**Konu : Kötü Amaçlı URL'ler**  
**AISECLAB Ödev - Elif Parlak / Machine Learning**

**Hazırlayan: Ceren Akbulut**

**Giriş**

Bu sunumda Kötü Amaçlı URL'ler Projesi'ni adım adım nasıl tamamladığımı ve elde ettiğim sonuçları anlatacağım.

**Proje Amaçları**

* Verilen "Malicious" veri seti üzerinde, kötü amaçlı URL'leri tespit etmek için bir makine öğrenimi modeli oluşturmak.
* Veri ön işleme, model seçimi, eğitim ve değerlendirme adımlarını uygulamak.
* Model başarı oranlarını değerlendirmek ve sonuçları sunmak.

**Veri Kümesi İncelemesi**

* Kullanılacak veri kümesinin özellikleri (satır ve sütun sayıları, etiket dağılımı, vb.) incelendi.
* Veri kümesinde boş veriler olup olmadığına bakıldı.

**Veri Seti**

* Bu veri kümesi toplam 651.191 URL'den oluşuyor bu URL türleri de şu şekilde gösterilebilir:

1. Zararsız veya Güvenli URL: 428.103
2. Tahrifat URL'si: 96.457
3. Kimlik Avı URL'si: 94.111
4. Kötü Amaçlı URL: 32.520

**Veri Ön İşleme**

* URL ve Etiket (Tür) sütunları kullanıldı.
* TF-IDF vektörleştirici ile metin verileri sayısal özelliklere dönüştürüldü.

**Veri Kümesi Bölünmesi**

* Veri kümesini eğitim ve test olmak üzere ikiye ayırdım.

**Model Oluşturma**

* Lojistik Regresyon modeli kullanıldı.
* Eğitim verileri üzerinde modeli eğitildi.
* Modelin performansı test verileri ile değerlendirildi.

**Sonuçlar ve Başarı Oranları**

* Modelin başarı oranı değerlendirildi.
* Model Başarı Oranı: 0.9461221293161035

**Sunumun Sonu**

Bu sunumda, Kötü Amaçlı URL'ler projesini tamamlamak için veri ön işleme, model oluşturma ve başarı oranlarını değerlendirme adımlarını uyguladım.