önerilen ve istatistiksel öğrenme teorisine dayanan bir makine öğrenme yöntemidir (Vapnik, 1999). SVM, sınıflandırma veya regresyon problemlerinde yardımcı olan denetimli bir makine öğrenimi algoritmasıdır. Olası çıktılar arasında optimal bir sınır bulmayı amaçlar. Basitçe söylemek gerekirse, destek vektör makinesi, seçilen çekirdek fonksiyonuna bağlı olarak karmaşık veri dönüşümleri yapar ve bu dönüşümlere dayanarak, tanımladığınız etiketlere veya sınıflara bağlı olarak veri noktalarınız arasındaki ayrım sınırlarını en üst düzeye çıkarmaya çalışır (Elen, Baş, & Közkurt, 2022) (Taher, Jisan, & Rahman, 2019).

2.2.3. Rastgele Orman (Random Forest - RF)

Rastgele ormanlar (Breiman, 2001) budanmamış sınıflandırma veya regresyon ağaçları topluluğudur. Rastgele orman birçok sınıflandırma ağacı üretir. Her ağaç, bir ağaç sınıflandırma algoritması kullanılarak orijinal verilerden farklı bir önyükleme örneğiyle oluşturulur. Orman oluşturulduktan sonra, sınıflandırma için ormandaki ağaçların her birine sınıflandırılması gereken yeni bir nesne konur. Her ağaç, ağacın nesnenin sınıfı hakkındaki kararını belirten bir oy verir. Orman, nesne için en çok oyu alan sınıfı seçer. Örnek bir Rastgele orman yapısı Şekil 4'te verilmiştir.



Şekil 4. Rastgele Orman

Figure 4. Random Forest

2.3. Veri Kümesi Tanımı

Bu çalışmada NSL-KDD veri kümesi kullanılmıştır. Veri kümesinde dört genel saldırı türü vardır: Denial of Service (DoS), Probe, User to Root (U2R) ve Remote to Local (R2L) (Tavallaee, Bagheri, Lu, & Ghorbani, 2009).

- Probe Saldırısı: Hedef sistem ile ilgili bilgi toplamak için yapılan saldırılardır. Örnek olarak; satan, ipsweep, nmap.

- DoS Saldırısı: Hizmet Reddi (DoS) saldırısı, bant genişliğini tüketerek veya kaynakları aşırı yükleyerek bir kaynağın hizmet vermesini engeller. Örnek olarak; smurf, neptune, teardrop.

- User to Root (U2R) Saldırısı: Ele geçirilen hesap ile hak yükselterek yetkili bir duruma gelmesidir. Örnek olarak; eject, load module ve Perl.

- Remote to Local (R2L) Saldırısı: Uzak makinede yetkisi olmayan saldırganın makinenin zafiyetlerinden faydalanarak lokal erişim elde etmesidir. Örnek olarak; ftp_write, şifre tahmin etme ve imap. Tablo 1'de veri kümesi dağılımı verilmiştir.

Tablo 1. Veri Kümesi

Table 1. Dataset

	TOPLAM	NORMAL	DOS	PROBE	R2L	U2R
KDDTRAİN+	125973	67343	45927	11656	995	52
KDDTEST+	22544	9711	7460	2421	2885	67

2.4. Veri Ön işleme

Veri ön işleme, veri kümesine bir model uygulanmadan önce yapılan birtakım işlemlere denmektedir. Bu işlemin yapılması modelimizin daha doğru ve performanslı olarak çalışmasını sağlamaktadır. Yapılan işlemler;

a. Sayısallaştırma

Tek sıcak kodlama yöntemi sayısallaştırma işlemini gerçekleştirmek için kullanılmıştır. "Protocol_type", "Service" ve "Flag" NSL-KDD veri setinde yer alan sembolik özniteliklerdir. Bu öznitelikler tek sıcak kodlama yöntemi ile sayısallaştırılmıştır. Bu işlem sonunda NSL-KDD veri kümesindeki öznitelik boyutu 122 olarak değişmiştir.

