www.dergipark.gov.tr
ISSN:2148-3736
El-Cezerî Fen ve Mühendislik Dergisi
Cilt: 8, No: 3, 2021 (1550-1564)

El-Cezerî Journal of Science and Engineering
Vol: 8, No: 3, 2021 (1550-1564)
DOI :10.31202/ecjse.971592

ECJSE

Research Paper / Makale

Kablosuz Sensör Ağlarda Makine Öğrenme Metotları Kullanılarak DOS Saldırıları Tespiti

Celil OKUR[1a], Murat DENER[1b*]

[1] Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Bilgi Güvenliği Mühendisliği, Ankara, Türkiye
celil.okur@gazi.edu.tr, muratdener@gazi.edu.tr

Received/Geliş: 15.07.2021 **Accepted/Kabul:** 19.09.2021

Öz: Son yıllardaki teknolojik gelişmelerle birlikte Kablosuz Sensör Ağlarının (KSA) kullanım alanları ve popülaritesi artmaktadır. Özellikle IoT teknolojisiyle birlikte çalışan sensör ağları; akıllı arabalar, akıllı evler, akıllı şehir gibi sivil uygulama alanlarında, askeri alanlarda ve endüstri-sanayide kullanılmaktadır. Kullanıldığı alanlar itibari ile saldırılara açık bir yapıya sahiptir. Bu saldırıların bazı fiziksel bazıları ise programsaldır. Hayat kalitesini arttıran ve hayatı kolaylaştıran bu teknolojilere yapılan saldırıları önlemek için bu alanda çeşit çalışmalar yapılmaktadır. Bu çalışmada KSA ağ saldırıları veri seti alınarak yapay zeka teknolojisinin alt dalı olan makine öğrenme modelleri ile analiz edilmiştir. Bu çalışmada WSN-DS saldırı veri seti kullanılmıştır. Veri seti, NS 2 benzetim ortamında oluşturulmuştur. Veri seti Grayhole, Blackhole, Flooding, TDMA gibi ağ saldırı trafiklerinden ve normal ağ trafiğinden oluşmaktadır. Bu veri seti makine öğrenme modellerinden gözetimli ve gözetimsiz modellerle incelenmiştir. Gözetimli öğrenme modellerinden; Decision Tree (J48), Random Forest, Naive Bayes algoritmalarıyla incelenmiş, gözetimsiz öğrenme modellerinden; Expectation Maximization (EM), Simple Kmeans, Filtered Clusterer,Canopy algoritmaları ile incelenmiştir. Sonuçları uygulama bölümünden tablolarla gösterilmiştir. Çalışma java tabanlı Weka 3.8.3 kullanılarak gerçekleştirilmiştir.

Anahtar Kelimeler: KSA, DOS saldırıları, makine öğrenmesi siber güvenlik, NS2

Detection of DOS Attacks Using Machine Learning Methods in Wireless Sensor Networks

Abstract: With the technological developments in recent years, the usage areas and popularity of Wireless Sensor Networks (WSNs) are increasing. Especially sensor networks working with IoT technology; It is used in civil application areas such as smart cars, smart homes, smart city, military fields and industry. In terms of the areas in which they are used, WSNs have a structure that is open to attacks. Some of these attacks are physical and some are programmatic. Various studies are carried out in this area to prevent attacks on these technologies that increase the quality of life and make life easier. In this study, WSNs attacks data set were taken and analyzed with machine learning models, which are sub-branches of artificial intelligence technology. The WSN-DS attack dataset used was created in the NS 2 simulation environment and consists of network attack traffics such as Grayhole, Blackhole, Flooding, TDMA and normal network traffic. In this study, the data set is analyzed with Supervised learning models (Decision Tree (J48), Random Forest, Naive Bayes) and Unsupervised learning models (Expectation Maximization (EM), Simple Kmeans, Filtered Clusterer, Canopy). The study was carried out using java based Weka 3.8.3, and the experimental results obtained are presented in detail.

Keywords: WSNs, DOS, attacks, machine learning, cyber security, NS2

1. Giriş

Kablosuz sensör ağların hayatı büyük oranda kolaylaştırmasından dolayı son yıllarda her alanda kullanımı artmaktadır. KSA market hacmi 2024 de 1,8 milyon dolara ulaşması beklenmektedir [1, 2, 3]. KSA'nın hem tek başına hem de IoT sistemlerde kullanılması KSA'nın popülaritesini ve

How to cite this article
Okur, C., Dener, M., "*Makine Öğrenme Metotları Kullanılarak KSA DOS Saldırıları Tespiti*" El-Cezerî Journal of Science and Engineering, 2021, 8(2); 1550-1564.

Bu makaleye atıf yapmak için
Okur, C., Dener, M., *"Detecting Wsn DOS Attacks Using Machine Learning Methods"* El-Cezerî Fen ve Mühendislik Dergisi 2021, 8(2); 1550-1564.
ORCID ID: [a]0000-0002-0773-6438; [b]0000-0001-5746-6141

