Bilgisayar Güvenliği Temelleri Dönem Ödevi (24-25 Güz)
Gerçek Zamanlı Ağ Trafiği Dinleme ve K-Means ile Anomali Tespiti

Ödevin Amacı

Bu ödev, öğrencilerin K-Means algoritmasını kullanarak bir anomali tespit sistemi (Intrusion Detection System - IDS) geliştirmelerini amaçlamaktadır. Bu sistemde öğrenciler, ağ trafiğini gerçek zamanlı izleyerek anomali tespiti yapacak, normal ve anormal veriyi ayırarak potansiyel saldırıları algılayacaktır.

Ödev İçeriği ve Aşamalar

Hazırlanacak dönem ödevi sekiz aşamadan oluşmaktadır. Her bir aşama aşağıda detaylandırılmıştır.

1. Teorik Temellerin Anlaşılması

Öğrenciler, öncelikle anomali tespiti, K-Means kümeleme algoritması ve küme merkezlerine göre anomali tespiti hakkında temel bilgilere sahip olmalıdır:

- **Anomali Tespiti:** Anomali tespiti, veri setinde normal veri akışından sapma gösteren olağandışı durumları belirlemeye yönelik bir süreçtir.
- **K-Means Kümeleme Algoritması:** K-Means algoritması, veri noktalarını belirli merkezlere göre kümeler ve her veri noktasını en yakın merkeze göre gruplar. Bu ödevde, normal trafiği temsil eden kümeler oluşturulacak ve merkezlerden çok uzak olan noktalar anomali olarak işaretlenecektir.
- **Eşik Değeri:** Kümeler oluşturulduktan sonra, küme merkezlerine olan uzaklıklara göre bir eşik değeri belirlenir. Bu eşik değerin üzerinde kalan veriler "anomali" olarak işaretlenir.

2. Veri Seti Seçimi ve Model Eğitimi

Öğrenciler, modelin doğruluğunu artırmak için önceden hazırlanmış veri setlerinden yararlanarak K-Means modelini eğitebilir.

- **Veri Seti Seçimi: NSL-KDD** veri seti kullanılacaktır. Bu veri seti, saldırı ve normal trafik verilerini içerdiği için K-Means algoritması ile anomali tespiti yapılması açısından uygundur.
- **Özellik Seçimi:** Öğrenciler, ağ trafiğindeki saldırı türlerini belirleyebilmek için veri setinde yer alan Kaynak Port, Hedef Port ve Paket Uzunluğu gibi özellikleri seçmelidir. Bu özellikler, normal ve anormal trafiği ayırmak için yeterli bilgi sağlayacaktır.
- **Veriyi İnceleme ve Hazırlık:** Seçilen veri setindeki eksik veya tutarsız veriler temizlenmeli ve yalnızca seçilen özellikler model eğitimine dahil edilmelidir. Ayrıca, verinin boyutları arasında tutarlılığı sağlamak amacıyla normalizasyon uygulanmalıdır.

3. Ağ Trafiği Dinleme ve Veri Toplama

Bu aşamada öğrenciler, gerçek zamanlı olarak ağ trafiğini dinleyecek ve veri toplayacaktır.

- **Ağ Trafiğini Dinleme:** Öğrenciler, Python kütüphanelerinden biri olan Scapy veya benzeri bir araç kullanarak ağdaki TCP paketlerini dinleyecektir. Trafikteki her paket için kaynak IP, hedef IP, kaynak port, hedef port ve paket uzunluğu gibi özellikler kaydedilmelidir. Bu bilgiler daha sonra anomali tespiti için kullanılacaktır.
- **Veriyi Yapılandırma:** Toplanan veriler, makine öğrenimi algoritmaları ile analiz edilebilmesi için Pandas DataFrame formatına dönüştürülebilir. Bu yapı, trafiği analiz etmek ve anormal durumları tespit etmek için kullanılacaktır.

4. Veri İşleme ve Normalizasyon

Toplanan veri üzerinde veri işleme ve normalizasyon adımları uygulanmalıdır.

- **Özellik Seçimi ve Normalizasyon:** Öğrenciler, belirledikleri özellikleri seçerek model için gerekli hale getirirler. Normalizasyon işlemi, veriyi belirli bir aralıkta ölçeklendirerek modelin doğruluğunu artırır ve farklı ölçeklerdeki verilerin kıyaslanmasını kolaylaştırır.
- **DataFrame Yapılandırma:** Elde edilen veriler bir DataFrame yapısına dönüştürülerek model eğitimi için hazırlanır. Bu yapı, özelliklerin daha hızlı işlenmesini sağlar ve model eğitiminde kullanılacak veriyi sadeleştirir.

5. K-Means Modeli ile Anomali Tespiti İçin Model Geliştirme

K-Means modeli, normal trafiği kümeleyerek, bu kümelerden uzak kalan verileri anomali olarak işaretleyecektir.

- **Küme Sayısını Belirleme:** Öğrenciler, modelin kaç adet küme oluşturacağına karar vermelidir. Örneğin, veri iki büyük grupta incelenecekse k=2 olarak belirlenebilir.
- **Model Eğitimi:** K-Means algoritması ile model eğitilir ve veri noktalarının küme merkezlerine olan uzaklıkları hesaplanır. Küme merkezlerinden çok uzak olan noktalar, anomali olarak işaretlenmek üzere belirlenir.
- **Anomali Eşik Değerinin Belirlenmesi:** Küme merkezlerine olan uzaklıklara göre bir eşik değeri belirlenir. En uzak %5'lik veri anomali olarak kabul edilerek eşik değer ayarlanabilir.