b. Normalizasyon

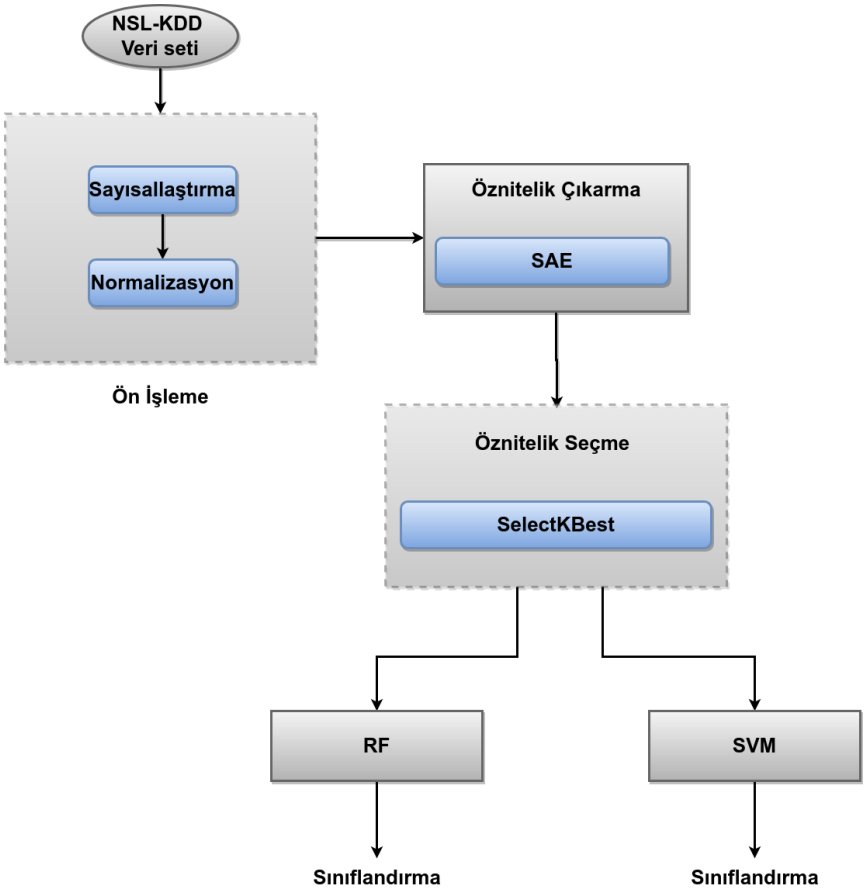
Normalleştirme, tüm değerlerin yeni 0 ve 1 aralığında olması için orijinal aralıktaki verilerin yeniden ölçeklendirilmesidir. Normalizasyon işlemi için Denklem 4'te verilen işlem veri kümesine uygulanmıştır.

$$Y = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad \textbf{Denklem 4}$$

"Y" normalize edilmiş değerini, "X" normalizasyon işlemi öncesi girdi değerini, "X_{min}" X değerlerinin en küçük değerini, "X_{max}" X değerlerinin en büyük değerini temsil etmektedir.

2.5 Önerilen Yöntem

Önerilen modele bu bölümde ayrıntılı olarak bahsedilmiştir. Çalışmanın ana fikri, NSL-KDD veri kümesini kullanarak sınıflandırma işlemi ile bir saldırı tespit sistemi ortaya çıkartmaktır. Bu sistem için sırası ile uygulanan işlemler şu şekildedir; veri ön işlemenin gerçekleştirilmesi, öznitelik azaltma yöntemlerinin uygulanması ve son basamak olarak makine öğrenme modelleri ile sınıflandırma işleminin gerçekleştirilmesi. Çalışmada önerilen modelin mimarisi Şekil 5'te gösterilmiştir. Bu çalışmada öznitelik azaltma olarak hem öznitelik çıkartma hem de öznitelik seçme yöntemi beraber kullanılmıştır. Öznitelik çıkartma yöntemi olarak SAE kullanılmıştır ve ardından SAE işleminden sonra elde edilen veri kümesine öznitelik seçme yöntemi olan SelectKBest yöntemi uygulanmıştır SelectKBest yöntemi ile SAE modellerinden elde edilen öznitelik sayısı sabit bir sayıya ayarlanmıştır. Öznitelik azaltma yönteminden sonra elde edilen yeni veri kümesi önerilen makine öğrenme yöntemlerine girdi olarak verilerek sınıflandırma işlemi yapılmıştır. Veri ön işleme sonucu veri kümemizin öznitelik sayısı 41'den 122'ye yükselmiştir. Derin sinir ağlarında, gizli katmanların sayısını ve her katmandaki nöron sayısını belirlemek için sabit bir yöntem yoktur. Farklı deneysel geçmişlere göre farklı ağ yapıları kurmamız gerekiyor. Gizli katman sayısının az olması ve her katmandaki nöron sayısının yetersiz olması modelin veri dağılımını etkin bir şekilde eşleştirememesine neden olabilir. Tersine, eğer gizli katmanların sayısı çok fazlaysa ve her katmandaki nöronların sayısı fazlaysa, modelin son derece karmaşık eğitim sürecine yol açabilir, bu da eğitim süresini ve bilgi işlem kaynaklarının tüketimini büyük ölçüde artırır. Aynı zamanda modelin aşırı öğrenmesine neden olabilir. Bu sebeple öznitelik çıkartma işleminde kullanılan SAE modeli için standart bir model olmadığı yoktur. Bu nedenden dolayı SAE'nin farklı modellerinin testi de gerçekleştirilmiştir.