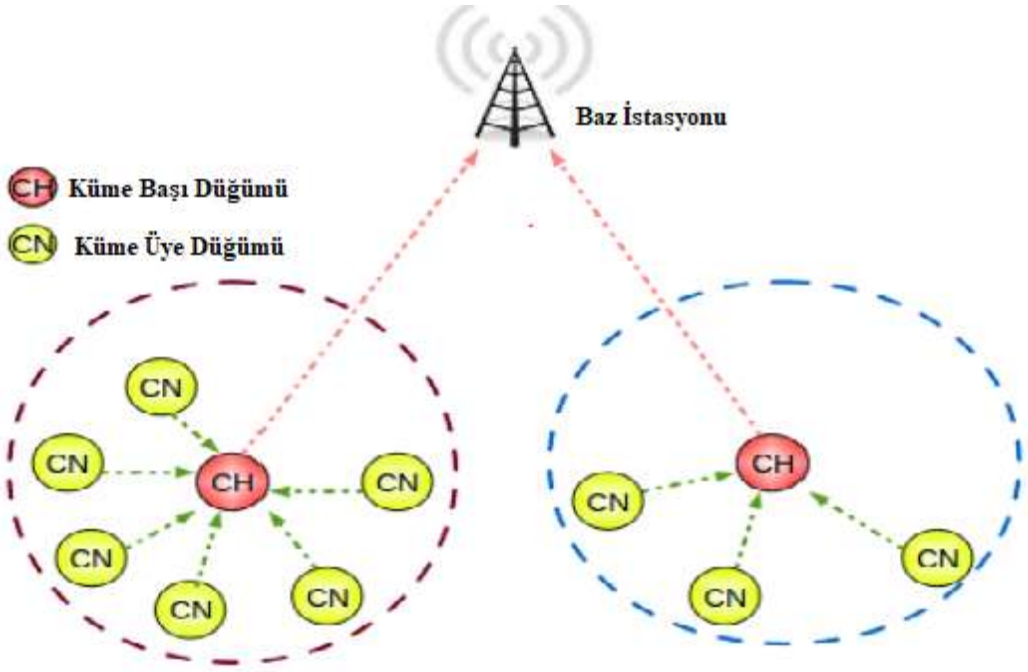
önemini daha da arttırmaktadır. KSA kablosuz servis ile çevrenin algılanmasını sağlayan sensörlerden oluşan senkronize çalışan bir organizasyondur [4, 5]. Sensör ağları isminden de anlaşılacağı üzere geliştiricinin belirli özelliklerdeki sensörleri ortak bir ağda toplayarak oluşturduğu algılayıcı sistemlerdir. Amaca uygun olarak birçok alanda kullanılabilen, modüler, kullanıcı dostu, kullanıcıya pratik çözümler sunan ve alt yapı gerektirmeyen bir sistemdir [6]. Kablosuz sensör ağları iç ortamlarda kullanılmasının yanında dış ortamlarda da özellikle erişilmesi güç, tehlikeli, zor alanlarda periyodik bakımı gerektirmediği için kullanılmaktadır. Bu kapsamda askeri alan, vahşi yaşam, endüstri, sağlık ve gündelik hayat gibi birçok farklı alanda uygulamaları görülmektedir [7]. KSA'nın avantajlarının yanında birtakım dezavantajları da bulunmaktadır. Sınırlı kaynakları, fiziksel ve ağ saldırılarına açık olmaları, bazı coğrafi koşullardan dolayı dış ortamlarda arızalanmaları gibi konularda dezavantajları vardır. Bu çalışmada KSA ya yapılan ağ saldırıları üzerinde durulmuştur. Klasik bilgisayar ağlarında oluğu gibi sensör ağlarda da sensörler birbirleri ile ve belirlenen merkez ile iletişim halinde çalışırlar. Bu haberleşme sırasında sensörler hem kendileri ile ilgili bilgileri hem de ortamdan algıladıkları bilgileri gönderip alırlar. Bu veri transferleri, gizlilik, bütünlük ve erişilebilirlik gibi bilgi güvenliğinin en temel ilkelerine göre gerçekleştirilmektedir [1]. Bu transfer gerçekleştirilirken sensör ağlarının içinden ya da dış ağdan bazı saldırılar olabilmektedir. Bu saldırılar amaçlarına göre bazen KSA için hayati öneme sahip kaynaklarını tüketmeye yönelik olurken bazıları iletişimi bozma ya da durdurma, bazıları ise ağı erişimsiz hale getirmeye yönelik olabilmektedir. Bu saldırıları tespit etmek ve engellemek için bilgisayar ağlarında güvenlik duvarları güvenlik programları çalıştırılmaktadır. Sensör ağlarda ise bu durum bir merkezden ya da hiyerarşik bir düzende küme başları düğümlerde çalışan algoritmalarla tespit edilebilmektedir. Bu algoritmalar geliştirilirken birçok yaklaşım kullanılmaktadır. Bu yaklaşımlar, zararlı ağ trafiğine göre geliştirilen, normal ağ trafiğine göre geliştirilen ve geçmiş saldırılardan edinilen bilgilere göre geliştirilen şeklinde sınıflandırılmaktadır. Bu çalışmada da farklı makine öğrenmesi modelleri ile normal ve farklı türlerdeki saldırı trafiği kategorize edilmiştir. Çalışmanın bu bölümünden sonraki ikinci bölümde KSA ağları ve ağda çalışan protokoller anlatılmıştır. Üçüncü bölümde KSA ağlarındaki saldırı türleri ve çözümleri ile ilgili bilgiler verilmiştir. Dördüncü bölümde makine öğrenmesi ve modelleri ilgili bilgiler verilmiştir. Beşinci bölümde Kablosuz Sensör Ağları saldırı verileri, makine öğrenmesi modelleri ile analiz edilmiştir. Son bölümde ise çalışmanın sonuçları sunulmuştur.

2. KSA

Sensor ağları alt yapı gerektirmeyen, pratik, modüler, iç-dış ortamlarda kullanılabilen, farklı sektörlerde uygulaması olan bir teknolojidir. Bu teknolojiyi anlamak için çalışma yapısını incelemek gerekmektedir. Öncelik KSA, sıcaklık hareketlilik nem sıcaklık gibi ortamdaki değişiklikleri ölçer. Her bir sensörün çalıştığı modüle düğüm denmektedir. Düğümler sensör, işlemci gönderici ve güç birimlerinden oluşur. Düğümlerde enerji, hafıza, işlemci gibi kaynaklar sınırlıdır [1]. Bu yüzden asgari kaynaklarla azami performans istenmektedir. Sensör ağlarındaki bu durumlar düşünülerek çeşitli önceliklere göre iletişim gerçekleşir. Bu iletişim belirlenen protokollere göre gerçekleşir. Protokoller düz, hiyerarşik, yerel temelli olmak üzere üç tip mimari yapı ile çalışır. Literatürde LEACH, PEGASIS, TEEN, APTEEN protokolleri bilinen hiyerarşik yönlendirme protokolleridir. Protokollerin çalışması bir algoritmaya dayanmaktadır. Bu algoritma geliştiricinin bu alandaki istekleri analiz ederek gerekleri belirleyip ona göre geliştirilir. KSA'da bazı protokoller güvenliği ön planda tutarken bazıları enerji tüketimini, bazıları düğüm konumunu ve hiyerarşiyi ön planda tutmaktadır. Bu çalışmada hiyerarşik bir mimariye sahip, enerji temelli çalışan LEACH protokolü kullanılmıştır. Şekil 1' de LEACH protokolünün KSA'da hiyerarşik çalışma şekli gösterilmiştir.

Low Energy Adaptive Clustering Hierarchy (LEACH) adından da anlaşılacağı gibi bu protokolde enerji temelli hiyerarşik bir yapı ile çalışmaktadır [8]. KSA'da hiyerarşi denildiğinde ilk akla gelen standart düğümler ile kümeler oluşturulup, bu kümenin başına çeşitli kriterlere göre bir küme başı

düğümü seçip tüm işlemlerin bu küme başı düğüm üzerinden sağlanması anlaşılmaktadır. Küme başı düğümlerde bir merkeze bağlı olup sistem bu hiyerarşik düzende çalışmaktadır. Protokol, enerji temelli çalışmaktadır. Küme başı düğümlerinde normal düğümlere göre daha çok enerji tükettiği bilindiğinden bu protokol enerjisi azalan küme başı düğümü belirli zamanlarda değiştirerek küme başı üzerinden haberleşen düğümün ağdan kopmaması sağlanır [9].