6. Gerçek Zamanlı Anomali Tespiti Uygulaması

Bu aşamada öğrenciler, modelin gerçek zamanlı veri üzerinde çalışmasını sağlayarak anomali tespiti yapacaktır.

- **Gerçek Zamanlı Paket İzleme:** Model, her gelen paketi analiz ederek belirlenen küme merkezine olan uzaklığına göre değerlendirir. Küme merkezinden belirli bir

mesafenin üzerinde olan paketler anomali olarak kabul edilir ve sistem tarafından uyarı olarak işaretlenir.

- **Anomali Bildirimi:** Model, anomali olarak sınıflandırdığı paketler için kullanıcıya gerçek zamanlı olarak bir bildirim verir. Bu bildirim, kaynak IP ve hedef IP bilgilerini içerebilir.

7. Saldırı Senaryoları ile Modelin Test Edilmesi

Öğrenciler, modelin doğruluğunu değerlendirmek amacıyla bazı temel saldırı senaryolarını simüle edecektir. Bu senaryolar, modelin belirlenen eşik değerine göre saldırıları tespit edebilme yeteneğini gözlemlemek için önemlidir.

A. SYN Flood Saldırısı

SYN Flood saldırısı, bir hedef IP'ye çok sayıda SYN (bağlantı başlatma) paketi göndererek sistem kaynaklarını tüketmeyi ve hedef sistemi meşgul etmeyi amaçlar. Bu saldırı, öğrencilere modelin SYN Flood gibi saldırıları nasıl tespit ettiğini gözlemleme imkanı sunar.

- **Simülasyon Amacı**: Modelin, yoğun SYN paketleri nedeniyle anormal trafiği nasıl tespit ettiğini görmek.
- **Değerlendirme**: Modelin SYN Flood saldırısını doğru bir şekilde "anomalik" olarak işaretleyip işaretlemediğini gözlemleyin.

B. Port Tarama Saldırısı

Port tarama, sistemdeki açık portları belirlemek için farklı portlara hızlı bir şekilde bağlantı talebi gönderilen bir saldırı türüdür. Bu senaryo, modelin çeşitli portlara hızlıca yapılan bağlantı taleplerini anomali olarak algılayıp algılayamayacağını gözlemlemek için kullanılabilir.

- **Simülasyon Amacı**: Farklı portlara yapılan hızlı bağlantı taleplerinin anomali olarak tespit edilip edilmediğini değerlendirmek.
- **Değerlendirme**: Model, sık ve hızlı bağlantı taleplerini anomali olarak işaretlemelidir. Böylece, port tarama saldırılarını algılama yeteneği değerlendirilmiş olur.

C. ICMP Flood (Ping Flood) Saldırısı

ICMP Flood saldırısı, hedefe çok sayıda ICMP (ping) paketi göndererek ağ üzerinde tıkanıklık yaratmayı amaçlayan bir saldırıdır. Bu saldırı modeli, modelin yüksek sayıda gelen ICMP paketlerini anomali olarak algılayıp algılamadığını test etmek için idealdir.

- **Simülasyon Amacı**: Hedefe yönlendirilmiş çok sayıda ICMP paketi modelin tespiti için nasıl değerlendirildiğini gözlemlemek.
- **Değerlendirme**: Modelin yoğun ICMP paketlerini "anomalik" olarak tanıyıp tanımadığı incelenir. Ping Flood saldırılarının model tarafından başarılı şekilde tespit edilmesi beklenir.

8. Sonuçların Değerlendirilmesi ve Rapor Hazırlama

Son aşamada, öğrenciler geliştirdikleri sistemin performansını analiz eden bir rapor hazırlamalıdır. Bu raporda:

- **Anomali Tespiti Başarı Oranları**: Modelin doğruluk oranları, yanlış pozitif ve yanlış negatif oranları analiz edilmelidir.
- **Saldırı Tespiti**: SYN Flood, Port Tarama ve ICMP Flood gibi saldırıların başarılı bir şekilde tespit edilip edilmediği gözlemlenmelidir.
- **Geliştirme Önerileri**: Modelin doğruluğunu artırmak için yapılabilecek iyileştirme önerileri belirtilmelidir.

Ödev Kriterleri

1. Doğrulama ve Çalışma (70 Puan): Her bir aşama kendi içerisinde puanlanacak ve öğrenci kodu modifiye edebilmesi sonucunda puan alacaktır.
2. Modülerlik ve Yeniden Kullanılabilirlik (10 Puan): Kod modüler olmalı ve farklı parçalar bağımsız olarak çalışabilmeli.
3. Kod Kalitesi ve Okunabilirlik (10 Puan): Kod anlaşılır olmalı ve uygun bir belgelendirme içermeli. Kod düzeni ve stili iyi olmalı.
4. Proje rapor (10 Puan)

Ödevin son teslim Tarihi: Finallerden bir önceki hafta Pazartesi günüdür. Salı günü savunmalar alınacaktır.