Şekil 5. Önerilen Model Mimarisi.

Figure 5. Recommended Model Architecture.

Bu araştırmada, öznitelik çıkarma olan SAE ve öznitelik seçme yöntemi olan SeleckKBest yöntemi art arda kullanılarak öznitelik azaltma işlemi gerçekleştirilmiştir. Daha sonra elde edilen yeni veri kümesi makine öğrenme algoritmaları olan RF ve SVM modellerine ayrı ayrı girdi olarak verilerek sınıflandırma işlemi yapılmıştır ve bu modellerin karşılaştırması yapılmıştır.
Çalışmada NSL-KDD veri kümesi kullanılmıştır ve sınıflar; Normal, DoS, R2L, U2R ve Probe olacak şekilde verilmiştir. Son olarak, önerilen modellerin performansı literatürdeki çalışmalar ile karşılaştırılmıştır. Ayrıca, KDDTrain+ ve KDDTest+ veri kümelerini birleştirdikten sonra veri kümesini eğitim (%66) ve test (%33) olarak iki bölüme ayırarak modeller test edilmiştir.

2.6 Hiper Parametre Kullanımı

SAE ile öznitelik çıkartma yönteminde gizli katman sayısı ve katmanların nöron sayısı için belirli bir yöntem yoktur. Bu belirsizlikten dolayı farklı SAE modelleri test edilmiştir. Böylelikle öznitelik azaltma işlemi gerçekleştirilerek yeni özniteliklere sahip yeni bir veri kümesi elde edilmiştir. Tablo 2'de SAE-3 modeli için kullanılan parametreler verilmiştir. SAE-1, SAE-2 ve SAE-4 için benzer parametreler kullanılmış olup sadece gizli katman yapıları farklıdır.

Tablo 2. SAE-3 için Kullanılan hiper parametreler

Table 2. Hyperparameters Used for SAE-3

KATMAN	ADI	PARAMETRELER	BOYUTLAR
0	Giriş	-	(122)
1	Kodlama	(122, 100)	(100)
	Aktivasyon	relu	-
2	Kodlama	(100, 70)	(70)
	Aktivasyon	relu	-
3	Kodlama	(70, 50)	(50)
	Aktivasyon	relu	-
4	Kodlama	(50, 35)	(35)
	Aktivasyon	relu	-
5	Kod Çözücü	(35, 50)	(50)
	Aktivasyon	relu	-
6	Kod Çözücü	(50, 70)	(70)
	Aktivasyon	relu	-
7	Kod Çözücü	(70, 100)	(100)
	Aktivasyon	relu	-
8	Kod Çözücü	(100, 122)	(122)
	Aktivasyon	sigmoid	-
9	Derleme	-	-
	Optimize Edici	adadelta	-
	Kayıp Fonksiyonu	En Küçük Kareler	-
	Metrik	Doğruluk	-
10	Eğitim	-	-
	Epoch	10	-
	Küme Sayısı	256	-

Tablo 3'te Önerilen model için kullanılan hiper parametreler verilmiştir, SAE-3 modeli için kullanılan parametreler verilmiştir. SAE-1, SAE-2 ve SAE-4 için benzer parametreler kullanılmış olup sadece gizli katman yapıları farklıdır.