Şekil 1. LEACH protokolünde düğüm yerleşimi

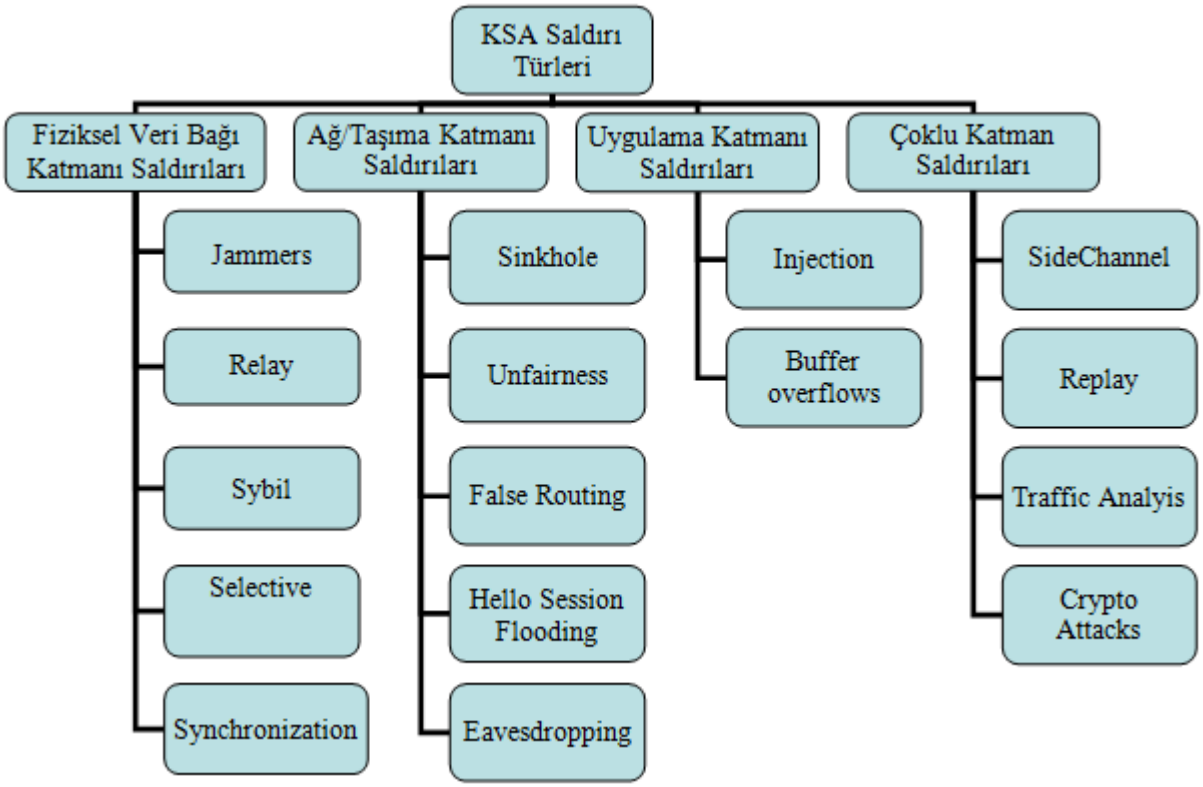
KSA'da iletişim tıpkı bilgisayar ağlarındaki gibi katmanlı protokollerle sağlanır. Bu katmanlar; Fiziksel Katman, Veri Bağı Katmanı, Ağ katmanı, Taşıma Katmanı ve Uygulama Katmanı olmak üzere 5 katmandan oluşmaktadır. KSA saldırılarının büyük bir bölümü bu katmanlara göre dizayn edilip yapılmaktadır. Bu yüzden saldırılar bu katmanlara göre kategorize edilmektedir. KSA 'da saldırının olup olmadığı ya da saldırgan düğümün ağdaki hangi düğüm olduğu belirli parametreler incelenerek tespit edilmektedir. Bu parametre, düğümlerin kalan enerji miktarları, düğümlerin düşürdüğü paket sayısı, düğümlerin aldığı paket sayısı, düğümlerin ilettikleri paket sayısı, düğümlerin gönderdiği paket miktarı gibi parametrelerdeki veriler, kodlanarak ya da yapay zekâ teknoloji ile incelenerek ağ trafiğinin ya da düğümün saldırgan olup olmadığı anlaşılabilir [5].

3. KSA Saldırı Türleri

KSA günlük hayatta birçok alanda kullanılmasından dolayı farklı gizlilik seviyelerinde bilgiler taşımaktadır. Bu bilgiler KSA'nı saldırganların hedefi haline getirmektedir. Saldırılar farklı şekillerde sınıflandırılabilirler. Bu çalışma için saldırılar katmanlara göre sınıflandırılmıştır. Genel olarak fiziksel katman saldırıları, veri bağı katman saldırıları, ağ saldırıları, transport katman saldırıları, uygulama katman saldırıları şeklinde yapılan saldırılar sınıflandırılmaktadır.

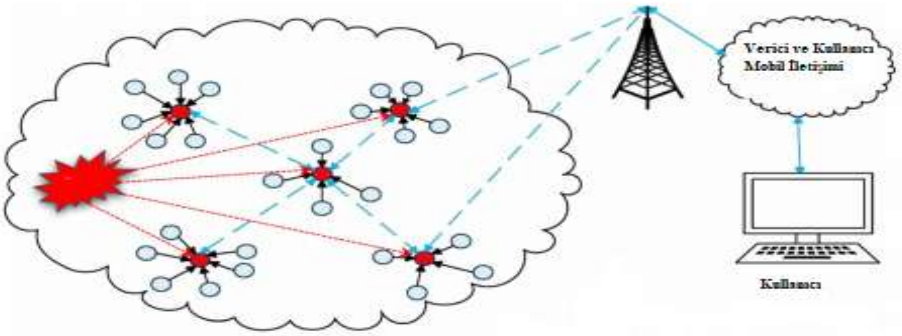
Veri setinin hazır olarak literatürden alınması ve literatürde de veri setinin DOS saldırı veri seti olarak geçmesinden dolayı çalışmada bu isimle kullanılmıştır. Ancak DOS ve DDOS saldırılarını birbirinden ayıran en önemli özellik, isminden de anlaşılacağa gibi birincisi servis reddi (DOS) diğeri dağıtık servis reddi (DDOS) saldırılarıdır [4]. İkisinde de amaç hedefi erişilmez kılmak ve

kullanıcıların hedefe ulaşması engellemektir. İlkinde bu atağı tek saldırgan yaparken ikincisinde birden çok saldırgan gerçekleştirmektedir. Son yıllarda DDOS saldırısının kullanımı zararlı yazılımlar ve bu yazılımları haberli-habersiz kullanan cihazların manipüle edilmesi ile artmıştır. DDOS, DOS saldırısına göre daha senkron ve zahmetli olsa da daha etkili ve zararlı sonuçlar meydana getirmektedir. Aşağıda belirtilen saldırı türleri DOS saldırılarında da DDOS saldırında da kullanılmaktadır.



Şekil 2. KSA katmanlarına göre saldırı türleri

Saldırıların gerçekleştirilme şekilleri farklı olsa da amaçları aynıdır. Literatürde DOS ve DDOS saldırılarının örnekleri bilgisayar ağlarında da çokça görülmekte ve amacı, uygulaması ve sonucu aynıdır. Bu çalışmada incelenen veri seti, ağ katmanına yönelik yapılmış saldırılardan oluşmaktadır.

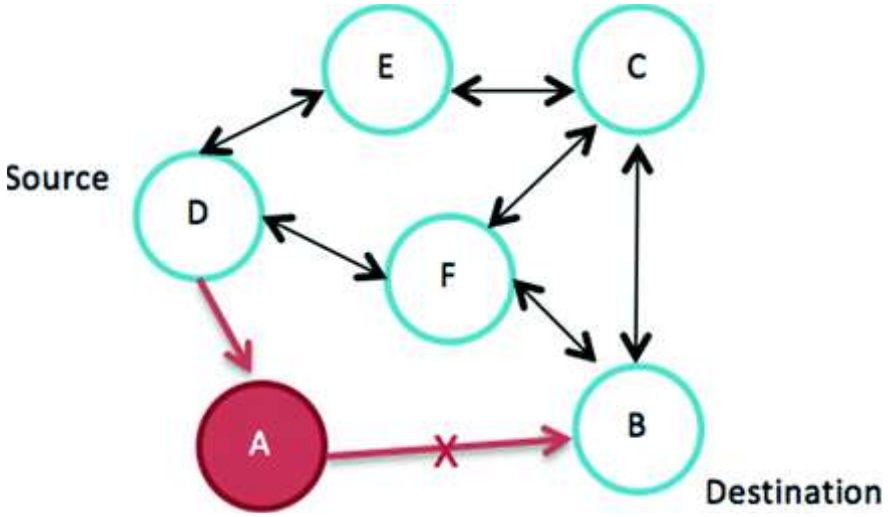


Şekil 3. Flooding saldırısı

Çalışmada DOS saldırılarından olan flooding, grayhole, blackhole, scheduling saldırı veri seti incelenmiştir. Makine öğrenme modelleri kullanılarak normal ve anormal ağ trafikleri kategorize edilmiştir. Anormal trafiğe neden olan ve bu çalışmada incelenmiş olan DOS saldırı türlerinden aşağıda kısaca bahsedilmiştir.

