

Bil 476 Veri Madenciliği Proje Önerisi

Proje İsmi: Türkçe NER Araçlarını Birleştirme Projesi

Proje Tanımı: Türkçe için Varlık İsmi Tanıma kütüphanelerini basit bir kullanıcı arayüzü ile bir araya getirme ve NER modelinin eğitim, test ve tahminleme işlemlerinin hızlandırılması ve kolaylaştırılması

Named Entity Recognition (Varlık İsmi Tanıma), verilen bir metindeki kelimelerin varlık türlerini tanıma işlemine denir. Buradaki varlık türleri metnin kategorisine göre farklılıklar gösterebilir. Doğal dil işleme sistemlerinde metnin çeşitli öznitelikleri çıkarılmaktadır. Bu özniteliklerin en önemlilerinden biri varlık türü tanımadır. Çünkü varlık türleri, cümleye anlam katan temel yapı taşlarındandır. Varlık türü tanıma işlemi her dil için özel olarak yapılmalıdır. Bu sebeple başka dillerde çok gelişmiş sistemler olsa da Türkçe için ayrıca veri setleri hazırlanıp, modeller üretilmelidir. Türkçe varlık türü tanıma işlemi için kısıtlı sayıda kütüphane bulunmaktadır.

Bu projenin amacı, Türkçe için geliştirilmiş varlık türü tanıma araçlarını, basit bir kullanıcı arayüzünde bir araya getirmektir. Bu arayüzde, veri seti okuma, okunan veri setini ön işleme, işlenmiş veri setinden model üretme, üretilmiş modeli kullanarak yeni bir verinin varlık türünü tahmin etme olarak dört ayrı bölüm bulunacaktır. Yapılacak arayüz kullanıcıya, aynı anda tek veri seti ile farklı modeller eğitme, ve farklı modellerden farklı tahminleri yapabilme olanağı sağlayacaktır. Kullanıcıya zaman konusunda fayda sağlayacaktır. Aynı zamanda önceden kullanılan veri setlerinin ve üretilen modellerin kaydedilmesine olanak sağlayacaktır. Kullanıcının süreç kontrolünü kolaylaştıracaktır.

Yapılacak olan bu uygulamada;

- Kullanıcı, eğitim için kullanılacak verileri JSON formatında yükler.
- Yüklenen veriyi eğitimde kullanılacak hale getirmek için ön işleme sokar.
- Ön işlenen veriyle istediği algoritmaları kullanarak model üretir.
- İsteddiği modelleri kullanarak yeni gelen veri için tahminleme yapar.



Görsel 1: Örnek Grafik Kullanıcı Arayüz Tasarımı

Proje Planı:

- Kullanıcı arayüz yapısını kurma
- Kurulan arayüzde tasarım iyileştirmeleri
- Arayüz fonksiyonlarının tamamlanması
- Belirlenecek kütüphanelerin uygulamaya eklenmesi
- Model üretme ve tahminleme işlemlerinin yapılması
- Verilerin ve modellerin yedeklenebilir hale getirilmesi
- Uygulamaya executable veya setup oluşturulması
- Uygulama dokümantasyonu ve proje raporu hazırlama
- Kod düzenlemesi ve yorum satırları ekleme

Proje Üyeleri:

M. Said Zengin

Merve Mallı

Yağmur Albayır