



**T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
FEN FAKÜLTESİ
BİLGİSAYAR BİLİMLERİ BÖLÜMÜ**

BİTİRME PROJESİ

**YEREL BELGELERDE HASSAS VERİ GİZLİLİĞİ:
DOĞAL DİL İŞLEME TABANLI HİBRİT BİR
ANONİMLEŞTİRME YAKLAŞIMI**

**Fatemehalsadat FATTAHI
Govher HYDYROVA**

Danışman: Prof. Dr. EFENDİ NASİBOĞLU

**MAYIS 2026
İZMİR**

ÖZET

YEREL BELGELERDE HASSAS VERİ GİZLİLİĞİ: DOĞAL DİL İŞLEME TABANLI HİBRİT BİR ANONİMLEŞTİRME YAKLAŞIMI

Fatemehalsadat FATTAHI

Govher HYDYROVA

Danışman: PROF. DR. EFENDİ NASİBOĞLU

Kişisel verilerin dijital ortamlarda hızla yaygınlaşması, bu verilerin korunmasına yönelik yasal düzenlemelerin etkin biçimde uygulanması ihtiyacını beraberinde getirmektedir. Bu çalışmada, Kişisel Verilerin Korunması Kanunu (KVKK) kapsamında kişisel verilerin metin içerisinde otomatik olarak tespit edilmesi ve anonimleştirilmesine yönelik hibrit bir sistem geliştirilmiştir. Geliştirilen sistem, kural tabanlı ve derin öğrenme tabanlı yaklaşımları bir arada kullanan çok katmanlı bir mimari üzerine inşa edilmiştir. Yapısal olarak tanımlanabilir kişisel veriler; Türkiye Cumhuriyeti kimlik numarası, IBAN, telefon numarası ve e-posta adresi gibi kategoriler için düzenli ifade (regex) motoru ile tespit edilirken, kişi adı, kurum adı ve yer adı gibi bağlama duyarlı varlıklar için Türkçe metinler üzerinde önceden ince ayar yapılmış bir BERT modeli kullanılmaktadır. Her iki bileşenden elde edilen tespitler, veri türü önceliklerine dayalı bir varlık birleştirme algoritması aracılığıyla birleştirilmekte ve olası çakışmalar giderilmektedir. Sistem ayrıca Wikipedia API entegrasyonu ile tespit edilen varlıkların kamuya açık olma durumunu kontrol eden ve KVKK ilkelerine dayalı kural tabanlı bir değerlendirme mekanizması ile her varlık için maskeleyen kararı üreten bir katmana sahiptir. Bununla birlikte sistem görüntü verilerinden metin çıkarımı için optik karakter tanıma (OCR) yaklaşımını da desteklemektedir. Kullanıcıya düz metin, PDF, DOCX ve görüntü formatlarında girdi imkânı sunan sistem, Streamlit tabanlı bir web arayüzü üzerinden çalışmaktadır. Geliştirilen sistem, farklı formatlardaki belgelerde kişisel verileri tespit ederek anonimleştirilmiş çıktılar üretebilmektedir. Bu çalışma, kural tabanlı ve model tabanlı yaklaşımların bütünleştirildiği hibrit bir mimari ile Türkçe metinlerde KVKK uyumlu veri anonimleştirmenin gerçekleştirilebileceğini göstermekte ve bu alanda yapılacak gelecekteki çalışmalar için bir temel oluşturmaktadır.

Anahtar Kelimeler: Kişisel veri anonimleştirme, KVKK, varlık ismi tanıma, BERT, hibrit sistem