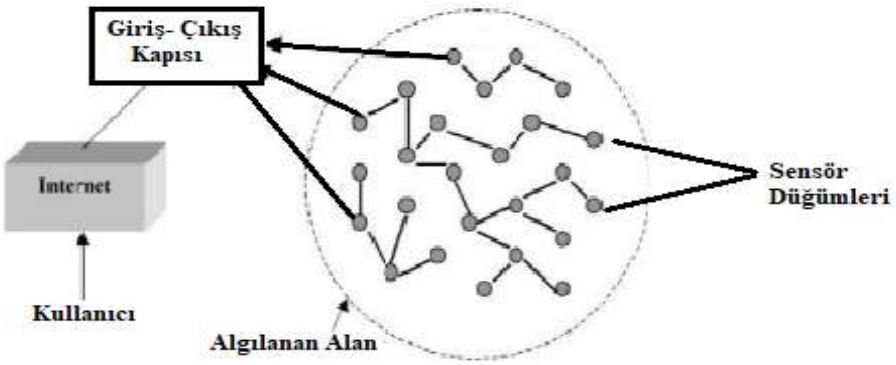
Flooding Saldırısı: Zararlı düğümün ağdaki diğer düğümlere yeni katılan normal bir düğüm gibi merhaba (hello) paketi göndererek yapılan saldırı çeşididir. Normal düğümlerden farklı olarak hello paketini sürekli göndererek ağın kaynaklarını tüketmeyi ve erişilmez kılmayı hedeflemektedir [10].

Blackhole Saldırısı: Saldırgan düğüm, ağdaki trafiği paket dağıtım oranı, kalan enerji miktarı, paket düşürme miktarı, güven değeri gibi değerlerini değiştirerek kendini en cazip düğüm haline getirip trafiği üzerine çekmektedir [5]. Bu düğüme düğer düğümlerden iletmesi için paketler gelmektedir. Saldırgan düğüm bu paketleri düşürerek trafikte kopukluklar ve bozulmalar meydana getirmektedir [11].



Şekil 4. Blackhole saldırısı

Grayhole Saldırısı: Saldırganın ağa iki zararlı düğüm koyarak çevredeki düğümlere bu iki düğümü cazip ve en hızlı iletim sağlayan düğümler olarak gösterip ağ trafiğini üzerine çekmesidir. Trafiği eline geçiren saldırgan istediği gibi veri kullanıp ağı manipüle edebilir [10].



Şekil 5. Grayhole saldırısı

Scheduling Saldırısı: Bu atak türü LEACH protokolünün kurulum aşamasında küme başı düğümleri veri transfer zaman slotu için TDMA planını hazırlarken meydana gelmektedir [12]. Saldırgan düğüm, kendini küme başı düğüm olarak ağ üye düğümlerine tanıtıp tüm düğümlere aynı zaman slotunu vermektedir [12]. Bu durumda çakışmalara, paket düşürme ve kaybolması gibi durumlara yol açmaktadır. Uygulamada bu atağın gerçekleşmesi ya tüm üyelerin paket gönderme zamanının aynı zamana ayarlanması ile ya da ikili üçlü ya da beşli düğümlerin aynı zaman dilimine ayarlanması ile olmaktadır. Bu saldırı türlerinin ortak özelliği herhangi veri sızdırma değil tamamen ya da kısmen sistemi bozma, erişilmez kılma ya da aksatmadır. Zaten bu saldırıların DOS çatısını altında toplamasının nedeni budur.

4. Makine Öğrenmesi

Özellikle son yıllarda her alanda üretilen verinin işlenerek farklı amaçlar doğrultusunda kullanılması verinin önemini bir kez daha göstermiştir. Veriler farklı amaçlar için farklı algoritmalarla işlenerek kullanılmaktadır. Bu veri işleme çeşitli model ve tekniklerle yapılmaktadır. Son yıllarda popülaritesi artan veri işleme modellerinden olan yapay zekâ teknoloji farklı alt dalları ile birçok alanda kullanılmaktadır. Yapay zekâ teknolojisinin bir alt olan makine öğrenmeleri bu çalışmada kullanılmıştır. Makine öğrenmesi, matematiksel ve istatistiksel işlemler ile veriler üzerinden çıkarımlar yaparak tahminlerde bulunan sistemlerin bilgisayarlar ile modellenmesidir [13]. Makine öğrenmesinde, nasıl ki insanlar olaylar karşısında çıkarım yaparak tecrübe kazanırlar hayat ile ilgili yeni şeyler öğrenirler, öylede makineler (algoritmalar) verileri kullanarak eğitilip sonuç odaklı doğru tahminler üretmeye çalışırlar. Tahminler gerçek sonuç ile ya da istenilen sonuç ile ne kadar örtüşür ise öğrenmenin o kadar başarılı olduğu söylenebilir. Eğitimin başarılı olup olmadığı test verileri ile kontrol edilir. Bu test verileri bazen eğitim verilerinden oluşabilirken bazen de farklı veri setleri ile test edilebilir. Makine öğrenmesi matematiksel ve istatistiksel işlemler ile veriler üzerinden çıkarımlar yaparak tahminlerde bulunan sistemlerin bilgisayarlar ile modellenmesidir [13]. Makine, giriş verilerinden öğrendiği edinim ile sonrasında gelen verileri kategorize eder. Bu öğrenme işlemleri literatürde farklı şekilde adlandırılmaktadır. Ancak temelde makine öğrenmeleri gözetimli ve gözetimsiz öğrenme olarak ikiye ayrılır [14]. Bu tamamen geliştirici tercihi ve işlenecek verinin uygunluğuna göre belirlenir. Gözetimli öğrenmede kısaca makine eğitilirken girdilerin ne olduğu ve bu girdilere göre çıktılarında ne olacağı bellidir. İstenilen sonuca ulaşmak için gerekli parametreler değiştirilir. Makineye bu çerçevede gelecek verileri sınıflandırması istenir. K-Nearest Neighbors (KNN), Decsion Tree(J48), Linear Regression, Support Vector Machine (SVM) gözetimli öğrenme modellerindendir [15]. Gözetimsiz öğrenmede ise veriler makinenin kullandığı algoritma çalışma sistemine göre girdi verileri gruplandırılarak eğitim gerçekleşir. Test aşamasında gelecek veriler de bu gruplardan uygun olanına göre isimlendirilir. Eğer mevcut gruptan hiçbirine uygun değilse farklı bir grup altında toplanır. Kısaca gözetimsiz öğrenmede veriler kullanılan algoritmaya göre gruplandırılır. Girdi verileri etiketsizdir. Nasıl bir sonuç çıkacağı makine inisiyatifindedir. Aynı kümede bulunan verilerin farklı açılardan benzerlikleri vardır. Zaten algoritmalarda bu benzerliklerden dolayı küme oluştururlar. Ancak farklı kümelerin benzerlikleri olamamaktadır. Kümeleme, PCA, EM gözetimsiz öğrenme modellerindendir.

