**Bitirme Projesi Önerisi**

**Malware nedir?**

Malware, ingilizcede Malicious Software’in kısaltmasıdır. Türkçede kötü amaçlı bütün yazılımlara verilen genel bir addır. Kötü amaçlı yazılım ise kişisel bilgilerinizi çalan, parola, kullanıcı adı gibi size özel verileri ele geçirmeyi hedefleyen, bilgisayarınızın çökmesi için uğraşan, bilgisayarınızı yavaşlatan, bilgisayarınızı sizden izinsiz başkalarının kontrolüne açan uygulamalar bütünüdür. Bu gruba virüs, trojan, Backdoor, worm (solucan) ve adware (reklam yazılımı) gibi kötü amaçlı uygulamalar girer.

**Proje Amacı**

Biz bu projede halihazırda geçmişteki verileri kullanarak bir makinenin gelecekte kötü bir yazılıma maruz kalıp kalmayacağını tahmin etmeyi amaçlıyoruz. Bu tahminin gerçekleştirilmesi için gereken algoritmalar tarafımızca makina öğrenmesi metotları kullanılarak yapılacaktır. Gerekli veri seti ise Microsoft’un kötü amaçlı yazılım tahmin yarışmasından tedarik edilecektir.

**Microsoft Kötü Amaçlı Yazılım Tahmin Yarışması**

Bu yarışmanın amacı, bir Windows makinesinin, bu makinenin farklı özelliklerine dayalı olarak, çeşitli kötü amaçlı yazılım aileleri tarafından etkilenme olasılığını tahmin etmektir. Bu özellikleri ve makine bulaşma (infection) bilgilerini içeren telemetri verileri, Microsoft'un uç nokta koruma çözümü Windows Defender tarafından toplanan kalp atışı (heartbeat) ve tehdit raporları birleştirilerek üretildi. Yarışmanın internet linki buradadır: https://www.kaggle.com/c/microsoft-malware-prediction/data

**Veri Kümesi İçeriği**

Bu veri kümesindeki her satır, benzersiz bir "MachineIdentifier" ile tanımlanmış bir makineye karşılık gelir. "HasDetections", temel gerçektir (ground truth) ve makinede kötü amaçlı yazılım tespit edildiğini gösterir. "Train.csv"deki bilgi ve etiketleri kullanarak, "test.csv"deki her makine için "HasDetections" değeri tahmin edilecektir.

Kötü amaçlı yazılım tespiti bir zaman serisi (time-series) sorunudur ancak yeni makineler, çevrimiçi ve çevrimdışı olan makineler, yamalar alan makineler, yeni işletim sistemleri alan makineler vb. tarafından daha karışık bir hale gelmiştir. Kabaca bölünmüş halde, yukarıda belirtilen komplikasyonlar ve örnekleme gereklilikleri, çapraz onaylama, genel ve özel puanlarınız arasında kusurlu bir anlaşma görebileceğiniz anlamına gelebilir! Ek olarak, bu veri kümesi Microsoft müşterilerinin vahşi ortamındaki makinelerini temsil etmemektedir; kötü amaçlı yazılım makinelerinin çok daha büyük bir bölümünü içerecek şekilde örneklenmiştir.

Gerçekleştirmek istediğimiz projemizle aynı zamanda Türkiye Siber Güvenlik Kümelenmesi proje yarışmasına da katılmak istiyoruz: <https://proje.siberkulupler.com/>

Saygılarımızla,

Furkan Karakaş & Mete Yıldırım