2.7. Değerlendirme Metrikleri

Model değerlendirme aşamasında kullanılan metrikler; Doğruluk, Kesinlik, Duyarlılık, F1 skor, MCC (Matthews Correlation Coefficient - MCC) ve Dengeli Doğruluk (Balanced Accuracy) olarak belirlenmiştir. Bu metriklerin matematiksel temsili sırasıyla 5, 6, 7, 8, 9 ve 10 numaralı denklemlere göre hesaplanmaktadır.

$$Doğruluk = \left(\frac{TN+TP}{TN+TP+FP+FN}\right)$$

Denklem 5

$$Kesinlik = \frac{TP}{TP+FP}$$

Denklem 6

$$Duyarlılık = \frac{TP}{TP+FN}$$

Denklem 7

$$F_1 - Skor = 2 * \left(\frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık}\right)$$

Denklem 8

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}}$$

Denklem 9

$$Dengeli\ Doğruluk = 0.5 * \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right)$$

Denklem 10

Tablo 3. Önerilen model için kullanılan hiper parametreler

Table 3. Hyperparameters used for the proposed model

KATMAN	ADI	PARAMETRELER	BOYUTLAR
0	Giriş	-	(122)
1	Kodlama	(122, 100)	(100)
	Aktivasyon	relu	-
2	Kodlama	(100, 70)	(70)
	Aktivasyon	relu	-
3	Kodlama	(70, 50)	(50)
	Aktivasyon	relu	-
4	Kodlama	(50, 35)	(35)
	Aktivasyon	relu	-
5	SelectKBest	Girdi: 35	-
	Yöntem	Chi-kare	10
6	Sınıflandırma	-	-
	Rastgele Orman	n_estimator=10	-
	Kriter	Entropi	-
	Destek Vektör Makineleri	Kernel=rbf	-

3. Bulgular ve Tartışma

Bu bölümde çalışma sonunda elde edilen değerlere ilişkin tablo ve grafiklerin tartışması yapılmıştır.

Tablo 4'te kullanılan SAE modellerinin katman sayısı, elde edilen yeni veri kümesinin öznitelik sayısı ve makalede adlandırılan isimleri gösterilmiştir.

Tablo 4. SAE Model Yapıları
Table 4. SAE Model Structures

Gizli Katman Yapısı	*SAE Sonrası Öznitelik Sayısı*	*SAE İsimleri*
[110, 95, 70, 55, 30, 15]	15	SAE-1
[105, 90, 70, 55, 30]	30	SAE-2
[100, 70, 50, 35]	35	SAE-3
[85, 50, 20]	20	SAE-4

Tablo 5'te yapılan çalışmadan elde edilen yeni veri kümesi ile RF modelinden elde edilen sonuçların kesinlik, duyarlılık, f1-skor, doğruluk, MCC ve Dengeli Doğruluk değerleri verilmiştir. Rastgele Orman için doğruluk açısından en iyi sonuç 3 gizli katmanlı ([85, 50, 20]) SAE-4-SKB-RF modeli %98,67 ile elde etmiştir.

Model	***MCC***	***Dengeli Doğruluk***	***Kesinlik***	***Duyarlılık***	***F1 Skor***	***Doğruluk***
SAE-1-SKB-RF	95,58%	76,06%	0,97	0,97	0,97	97,40%
SAE-2-SKB-RF	97,53%	78,48%	0,99	0,99	0,99	98,54%
SAE-3-SKB-RF	97,49%	78,03%	0,98	0,99	0,98	98,52%
SAE-4-SKB-RF	97,75%	78,26%	0,99	0,99	0,99	98,67%

Tablo 5. SAE-SKB-RF Model Sonuçları
Table 5. SAE-SKB-RF Model Results

Tablo 6'da yapılan çalışmadan elde edilen yeni veri kümesi ile SVM modelinden elde edilen sonuçların kesinlik, duyarlılık, f1-skor, doğruluk, MCC ve Dengeli Doğruluk değerleri verilmiştir. SVM için doğruluk açısından en iyi sonuç 3 gizli katmanlı ([85, 50, 20]) SAE-4-SKB-SVM modeli %95,26 ile elde etmiştir.