4.1. Decision Tree (J48)

Decision Tree algoritması makine öğrenmesinde kullanılan büyük boyuttaki verileri küçük boyutlara belirlenen karar ışığında sınıflandırmasıdır. Sınıflandırma yapısı ağaç dallarına benzediğinden ve sınıflandırmanın karar mekanizmasına göre yapıldığında bu şekilde isimlendirilmektedir. Algoritmada verinin içeriğine göre çok sayıda kategori oluşturabilir. Bu durum avantaj olduğu gibi karmaşıklığa da neden olabilmektedir. Algoritmanın ağaç yapısı görselleştirilebilir. Weka platformunda Decision Tree algoritması çok büyük verilerde boyutunda

dolayı hata verebilmektedir. Ayrıca öğrenmede ezber yaşanabilmektedir. Bu durumda elde edilen sınıflandırma doğruluk yüzdesi anlam ifade etmemektedir. Bunu engellemek için parametre kısıtlamasına gidilebilir.

4.2. Naive Bayes

Çalışma prensibi olasılık üzerine kuruludur. Eğer veri setinin boyutu az ise ve öğrenme içinde az veri ayrılmış ise kullanılması ideal bir algoritmadır. Algoritma her veri örneği için aitlik olasılığı hesaplar ve en yüksek olasılık değerine göre sınıflandırılır. Tembel bir öğrenme modelidir. Temeli Bayes teoremine dayanmaktadır. A değeri sınıf ifade ederken, B değeri veriyi ifade etmektedir.

4.3. Random Forest

Temelinde Decision Tree mantığı yatmaktadır. Ancak Decision Tree modelindeki ezberleme dezavantajı bu modelde giderilmiştir. Ezberlemeyi ortadan kaldırmak için veri setinde, Decision Tree algoritmasından farklı sayılarda seçerek eğitimleri gerçekleştirilmektedir. Decision Tree algoritmaların her birinin sonuçlarına bakılarak en yüksek oy alan ağacın sonucu kullanılmaktadır. Algoritma birden fazla ve farklı boyutlardaki veri seti ile birden fazla Decision Tree ile öğrenme gerçekleştirilip sağlaması yapılarak ezberlemenin önüne geçilmektedir. Decision Tree modelinden daha kompleks bir yapıya sahiptir. Modelin detaylı olarak matematiksel formüle edilmiş hali ilgili çalışmada bulunmaktadır.[16]

4.4. Simple K-Means Clustering

Veri işleme ve analizinde kullanılan bu sade teknik bütün bir veri setini belirli sayıda alt veri gruplara ayırma işlemini gerçekleştirmektedir. Ayrılan verilerden aynı grupta olanlar benzer özellikte farklı grupta olanlar ise farklı özellik taşımaktadır. Küme içi benzerlik en üst seviyede iken kümeler arası benzerlik minimumdur. Büyük boyutlu verilerde kullanılıp hızlı bir şekilde sonuç çıkarabilmektedir. Veri analiz işlemi başlamadan önce K değeri belirlenmiş olmalıdır. K değeri bütün veriyi kendi içinde benzer kaç kümeye ayrılacağını belirlemektedir. Tekrarlı olarak çalıştığı için genellikle yüksek doğruluk oranları ile sonuçlar üretilmektedir. Modelin detaylı olarak matematiksel formüle edilmiş hali ilgili çalışmada bulunmaktadır.[17]

4.5. Expectation Mazimation

İstatistikte kullanılan beklenti maximizasyonu olarak bilinen bu modelde gözlemlenemeyen gizli değişkenlere bağlı istatistiksel modellerin parametrelerinin en büyük olabilirlik ya da en büyük artçıl tahminlerinin bulunması için kullanılan bir yinelemeli arama yöntemidir. Beklenti maksimizasyonu, beklenti (B) adımı ve maksimizasyon (M) adımı olarak iki adımın art arda tekrarlanmasıyla gerçekleşir.

4.6. Canopy

Canopy kümeleme algoritması, veri ön işleme aşamasında sıklıkla kullanılmaktadır. Genellikle K-Means ve Hiyerarşik kümeleme modellerinden önce kullanılmaktadır.

5. Saldırı Tespit Uygulaması

Bu çalışmada, NS 2 ortamında elde edilen WSN-DS veri seti kullanılmıştır. Çalışmada literatürde yaygın olarak kullanılan enerji temelli çalışan ve hiyerarşik yapıya sahip olan LEACH protokolü

kullanılmıştır. Veri seti, 100 düğüm ile gerçekleştirilen uygulamadan elde edilmiştir. Verinin elde edildiği uygulamanın ağ mimari hiyerarşik bir yapıda olduğundan ağ 5 kümeden oluşmaktadır. Her kümeden sorumlu bir küme başı düğüm bulunmaktadır. Düğümler 100*100 metre alanına konuşlandırılmıştır. Paket boyutları 500bytes boyutundadır. Her düğüm için başlangıç enerji değeri 5,50 joule'dür. Çalışma 3600 saniye olarak gerçekleştirilmiştir. Veri seti 374661 kayıttan oluşmaktadır. Veri seti Blackhole, Grayhole, Flooding, Scheduling atakları şeklinde dört tip DOS atak ve normal ağ trafiği içermektedir. Veri setinin boyutu 25.3 MB'dır. Veri seti 19 özellik boyutuna sahiptir. Tablo I de veri setine ait özellikler gösterilmiştir. Özelliklerin açıklamalarına [16]'dan erişilebilir. Ancak elde edilen verilerin benzetim ortamından alındığını belirtmek bu noktada önemlidir. Standart (normal) düğümlerden benzetim ortamında ya da gerçek ortamda Tablo I deki özelliklere ait veriler alınabilir.

Tablo 1. Özellik tablosu.

No	Uygulamada kullanılan 19 özellik listesi	Kısaltmaların Açıklaması
1	ID	Sabit düğüm numarası
2	Time	Düğümün çalışma süresi
3	Is_CH	1 ise küme başı düğüm 0 ise normal düğümdür
4	Who CH	O raunttaki küme başı düğümün numarası.
5	Dis_To_CH	İlgili düğüm küme başı düğüm arasındaki RSSI değeri
6	ADV_S	Küme başı düğümlerin normal düğümlere gönderdiği yayım mesaj numarası
7	ADV_R	Küme başı düğümlerden alınan küme başı yayın mesaj numarası
8	JOIN_S	Normal düğümden küme başına gönderilen katılım istek mesaj numarası
9	JOIN_R	Küme başı düğümün normal düğümden aldığı katılım istek mesaj numarası
10	SCH_S	Düğümlere gönderilen TDMA yayın mesaj numarası
11	SCH_R	Küme başı düğümlerden alınan TDMA yayın mesaj numarası
12	Rank	O düğümün TDMA'daki hiyerarşik derecesi
13	DATA_S	Düğümün, küme başına gönderdiği paket sayısı
14	DATA_R	Küme başı düğümden alınan paket sayısı
15	Data_Sent_To_BS	Baz istasyonuna gönderilen paket sayısı
16	Dist_CH_to BS	Küme başı ile baz istasyonu arasındaki uzaklık
17	Send_code	Kümelere kod gönderimi
18	Consume Energy	Önceki düğümde harcanan enerji miktarı
19	Attacks Type	Belirlenen saldırı tipi

