

Siber Zorbalık Nedir?

Siber zorbalık; sosyal medya, mesajlaşma uygulamaları ve diğer dijital platformlar üzerinden kişilere zarar vermek amacıyla yapılan tehdit, aşağılama, dışlama veya taciz gibi eylemlerdir.

Amacımız

Projemizin temel amacı, sosyal medya üzerindeki Türkçe paylaşımlarda yer alan siber zorbalık içeriklerini otomatik olarak tespit edebilen bir makine öğrenmesi modeli geliştirmektir.

Veri Setimiz

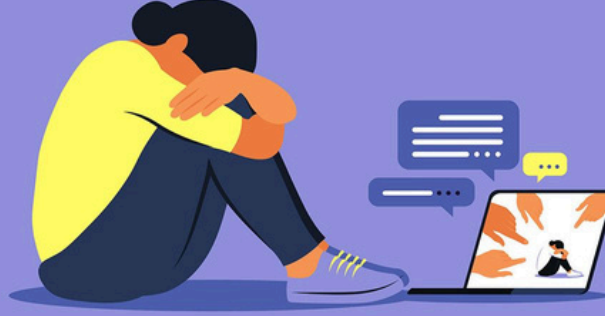
Kaggle'da bulunan pozitif ve negatif etiketli 11.000+ sosyal medya paylaşımından oluşan 'Türkçe Sosyal Medya Paylaşım Veri Seti' başlıklı veri setini kullanılmıştır.



Nasıl Yaptık?

11.000+ etiketli Türkçe sosyal medya paylaşımını analiz ettik. Metinleri TF-IDF yöntemiyle sayısal verilere dönüştürdük. Ardından SVM, Lojistik Regresyon ve Random Forest algoritmalarını uygulayarak modellerin performansını karşılaştırdık. En yüksek başarıyı %84 doğruluk ve 0.87 F1-score ile Logistic Regression modeli verdi. LDA algoritmasıyla negatif içerikleri tematik olarak kümeledik ve çeşitli kategorilere ayırdık.

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE SİBER ZORBALIK TESPİTİ



“Dijital kelimeler, fiziksel darbeler kadar yaralayıcı olabilir.
Siber zorbalık görünmez ama gerçektir.”

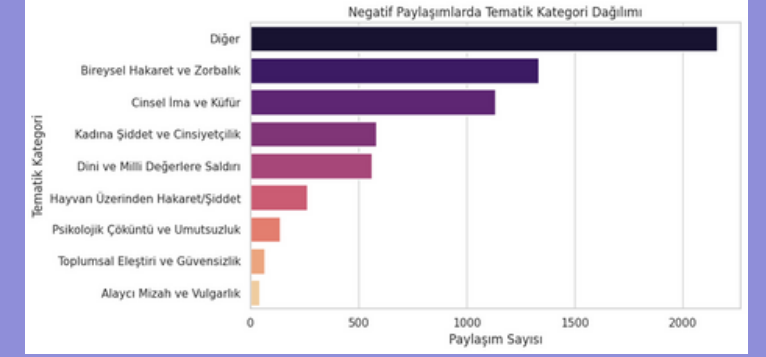
Zorbalık Kendini Tekrarlıyor!

Vocabulary Richness (sözcük çeşitliliği) ölçümlerimiz, negatif içeriklerin daha sınırlı ve kalıplaşmış bir kelime dağarcığıyla üretildiğini ortaya koydu. Bu bulgu, siber zorbalığın dilsel olarak da tahmin edilebilir bir yapıya sahip olduğunu destekliyor.



Zorbalık Temalarla Kendini Ele Veriyor!

Negatif içerikli paylaşımlar üzerinde yaptığımız analizde, içeriklerin büyük çoğunluğunun üç ana tema etrafında toplandığını keşfettik: bireysel hakaret, cinsiyetçilik ve küfür. Bu durum, siber zorbalığın yapısal ve tekrarlayan kalıplar içerdiğini gösteriyor.



Saldırgan Dil Daha Kısa Konuşuyor!

Yaptığımız kelime ve karakter uzunluğu analizleri sonucunda, negatif içeriklerin pozitiflere göre belirgin şekilde daha kısa yazıldığını gördük. Bu durum, saldırgan mesajların genellikle daha doğrudan ve az sözcükle iletildiğini gösteriyor.