Model	***MCC***	***Dengeli Doğruluk***	***Kesinlik***	***Duyarlılık***	***F1 Skor***	***Doğruluk***
SAE-1-SKB-SVM	81,51%	56,56%	0,89	0,89	0,89	89,28%
SAE-2-SKB-SVM	89,10%	60,02%	0,94	0,94	0,93	93,61%
SAE-3-SKB-SVM	91,39%	62,95%	0,95	0,95	0,94	94,95%
SAE-4-SKB-SVM	91,93%	69,67%	0,95	0,95	0,95	95,26%

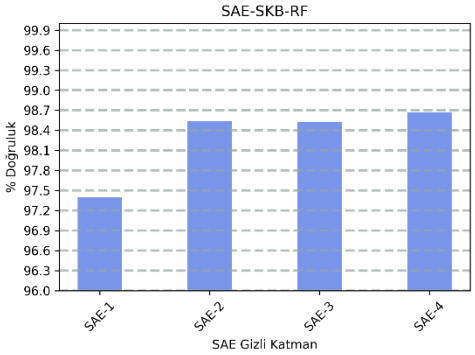
Tablo 6. SAE-SKB-SVM Model Sonuçları

Table 6. SAE-SKB-SVM Model Results

Genel olarak en iyi sonucu 3 gizli katmana ([85, 50, 20]) sahip olan SAE-4-SKB-RF modeli %98,67 doğruluk oranı ile sağlamıştır. SAE öznitelik çıkartma yöntemi ile ilgili belirli bir model yoktur. Buradaki amacımız ilk olarak SAE modelleri ile öznitelik çıkartması yapmak. SAE sonucu elde edilen öznitelikler direkt olarak sınıflandırmada kullanılmamıştır. SAE modelinden elde edilen yeni veri kümesini SelectKBest öznitelik seçme yöntemi kullanılarak yeni veri kümesinin öznitelik sayısını belirli bir sayıya çekiyoruz. Böylelikle SAE modelinden elde edilen veri kümesindeki az öneme sahip özniteliklerin elemesini gerçekleştirmiş oluyoruz. Bu işlemden sonra elde edilen yeni veri kümesi ile makalede belirtilen makine öğrenme yöntemleri ile sınıflandırma işlemini gerçekleştiriyoruz. Normal şartlarda SAE modeli için gizli katman sayısının çok olması performansın artmasını sağlamaktadır (He, ve diğerleri, 2019) yani SAE-1-SKB-SVM modelinin daha başarılı

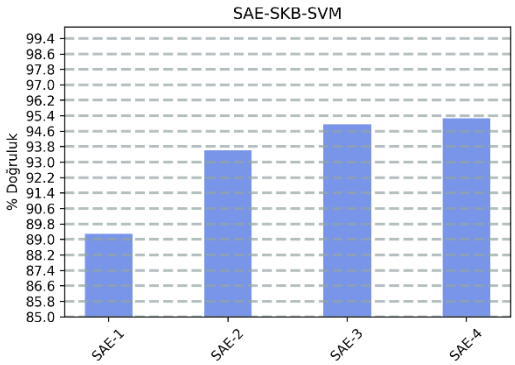
olması beklenmektedir ama makalede yapılan çalışmada SAE yönteminden sonra öznitelik seçme yöntemi kullanılmış olup en iyi sonucu SAE-1-SKB-SVM modeline göre daha az gizli katman yapısına sahip SAE-1-SKB-SVM modeli elde etmiştir. Bu çalışmaya göre daha az gizli katman yapısına sahip SAE modeli ile daha yüksek bir doğruluk oranı elde etmekteyiz. SAE modelinin gizli katman sayısının az olması modelin daha hızlı çalışmasına olanak sağlamaktadır.

Şekil 6 ve Şekil 7'de öznitelik azaltma sonrası SAE gizli katmanlarına göre RF ve SVM modellerinin elde etmiş olduğu doğruluk değerlerinin grafikleri gösterilmiştir.