Ancak saldırgan düğümlere ait bilgilerin gerçek ortamlarda alınması mümkün değildir. Saldırıların tespitinde düğümlere ait özellikler ve bu özelliklerdeki sayısal değişimler hayati önem sahiptir. Hepsi olmasa da bu özellikler kullanılarak düğümün saldırgan mı değil mi ya da hangi saldırı tipinin gerçekleştirildiği tespit edilebilir. Eğer düğümün ilgili özelliklerine ulaşılamaz ise böyle bir tespit yapılamaz. Benzetim ortamında düğümlere ait özellikler elde edilse de gerçek ortamda böyle bir senaryo mümkün değildir. Zararlı ya da herhangi bir düğümün bilgisine ulaşamamak düğümler arasında güven değeri açısın bir faktör olmaktadır. Yani baz istasyonu için bilmediği bir düğümün standart bildiği düğüme göre her zaman daha güvenilmezdir. Bu durumda saldırgan düğüm daha

kolay tespit edilebilir. Ya da düğümler Tablo I de belirtilen özelliklere göre değerlendirilerek düğümlerin güvenirlikleri seviyelendirilebilir. Bir saldırı durumunda güvenilirlik değeri en düşük olan saldırgan düğüm olma durumu üzerinde durulabilir. Dolayısıyla geliştiricilerin gerçek ortamda saldırı tespitinde zararlı düğümlere ait bilgilere ihtiyaç duymayan tespit yöntemi geliştirmeleri gerekmektedir. Tablo I deki özellikleri KSA ağ trafik veri setinin türlerine göre kategorize edilmesi işlemi için gözetimli ve gözetimsiz öğrenme gerçekleştirilmiştir. Bu işlem için WEKA 3.8 programı kullanılmıştır. Program aracılığıyla makine öğrenme algoritmalarından gözetimli öğrenme için Decision Tree (J48), Naive Bayes, Random Forest algoritmaları kullanılarak verinin sınıflandırılması gerçekleştirilmiştir. Gözetimsiz öğrenme için ise EM, Simple K-Means, Filtered Clustered, Canopy algoritmaları kullanılarak verinin kümeleme işlemi gerçekleştirilmiştir. Ayrıca test option kısmında verinin nasıl kullanılacağı belirlenmiştir. Split percentage seçeneği ile verinin hangi yüzde ile eğitim ve testin için kullanılacağı belirlenmiştir. Diğer bir seçenek olan 10-fold cross-validation seçeneği işaretlenerek eldeki verinin 10 eşit parçaya ayırılması sağlanmıştır. Her eğitim ve test işlemi için 10 parça ayrı ayrı kullanılarak öğrenme gerçekleştirilmiştir. [18].

Veri seti, her algoritma ile ayrı ayrı analiz edilmiş ve farklı doğruluk yüzdeleri ile ataklar ve normal trafik şeklinde kategorize edilmiştir. Uygulamada öncelikle gözetimli öğrenme algoritmaları sonra gözetimsiz öğrenme algoritması kullanılmıştır. Gözetimli Öğrenme modellerinden Decision Tree algoritması kullanılırken J48 algoritması seçilerek veri işlenmiş ve %99,66 doğruluk ile veri sınıflandırılmıştır. Decision Tree algoritması ile incelenen verinin Tablo 2'de atak türüne göre dağılımı gösterilmiştir.

Tablo 2. Veri seti dağılımı (Decison Tree J48).

J48 Actual Set	**Normal**	**Flooding**	**TDMA**	**Schedule**	**Grayhole**	**Blackhole**
Normal	339719	159	27	150	11	
Flooding	83	3229	0	0	0	
TDMA	478	0	6151	5	4	
Grayhole	120	0	3	14332	141	
Blackhole	2	0	1	76	9970	

J48 sınıflandırma işlemini daha detaylı olarak Tablo 3' de gösterilmiştir. Satırlar, veri setinin gerçek (ham) halini göstermekte, sütunlarda ise makine öğrenme modelinin tespit ettiği veriler gösterilmektedir. F-Measure False Pozitif, True Pozitif, Precision, Recall, Matthews correlation coefficient (MCC), Receiver Operating Charestic Curve (ROC) Area, Precision Recall Curve (PRC) Area gibi doğruluk yüzdesini etkileyen faktörlerde detaylı olarak gösterilmiştir.

Tablo 3. Detaylı sonuçlar (Decision Tree J48).

Class	**TP Rate**	**FP Rate**	**Precision**	**Recall**	**F Measure**	**MCC**	**ROCArea**	**PRC Area**
Normal	0.999	0.020	0.998	0.999	0.998	0.984	0.990	0.998
Flooding	0.975	0.00	0.953	0.975	0.964	0.964	0.995	0.972
TDMA	0.927	0.00	0.995	0.927	0.960	0.960	0.960	0.935
Grayhole	0.982	0.001	0.984	0.982	0.983	0.982	0.997	0.977
Blackhole	0.992	0.00	0.985	0.992	0.988	0.988	0.998	0.993
Average Weight	0.997	0.018	0.997	0.997	0.997	0.983	0.991	0.995

Uygulamada kullanılan diğer gözetimli öğrenme modeli ise Naive Bayes algoritmasıdır. Bu algoritma ile yapılan çalışma sonucu %95,35 oranında doğru sınıflandırma işlemi yapılmıştır. Naive Bayes algoritması ile incelenen veri Tablo 4'de atak türüne göre dağılımı gösterilmiştir.

Tablo 4. Veri seti dağılımı (Naive Bayes).

Naive Bayes Actual Set	Normal	Flooding	TDMA Schedule	Grayhole	Blackhole
Normal	330336	3200	45	6446	39
Flooding	0	3311	0	1	0
TDMA	423	242	5009	41	923
Grayhole	0	311	12	8604	5669
Blackhole	0	0	2	65	9982

Tablo 5'de Naive Bayes modeli için F-Measure False Pozitif, True Pozitif, Precision, Recall, MCC, ROC Area PRC Area değerleri gösterilmiştir. Satırlar veri setinin gerçek (ham) halini göstermekte, sütunlarda ise makine öğrenme modelinin tespit ettiği veriler gösterilmektedir.

Tablo 5. Detaylı sonuçlar (Naive Bayes).

Class	TP Rate	FP Rate	Precision	Recall	F Measure	MCC	ROCArea	PRC Area
Normal	0.971	0.012	0.999	0.971	0.985	0.863	0.980	0.997
Flooding	1.000	0.010	0.469	1.000	0.638	0.861	1.000	0.903
TDMA	0.755	0.000	0.988	0.755	0.856	0.862	0.956	0.834
Grayhole	0.589	0.018	0.568	0.589	0.578	0.561	0.983	0.620
Blackhole	0.993	0.018	0.601	0.993	0.749	0.765	0.992	0.697
Average Weight	0.954	0.012	0.966	0.954	0.957	0.847	0.980	0.971

Random Forest modeliyle yapılan öğrenme %99,72 doğruluk oranı ile gerçekleştirilmiştir. Veri setinin dağılımı Tablo 6'da gösterilmiştir.

Tablo 6. Veri seti dağılımı (Random Forest).

Random Forest Actual Set	Normal	Flooding	TDMA Schedule	Grayhole	Blackhole
Normal	339740	186	24	111	5
Flooding	42	3270	0	0	0
TDMA	476	0	6155	3	4
Grayhole	73	0	0	14439	8
Blackhole	0	0	1	32	10016

Tablo 7. Detaylı sonuçlar (Random Forest)

Class	TP Rate	FP Rate	Precision	Recall	F Measure	MCC	ROCArea	PRC Area
Normal	0.999	0.017	0.998	0.999	0.999	0.985	0.997	0.999
Flooding	0.987	0.001	0.946	0.987	0.966	0.966	1.000	0.994
TDMA	0.927	0.000	0.996	0.927	0.960	0.960	0.983	0.957
Grayhole	0.989	0.000	0.990	0.989	0.989	0.989	1.000	0.999
Blackhole	0.997	0.000	0.991	0.997	0.994	0.994	1.000	1.000
Average Weight	0.997	0.016	0.997	0.997	0.997	0.905	0.997	0.999

Tablo 7' de Random Forest sınıflandırma işlemi sonucu ile ilgili değerlere yer verilmiştir. F-Measure False Pozitif, True Pozitif, Precision, Recall, MCC, ROC Area PRC Area gibi doğruluk yüzdesini etkileyen faktörler de detaylı olarak gösterilmiştir. Satırlar, veri setinin gerçek (ham) halini göstermekte, sütunlarda ise makine öğrenme modelinin tespit ettiği veriler gösterilmektedir.

