Merhaba arkadaşlar!

Bugün sizlerle "Kötü Amaçlı URL Tespiti" projesinin heyecan verici sonuçlarını paylaşmak istiyorum. Evet, siber güvenlik gibi büyük bir konu hakkında konuşacağız!

Veri Seti ve Macera Başlıyor!

İşe, "Malicious URLs Dataset" veri setini inceleyerek başladık. Bu veri seti, tamı tamına 651.191 tane URL'den oluşuyor. Zararsız, tahrifat, kimlik avı ve kötü amaçlı URL'ler gibi korkutucu türler içeriyor. Hani o kötü amaçlı virüslerden, kötü niyetli yazılımlardan bahsediyoruz. Peki, bu veri setinde neler döndü?

Verileri Anlamak ve Temizlemek

Öncelikle "pandas" kütüphanesini çağırdık ve veri setini çekip elde ettik. Sonra "veri.head()" diyerek dosyanın ilk 10 satırına bakarak işe başladık. Ardından "veri.info()" ile de verilerin ne durumda olduğuna dair bir göz atış yaptık.

Eksik veya aynı olan verileri temizlemek de cabası. Bu kısmı "Ben buradayım!" diye düşünüp eksikleri doldurduk ve aynı olanları çıkardık. Neyse ki, veriler biraz daha temiz hale geldi!

Model Oluştur ve Sonuçları Değerlendir!

Ve işte geldik heyecanlı kısma! Verileri "train" ve "test" olarak ayırdık. İşte burada "Logistic Regression" adında bir süper kahraman modelimiz var. Ona verileri öğretip, bu kötü amaçlı URL'leri yakalama görevini veriyoruz.

Ve tahmin et ne oldu? Doğruluk oranımız tam da "0.946429685550287"! Evet, yanlış duymadınız. Yüzde 94 civarında bir doğrulukla, kötü niyetli URL'leri neredeyse tüm zamanlarımızda yakalıyoruz.

Sonuç Olarak...

Arkadaşlar, bugün gerçekten heyecan verici bir yolculuğa çıktık. Siber güvenlik dünyasında bize karşı gelen tehditlerden biraz daha haberdar olduk ve nasıl mücadele edebileceğimizi gördük. "Kötü Amaçlı URL Tespiti" projesi sayesinde verileri anlama, temizleme ve modele besleme gibi adımları öğrendik.

Herkesin güvende kalması bizim için çok önemli ve böyle projeler, gelecekte daha güvenli bir dijital dünya için büyük bir adım olabilir. Umarım bu projeden öğrendiğimiz bilgileri, günlük hayatta da kullanabiliriz.