Şekil 6. SAE-SKB-RF doğruluk oranları

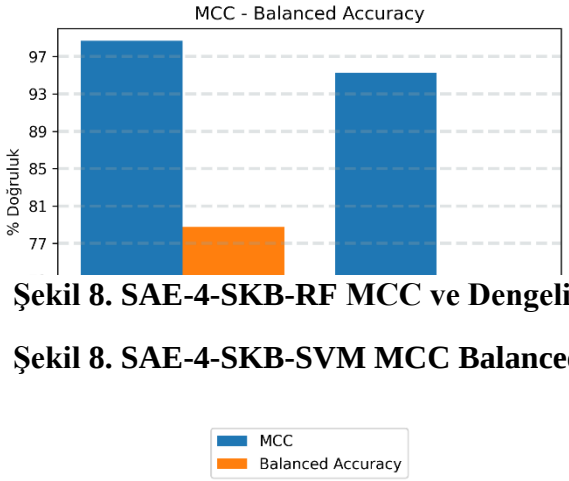
Figure 6. SAE-SKB-RF accuracy rates



Şekil 7. SAE-SKB-SVM doğruluk oranları

Figure 7. SAE-SKB-SVM accuracy rates

Şekil 8'de SAE-SKB öznitelik azaltma yöntemi sonrası elde edilen doğruluk açısından en iyi RF ve SVM modellerinin sırası ile SAE-4-SKB-RF ve SAE-4-SKB-SVM modellerinin MCC ve Dengeli Doğruluk değerlerinin grafiği gösterilmiştir. Bu sonuçlar doğrultusunda RF dengesiz veri kümesine karşı SVM modeline göre daha başarılı sonuç elde etmiştir.



Şekil 8. SAE-4-SKB-RF MCC ve Dengeli Doğruluk Değerleri

Şekil 8. SAE-4-SKB-SVM MCC Balanced Accuracy Values

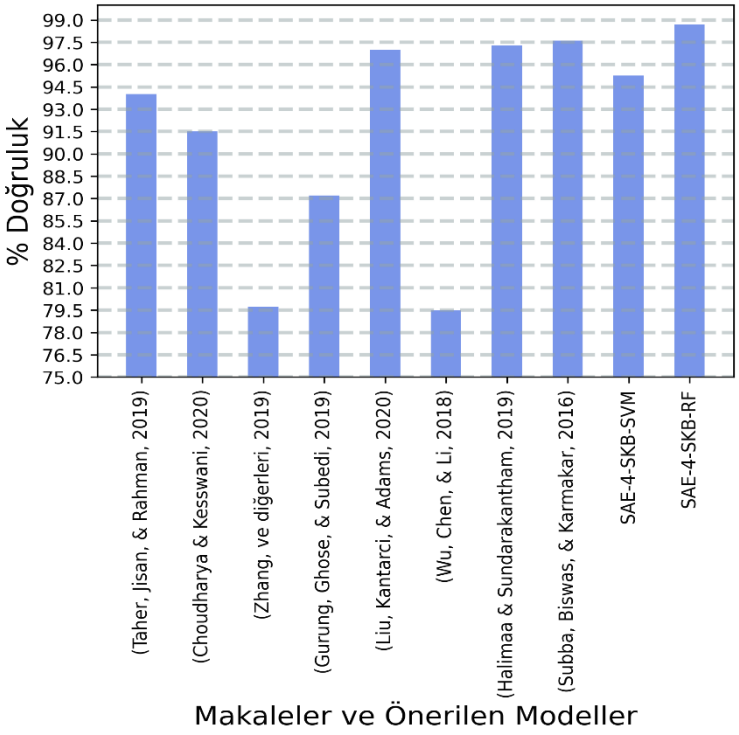
Tablo 7'de literatürde yer alan çalışmalar ile önerilen modellerin karşılaştırılması verilmiştir. Karşılaştırma NSL-KDD veri kümesine göre ve makalelerde elde edilen en yüksek doğruluk oranına göre yapılmıştır. Önerilen modellerden SAE-4-SKB-RF modeli %98,67 ile en iyi doğruluk oranını elde etmiştir.

Tablo 7. Literatür Çalışmaları ve Önerilen Modelin Karşılaştırması

Table 7. Comparison of Literature Studies and Proposed Model

Metot	Öznitelik Sayısı	Veri Kümesi	Doğruluk (%)
ANN (Taher, Jisan, & Rahman, 2019)	17	NSL-KDD	94,02
DNN (Choudharya & Kesswani, 2020)	8	KDD-Cup'99, NSL-KDD ve UNSW-NB15	91,50
AE (Zhang, ve diğerleri, 2019)	122	NSL-KDD	79,74
Sparse AE (Gurung, Ghose, & Subedi, 2019)	10	NSL-KDD	87,20
XGBoost (Liu, Kantarci, & Adams, 2020)	-	NSL-KDD	97,00
CNN (Wu, Chen, & Li, 2018)	122	NSL-KDD	79,48
SVM (Halimaa & Sundarakantham, 2019)	17	NSL-KDD	97,61
SVM (Subba, Biswas, & Karmakar, 2016)	-	NSL-KDD	97,29
Önerilen Model (SAE-4-SKB-RF)	10	NSL-KDD	**98,67**
Önerilen Model (SAE-4-SKB-SVM)	10	NSL-KDD	95,26

Şekil 9'da literatürde yer alan çalışmalar ile önerilen modellerin doğruluk değeri açısından grafiği verilmiştir.



