**Bölüm 1: Veri Madenciliği Tekniği Seçimi**

**1.** **Veri kümesini analiz etmek için bir veri madenciliği tekniği seçin ve seçiminizi gerekçelendirin. Sınıflandırma, Kümeleme, Anomalilerin Tespiti veya İlişki Kuralları Madenciliği gibi teknikleri düşünün**.

=> Veri setindeki benzer özelliklere sahip veri noktalarını gruplamak, belirli kalıpları ve davranışları tanımlamak için kullanışlı olabilir. Bu, siber güvenlik tehditlerini tespit etmek ve izole etmek için faydalı bir strateji olabilir. Kümeleme tekniklerinden ise K-Means algoritması seçtik.

K-Means: K-ortalama kümeleme ya da K-means kümeleme (K-means clustering) yöntemi N adet veri nesnesinden oluşan bir veri kümesini giriş parametresi olarak verilen K adet kümeye bölümlemektir. Amaç, gerçekleştirilen bölümleme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin ise minimum olmasını sağlamaktır.

**2.** **Seçiminizin arkasındaki mantığı ve siber güvenlikle ilgisini ayrıntılı olarak açıklayarak beklenen sonuçları detaylandırın.**

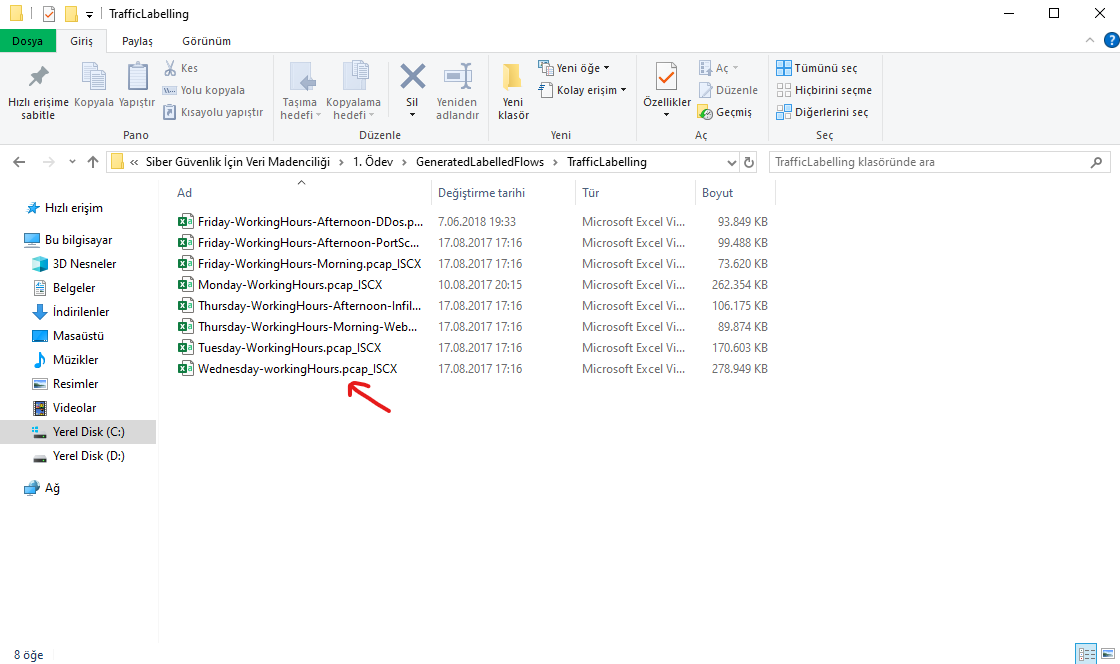
=> Kümeleme, benzer özelliklere sahip veri noktalarını bir araya getirerek anlamlı gruplar oluşturur. Kümeleme, siber güvenlik alanında potansiyel tehditleri ve anormallikleri tespit etmek için etkili bir yöntemdir. Benzer özelliklere sahip ağ trafiği, saldırı girişimleri veya diğer güvenlik olayları genellikle aynı kümelere aittir. Bu nedenle, kümeleme algoritmaları kullanarak veri setini kümelere bölmek, güvenlik tehditlerini sınıflandırmak ve belirli özelliklere odaklanmak için bir yol sağlayabilir.

**Bölüm 2: Güvenlik Tehdidi Tespiti Uygulama**

**1. Seçilen veri madenciliği tekniklerini kullanarak veri seti içindeki potansiyel güvenlik tehditlerini veya anormallikleri tespit edin ve sınıflandırın.**

Veri setimizde yer alan GeneratedLabelledFlows\TrafficLabelling\ Wednesday-workingHours.pcap\_ISCX veri setimiz için K-Means algoritması kullanılarak kümeleme yapıldı. K-Means kümeleme algoritması kullanılarak, ' Wednesday-workingHours.pcap\_ISCX ' veri setindeki benzer ağ trafiği özelliklerine sahip veri noktalarını altı farklı küme (cluster) içinde grupladık.