Uygulamanın ikinci kısmında gözetimsiz öğrenme modelleri ile veri seti incelenmiştir. Gözetimsiz Öğrenme modellerinden ilki Simple K-Means algoritması ile yapılan çalışma sonucu %61,37 oranında doğru kümeleme işlemi gerçekleştirilmiştir. Tablo 8 Simple K-Means ile incelene verinin dağılımı gösterilmiştir.

Tablo 8. Veri seti dağılımı (Simple K-Means)

Simple K Means Actual Set	**Normal**	**No Class**	**TDMA Schedule**	**Grayhole**	**No Class**
Normal	215293	13246	63257	9201	39069
Flooding	0	0	0	3312	0
TDMA	370	9	25	6215	19
Grayhole	0	0	0	14596	0
Blackhole	0	0	0	10049	0

Uygulamada kullanılan diğer gözetimsiz öğrenme modeli olan EM algoritması ile incelenen veri Tablo 9 de atak türüne göre dağılımı gösterilmiştir. Satırlar, veri setinin gerçek (ham) halini göstermekte, sütunlarda ise makine öğrenme modelinin tespit ettiği veriler gösterilmektedir. Bu algoritma ile yapılan çalışma sonucu %69,14 oranında doğru sınıflandırma işlemi yapılmıştır.

Tablo 9. Veri seti dağılımı (EM)

EM Actual Set	**Normal**	**No Class**	**TDMA Schedule**	**Grayhole**	**No Class**
Normal	244329	13634	32204	11066	38833
Flooding	0	0	0	3312	0
TDMA	303	13	88	6215	19
Grayhole	0	0	0	14596	0
Blackhole	0	0	0	10049	0

Uygulamada kullanılan diğer gözetimsiz öğrenme modeli ise Canopy algoritmasıdır. Bu algoritma ile yapılan çalışma sonucu %77,80 oranında doğru sınıflandırma işlemi yapılmıştır. EM algoritması ile incelenen veri Tablo 10 de atak türüne göre dağılımı gösterilmiştir.

Tablo 10. Veri seti dağılımı (Canopy)

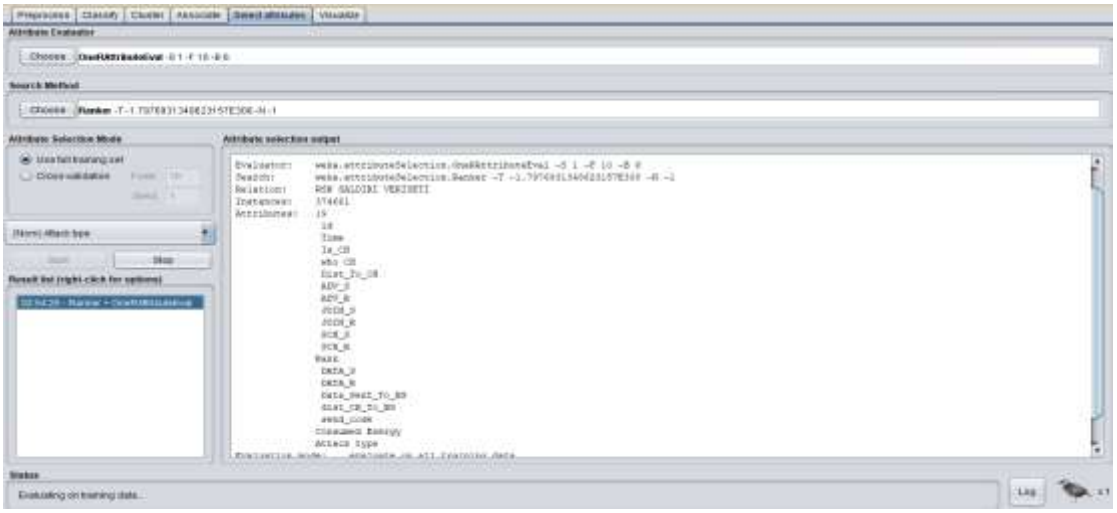
Canopy (Real Set)	Normal	No Class	TDMA Schedule	Grayhole	No Class
Normal	278550	13246	5371	3830	39069
Flooding	0	0	766	2546	0
TDMA	395	9	871	5344	19
Grayhole	0	0	2553	12043	0
Blackhole	0	0	771	9278	0

Son gözetimsiz öğrenme modeli olan Filtered Clusterer ile yapılan çalışmadaki trafiğin türüne göre veri dağılımı Tablo 11'de gösterilmiştir. Satırlar, veri setinin gerçek (ham) halini göstermekte, sütunlarda ise makine öğrenme modelinin tespit ettiği veriler gösterilmektedir. Bu algoritma ile yapılan çalışma sonucu %61,37 oranında doğru kümeleme işlemi yapılmıştır.

Tablo 11. Veri seti dağılımı (Filtered Clusterer)

Filtered Cluster Actual Set	**Normal**	**No Class**	**TDMA Schedule**	**Grayhole**	**No Class**
Normal	215293	39069	63257	9201	13246
Flooding	0	0	0	3312	0
TDMA	370	19	25	6215	9
Grayhole	0	0	0	14596	0
Blackhole	0	0	0	10049	0

19 özellik boyutlu WSN-DS veri seti 3 farklı gözetimli öğrenme modeli ve 4 farklı gözetimsiz öğrenme modeli ile incelenerek uygulama gerçekleştirilmiştir [16]. Ayrıca Gözetimli öğrenme kısmında veriler 10-fold cross-validation yaklaşımıyla incelenmiştir. Makine Öğrenmelerinde incelenen verinin büyüklüğü ile doğru orantılı olmak üzere her veri farklı sayıda özellik boyutuna sahiptir. Fazla sayıda özelliğe sahip verilerde işlem karmaşasını azaltmak için ya da doğruluk değerini (belirli şartlar bağlı kalmak kaydıyla) arttırmak için özellik boyutunu azaltma yoluna gidilebilmektedir. Bu işlem gerçekleştirilirken sonucu en az etkileyen özellikler göz ardı edilerek hem doğruluk değer sonucu optimize edilir hem de karmaşık işlemlerden azaltılarak donanımsal yetersizliklerin önüne geçilir. Bu özellik boyutu azaltma işlemi özelikler kendi aralarında karşılaştırılarak sonucu en çok etkileyenden başlayarak sıralanır. Geliştirici bu işlem sonucuna göre boyut azaltma işlemini gerçekleştirir. Bu çalışmadaki özellik sayısı kaynakları zorlayacak büyüklükte değildir. Yine de bu azaltma işleminin sonucu nasıl etkilediğini görmek adına gerçekleştirilmiştir. Weka'da bulunan hazır araçlardan olan WEKA One-R Attribure Evaluator kullanarak özellik azaltma işlemi gerçekleştirilmiştir.



Şekil 6. Weka platform ara yüzü

Veri seti 19 özellik boyutundan 2 veri boyutuna (Consumed Energy, ADV_S) indirgenmiştir. Tablo 12' de özellikler gösterilmiştir.