Şekil 9. Literatür ve Önerilen Model Doğruluk Oranları
Figure 9. Literature and Suggested Model Accuracy Rates

4. Sonuç

Bu çalışmada, Destek Vektör Makineleri ve Rastgele Orman makine öğrenimi sınıflandırıcılarının verimliliğini ve performansını değerlendirmek için çeşitli deneyler yapılmış ve test edilmiştir. Tüm testler NSL-KDD saldırı tespit sistemi veri kümesi kullanılarak gerçekleştirilmiştir. Çalışmada çok boyutluluk problemi ele alınarak öznitelik azaltma yöntemleri uygulandıktan sonra Rastgele Orman ve Destek Vektör Makineleri makine öğrenme modelleri ile sınıflandırma işlemi yapılmıştır. Öznitelik azaltma işlemi, öznitelik çıkartma yöntemi olan SAE ve öznitelik seçme yöntemi olan SelectKBest işlemleri sıra ile art arda kullanılarak gerçekleştirilmiştir. Bu işlemden sonra elde edilen yeni veri kümesinin öznitelik sayısı sabit bir sayıya çekilmiştir. Böylelikle daha az öznitelik sayısı daha yüksek doğruluk oranı elde ettiği gösterilmiştir. Yapılan çalışmalar sonucunda önerilen model olan SAE-4-SKB-RF modeli %98,67 ile en yüksek doğruluk oranını elde etmiştir ve bu model için kullanılan SAE modeli SAE-4 olarak belirtilen ve gizli katman yapısı ([85, 50, 20]) olan modeldir. Veri kümesinin boyutunu küçültmek, yalnızca algılama sisteminin öğrenme performansını iyileştirmekle kalmaz; ayrıca veri kümesinin fazlalığını da azaltabilir. Literatür çalışmalarından farklı olarak öznitelik çıkartma ve seçme yöntemleri bir arada kullanılarak boyut azaltma işlemi gerçekleştirilmiştir. Bu işlemlerin kullanılması sınıflandırma açısından literatür çalışmalarına oranla daha yüksek doğruluk elde etmiştir. Çalışmada veri kümesinin dengesizliği de göz önüne alınarak MCC ve Dengeli Doğruluk değerleri de hesaplanmıştır ve SAE-4-SKB-RF modelinin en yüksek değerleri elde ettiği gözlemlenmiştir. Geleneksel saldırı tespit sistemi yalnızca bilinen saldırıları tespit edebilir. Yeni saldırıların veya sıfır gün saldırılarının tespiti, mevcut sistemlerin yüksek yanlış pozitif

oranı nedeniyle hala bir araştırma konusu olmaya devam etmektedir. Önerilen yöntemin dezavantajı, R2L ve U2R düşük frekanslı saldırı örneklerini etkin bir şekilde tespit edememesi, yani dengesiz veri dağılımından kaynaklanan olumsuz etkilerin üstesinden gelememesidir. Gelecekteki araştırmalarda, öznitelik çıkarma sürecindeki dengesiz veri sorununu ele almak için mevcut yöntemlerin nasıl kullanılacağı veya yeni yöntemler önerileceği konusu araştırılacaktır ayrıca geliştirilen modelin canlı bir ortamda testinin yapılması da amaçlanmaktadır.

5. Kaynaklar

Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowledge-Based Systems*.

Breiman, L. (2001). Random Forests. *Machine learning*, s. 5-32.

Cahyo, A. N., Sari, A. K., & Riasetiawan, M. (2020). Comparison of Hybrid Intrusion Detection System. *In 2020 12th International Conference on Information Technology and Electrical Engineering*, s. 92-97.

Choudharya, S., & Kesswani, N. (2020). Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT. *Procedia Computer Science*, s. 1561-1573.

