

# ABRA MUHARA

PROJE ÖZETİ: **Bu proje, kullanıcıların sisteme yükledikleri Türkçe kitapların PDF dosyalarını analiz ederek kitapların uygun yaş aralıklarını, içerilerinde kaç uygunsuz cümle ve kelime geçtiği vb. bilgileri belirlemeyi ve bunları kullanıcıya bildirmeyi amaçlayan bir uygulamadır.**

GİRİŞ: **Ele alınan problem, özellikle çocuk ve genç okurlar için uygun kitapların seçilmesi sürecindeki eğitimcilerin ve yetişkinlerin yaşadıkları belirsizliktir. Kitapların içerdiği dil, temalar ve uygunsuz öğeler, yaş gruplarına göre farklı etkiler yaratmaktadır. Bu bağlamda, öğretmenler, kütüphaneciler, ebeveynler ve okurlar için kitapların içeriklerinin değerlendirilmesi ve uygunluk derecelerinin belirlenmesi oldukça önem arz etmektedir.**

YÖNTEM: **PDF formatındaki kitap, metin çıkarma işlemi ile analiz edilir. BERTürk modeli kullanılarak uygunsuz cümleler ve ortalama uygunsuzluk olasılığı hesaplanır ve Excel tablosuna eklenir. Zemberek ve regex kütüphaneleri ile kelimeler köklerine ayrılır ve uygunsuz kelimeler tespit edilerek Excel'e eklenir. Catboost model tipini Optuna ile optimize ettik ve %95 doğruluk elde ettik. Python'da gui yazarak lokalde çalıştırdık ve fast-api + Hugging Face Space ile kolay kullanılabilir bir API sitesi yaptık.**

SONUÇLAR:

Yaş kategori konusunda Catboost modelinin doğruluğu %95,65 ile birinci oldu. Uygunsuzluk modelinde Fine-tuned Berturk modeli %93 ile birinci oldu.

KAYNAKLAR:

**ATEŞMAN, Ender. (1997). Türkçe’de okunabilirliğin Ölçülmesi. A.Ü. Tömer Dil Dergisi, sayı:58,s.171-174.**

**Cetinkaya, B. (2008). Türkçe Metinlerde Okunabilirlik Analizi.**

**Flesch, R. (1948). A New Readability Yardstick. Journal of Applied Psychology, 32(3), 221-233. https://doi.org/10.1037/h0057532**