Ranking değerine göre en yüksek ilk iki özellik alınarak uygulama adımları tekrar gerçekleştirilmiştir. Sadece bu 2 özellik ile alınan sonuçlar ile 19 özellikle alınan sonuçlar karşılaştırılmış, Tablo 13'de sunulmuş ve sonuca etkisi gösterilmiştir. 2 özellikle yapılan çalışmada

Naive Bayes %95,07 Random Forest %96,99, Decision Tree %97,37, EM %92,66, Simple K-Means %81,08, Filtered Clusterer %81,08, doğruluk sonuçları ile veri kategorize edilmiştir. Veri seti 10-fold cross-validation tekniği ile incelenmiştir.

Tablo 12. Azaltma işleminden sonra kullanılan özellikler.

No	19 özellik arasından seçilen 2 özellik	Yüzde %
1	Consumed Energy	94.19
2	ADV_S	93.089

Tablo 13. 2 ve 19 özellikle yapılan çalışmaların doğruluk yüzdeleri.

ML Modelleri	19 Özellikli çalışmanın doğruluk yüzdeleri %	2 Özellikli çalışmanın doğruluk yüzdeleri %
Decision Tree	99.66	97.37
Random Forest	99.72	96.99
Naïve Bayes	95.35	95.07
Simple K Means	61.37	81.08
EM	69.14	92.66
Canopy	77.80	91.29
Filtered Clusterer	61.37	81.08

6. Sonuç

Bu çalışmada KSA ağlarında ağ katmanı saldırısı olarak bilinen DOS saldırı tiplerinden Flooding, TDMA, Grayhole, Blackhole atak türleri incelenmiştir. Bu atak türlerine ait ağ saldırı trafiği ile normal ağ trafiği içeren veri setiyle uygulama gerçekleştirilmiştir. WSN-DS veri seti, NS 2 benzetim ortamında oluşturulmuştur. Saldırı trafikleri türlerine göre makine öğrenme modelleri yardımı ile incelenerek analiz edilmiştir. Veri seti Random Forest, Decision Tree, Naive Bayes, EM, Filtered Clusterer, Simple K-means, Canopy makine öğrenme algoritmaları ile analiz edilmiştir. Gözetimli öğrenmede en yüksek doğruluk yüzdesine Random Forest ile ulaşılırken gözetimsiz öğrenme modelinde ise en yüksek doğruluk yüzdesine Canopy ile ulaşılmıştır. Tüm öğrenme modellerinde ise Random Forest %99,72 ile en yüksek doğruluk yüzdesi öğrenme işlemini gerçekleştirmiştir. Bu veri işleme ve makine öğrenme modelleri için Weka platformu kullanılmıştır. Ayrıca yukarıdaki yüzdeler özellik boyut azaltma işlemi gerçekleştirilerek 2 boyuta indirgenmiş (Consumed energy ve AODV_S) adımlar tekrarlanarak sonuçlar karşılaştırılmıştır. 2 boyutta ise, Gözetimli öğrenmede en yüksek doğruluk yüzdesine Decision Tree ile ulaşılırken gözetimsiz öğrenme modelinde ise en yüksek doğruluk yüzdesine EM ile ulaşılmıştır. İki boyutlu veri setinde tüm öğrenme modellerinde ise Decision Tree %97,37 ile en yüksek doğruluk yüzdesi ile öğrenme işlemini gerçekleştirmiştir. Tablo 13 incelendiğinde özellik azaltma işleminin gözetimli öğrenme algoritmalarının doğruluk sonuçlarını azalttığı ancak gözetimsiz öğrenme doğruluk sonuçlarını arttırdığı görülmüştür.

Yazarların Katkıları

CO çalışmanın temel halini oluşturdu ve deneysel çalışmaları gerçekleştirdi. MD çalışmayı kontrol ederek gerekli düzeltme ve düzenlemelerin yapılmasını sağladı.

Her iki yazarda makalenin son halini okudu ve onayladı.

Çıkar Çatışması

Yazarlar, çıkar çatışması olmadığını beyan eder.

Kaynaklar

[1]. M. Dener and O. Bay, "TeenySec: a new data link layer security protocol for WSNs", Security and Communication Networks, 2016, 9(18): 5882-5891. Available: 10.1002/sec.1743.

[2]. D. Deif and Y. Gadallah, "An ant colony optimization approach for the deployment of reliable wireless sensor networks", IEEE Access, 2017, 5, 10744-10756. Available: 10.1109/access.2017.2711484.

[3]. Z. Sheng, C. Mahapatra, C. Zhu and V. Leung, "Recent advances in ındustrial wireless sensor networks toward efficient management in IoT", IEEE Access, 2015, 3: 622-637. Available: 10.1109/access.2015.2435000.

[4]. M. Abazeed et al., "A review of secure routing approaches for current and next-generation wireless multimedia sensor networks", International Journal of Distributed Sensor Networks, 2015, 1-22. Available: 10.1155/2015/524038.

[5]. M. Selvi, K. Thangaramya, S. Ganapathy, K. Kulothungan, H. Khannah Nehemiah and A. Kannan, "An energy aware trust based secure routing algorithm for effective communication in wireless sensor networks", Wireless Personal Communications, 2019, 105 (4): 1475-1490. Available: 10.1007/s11277-019-06155-x.

[6]. C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures", Ad Hoc Networks, 2003, 1 (2-3): 293-315. Available: 10.1016/s1570-8705(03)00008-8.

[7]. M. Dener, "Security analysis in wireless sensor networks", International Journal of Distributed Sensor Networks, 2014, 10 (10): 303501. Available: 10.1155/2014/303501.

[8]. W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” in Proceedings of the 33rd IEEE Annual Hawaii International Conference on System Sciences, 1–10, Maui, Hawaii, USA, January 2000.

[9]. S. Tyagi and N. Kumar, “A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks,” Journal of Network and Computer Applications, 2013, 36 (2): 623–645.

[10]. S.Kannan,T.Maragatham,S.Karthik and V. P. Arunachalam,“A study of attacks,attack detection and preevention methods in proactive and reactive routing protocols”International Business Management, 2011, 5(3): 4178-183.

[11]. F. Tseng, L. Chou, & H. Chao, “A survey of black hole attacks in wireless mobile ad hoc networks”. Human Centric Computing Information Sciences, 2011, 1(4): https://doi.org/10.1186/2192-1962-1-4.

[12]. I. Almomani and B. Al-Kasasbeh, "Performance analysis of LEACH protocol under Denial of Service attacks," *2015 6th International Conference on Information and Communication Systems (ICICS)*, 2015, 292-297. doi: 10.1109/IACS.2015.7103191.

[13]. “Makine Öğrenmesi Nedir ?”, Medium, 2020. [Online]. Available:https://medium.com/t%C3%BCrkiye/makine%C3%B6%C4%9Frenmesi-nedir-20dee450b56e. [Accessed: 24- Jun- 2020].

[14]. O. Salman, I. Elhajj, A. Chehab and A. Kayssi, "A machine learning based framework for IoT device identification and abnormal traffic detection", Trans on Emerging Telecomm. Techn, 2019. Available: 10.1002/ett.3743.

[15]. S. Patil and U. Kulkarni, "Accuracy Prediction for Distributed Decision Tree using Machine Learning approach”, 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, 1365-1371. doi: 10.1109/ICOEI.2019.8862580.

[16]. L., Breiman, "Random Forest", Machine learning, 2001, 45, 5–32. Available: https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf

[17]. M. C., Hung, J. Wu, J. Chang and D.L., Yang, "An efficient k-Means clustering algorithm using simple partitioning", Journal of Information Science and Engineering, 2005, 21: 1157-1177.

[18]. I. Almomami, B. Al-Kasasbeh, M. Al-Akhras, " WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks", Journal of Sensors, 2016, 1-16. http://dx.doi.org/10.1155/2016/4731953