Chumerin, N., & Hulle, M. M. (2006). Comparison of Two Feature Extraction Methods Based on Maximization of Mutual Information. *In 2006 16th IEEE signal processing society workshop on machine learning for signal processing*, s. 343-348.

Cisco. (2019). *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*. www.cisco.com: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networkingindex-vni/complete-white-paper-c11-481360.html adresinden alındı

Elen, A., Baş, S., & Közkurt, C. (2022, 01 30). An Adaptive Gaussian Kernel for Support Vector Machine. *Arabian Journal for Science and Engineering*, s. 10579–10588.

Fujita, H., Gaeta, A., Loia, V., & Orciuoli, F. (2019, 5). Resilience Analysis of Critical Infrastructures: A Cognitive Approach Based on Granular Computing. *IEEE Transactions on Cybernetics*, s. 1835-1848.

Gurung, S., Ghose, M. K., & Subedi, A. (2019). Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset. *International Journal of Computer Network and Information Security*, s. 8-14.

Halimaa, A. A., & Sundarakantham, K. (2019). Machine Learning Based Intrusion Detection System. *In 2019 3rd International conference on trends in electronics and informatics (ICOEI)*, s. 916-920.

He, D., Qiao, Q., Gao, Y., Zheng, J., Chan, S., Li, J., & Guizani, N. (2019). Intrusion Detection Based on Stacked Autoencoder for Connected Healthcare Systems. *IEEE Network*, s. 64-69.

HINTON, G. E., & SALAKHUTDINOV, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *science*, s. 504-507.

Hubballia, N., & Suryanarayanan, V. (2014). False alarm minimization techniques in signature-based intrusion detection systems. *A survey. Computer Communications*, s. 1-17.

Jyothsna, V., & Vaddella, R. P. (2011). A Review of Anomaly based IntrusionDetection Systems. *International Journal of Computer Applications*, s. 26-35.

Liao, H.-J., Lin, C.-H. R., Lin, Y.-C., & Tung, K.-Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, s. 16-24.

Liu, J., Kantarci, B., & Adams, C. (2020). Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset. *In Proceedings of the 2nd ACM workshop on wireless security and machine learning*, s. 25-30.

Miao, J., & Niu, L. (2016). A Survey on Feature Selection. *Procedia Computer Science*, s. 919-926.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grise, O., . . . Duchesnay , E. (2011). Scikit-learn: Machine Learning in Python. *the Journal of machine Learning research*, s. 2825-2830.

Scarfone, K., & Mell, P. (2007). Guide to intrusion detection and prevention systems. *NIST special publicatio*, s. 94.

Subba, B., Biswas, S., & Karmakar, S. (2016). Enhancing performance of anomaly based intrusion detection systems through dimensionality reduction using principal component. *In 2016 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, s. 1-6.

Taher, K. A., Jisan, B. M., & Rahman, M. (2019). Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection. *In 2019 International conference on robotics, electrical and signal processing techniques*, s. 643-646.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *In 2009 IEEE symposium on computational intelligence for security and defense applications*, s. 1-6.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *In 2009 IEEE symposium on computational intelligence for security and defense applications*, s. 1-6.

Tavallaee, M., Stakhanova, N., & Ghorbani, A. A. (2010). Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods. *IEEE Transactions on Systems, Man, and Cybernetics*, s. 516-524.

Vapnik, V. N. (1999). The nature of statistical learning theory. *Springer science & business media*.

Wu, K., Chen, Z., & Li, W. (2018). A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks. *Ieee Access*, s. 50850-50859.

Yan, Y., Qi, L., Wang, J., Lin, Y., & Chen, L. (2020). A Network Intrusion Detection Method Based on Stacked Autoencoder and LSTM. *In ICC 2020-2020 IEEE International Conference on Communications (ICC)*, s. 1-6.

Zamani, M., & Movahedi, M. (2013). Machine learning techniques for intrusion detection. *arXiv preprint arXiv:1312.2177*.

Zhang, C., Ruan, F., Yin, L., Chen, X., Zhai, L., & Liu, F. (2019). A Deep Learning Approach for Network Intrusion Detection Based on NSL-KDD Dataset. *In 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification*, s. 41-45.

Zulfiker, M. S., Kabir, N., Biswas, A. A., Nazneen, T., & Uddin, M. S. (2021). An in-depth analysis of machine learning approaches to predict depression. *Current research in behavioral sciences*.