Cluster değerleri 6 farklı kümede gruplamamızın sebebi GeneratedLabelledFlows\TrafficLabelling içerisindeki 8 adet veri setinden **Wednesday-workingHours.pcap\_ISCX** isimli veri setini kullanmamızın sebebi kümeleme olarak en çok unique (Label sütunu baz alındı) değeri içeren veri setiydi. Yani gözle görülür bir kümeleme yapmak istiyorduk.



Bununla ilgili gerekli unique hesaplatan kodu yazdık.(unique.py github’a yüklendi). Aşağıdaki resimdede görüldüğü üzere her veri setinin ‘Label’ sütununda kaç farklı değer olduğu bulundu(6).

metin, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

BENIGN: iyi huylu

DoS slowloris: DDoS saldırılarının bir türüdür ve web sunucularına karşı gerçekleştirilir.

DoS Slowhttptest: slow tekniğini kullanarak web sunucularına dos saldırıları yapmaya yaran bir araçtır.

DoS Hulk: DoS Hulk saldırısı özellikle ağ kaynaklarını, genellikle işlemci gücünü veya bant genişliğini aşırı şekilde kullanarak sistem performansını düşürmeyi hedefler.

DoS GoldenEye: DoS GoldenEye saldırısı, genellikle bir bot ağı (botnet) kullanılarak gerçekleştirilir. Saldırganlar, bu botnet içinde kontrol ettikleri birçok cihazı koordine ederek hedefe yoğun trafik gönderirler. Bu trafik, hedef sistem üzerinde aşırı yük oluşturarak normal işleyişini engeller ve kullanıcıların hizmet almasını önler.

Heartbleed: Heartbleed saldırısı, OpenSSL kriptografik kütüphanesindeki bir güvenlik açığından kaynaklanan bir saldırı türüdür. Bu saldırı, TLS (Transport Layer Security) protokolü üzerinde çalışan sunucularda meydana gelir.

Ardından veri setine K-Means algoritması uyguladığımız kodda(vize.py) aşağıdaki kilit noktalar belirlendi.Bu kodun amacı her bir kümenin temsil ettiği benzer özellikler, aynı türden ağ trafiği davranışlarını gösteren veri noktalarını içermektedir veri seti içerisinde bulunan saldırı anındaki bazı bilgiler olarak şunlar; 'Flow ID', 'Source IP', 'Destination IP', 'Timestamp', 'Label' değerlerine göre nitelikler seçildi.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

ve K-Means niteliği olarak Label sütunu baz alındı.



**4. Tespit edilen tehdit türleri hakkında içgörüler sunun ve bunların siber güvenlik için etkilerini tartışlayın.**

Tespit edilen altı küme, belirli güvenlik tehditleri veya anormallikleri temsil eder. Bu kümeleme sonuçları, belirli bir küme içindeki veri noktalarının birbirine benzer özelliklere sahip olduğunu gösterir. Örneğin, belirli bir küme içindeki ağ trafiği yoğunluğu, iletişim protokolleri veya bağlantı noktaları gibi özellikler benzer olabilir. Bu durum, aynı türden bir saldırı veya anormallik tespit edildiğinde, bu kümenin incelenmesini sağlar.

**Bölüm 3: Sonuçlar**

**5. Analiz sonuçlarınızı özetleyin ve uygulanan veri madenciliği tekniklerinin doğruluğunu ve etkinliğini tartışın.**

Özetle, 6 farklı Cluster değeri renk olarak (BENIGN DoS slowloris,DoS Slowhttptest,DoS Hulk,DoS GoldenEye,Heartbleed) değerlerine yakın olan/benzer farklı veriler 6 farklı renkle kümelendi(K-Means) ve grafik olarak aşağıdaki gibi gösterildi.

metin, ekran görüntüsü, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

**6. Bulgularınıza dayanarak, ağ güvenliğini, tehdit tespitini veya olay yanıtı stratejilerini geliştirmek için önerilerde bulunun.**

Küme Analizi ile Sınıflandırma: Belirlenen altı küme, ağ trafiğinin benzer özelliklere sahip gruplara ayrıldığını göstermektedir. Her bir küme, belirli bir türdeki ağ etkinliğini temsil edebilir. Bu kümeleme sonuçlarını kullanarak, ağ trafiğini daha iyi anlamak ve farklı etkinlik türlerini tanımlamak için sınıflandırma algoritmaları geliştirebiliriz.

Anomalilerin Tespiti: Küme analizi, normalliğin ötesine geçen anormal ağ etkinliklerini belirleme konusunda yardımcı olabilir. Oluşturulan küme merkezlerinden uzakta bulunan örnekler, anormal davranışları işaret edebilir. Bu anormallikleri izlemek ve tanımlamak için belirli bir küme içindeki ve dışındaki örnekleri değerlendirebiliriz.

Eğitim ve İyileştirme: Kullanılan modeli daha da iyileştirmek için daha fazla etiketli veri kullanarak modeli eğitebiliriz. Bu, modelin belirli tehdit türlerini daha iyi tanımlamasına yardımcı olabilir. Ayrıca, belirlenen küme sayısını ve özelliklerini gözden geçirerek daha optimize edilmiş bir model elde etmeye çalışabiliriz.