

Kocaeli Üniversitesi, Teknoloji Fakültesi
Bilişim Sistemleri Mühendisliği Bölümü
TBL331: Yazılım Geliştirme Laboratuvarı I
2024-2025 Güz
Proje I

YAPAY ZEKA UYGULAMARINA YÖNELİK VERİ TOPLAMA ve VERİ ÖN İŞLEME UYGULAMALARI

Proje son teslim tarihi: 15 Kasım 2024 Cuma Saat: 17.00

GİRİŞ

Bu projede, dört farklı veri tipinden biri (ses verisi, zaman serisi verisi, görüntü verisi veya metin verisi) ile ilgili veri toplamanız ve ön işleme adımlarını gerçekleştirmeniz beklenmektedir. Hangi veri ile çalışacağınız otomatik olarak belirlenmiştir. Örnek platformlar ve veri tiplerine dair örnekler aşağıda verilmiştir.

Veri tipindeki her bir sınıf için minimum 5000 kayıt toplamanız gerekmektedir. Örneğin görsellerden kedi-köpek ayrımını yapabilen sınıflandırıcı yapay zeka modeli geliştirmek için toplamda 10000 adet görsel toplamanız gerekir. **Toplanan verilerin, bir sonraki Yazılım Geliştirme Laboratuvarı projesinde kullanılması zorunludur.**

PROBLEM

İnternet üzerindeki büyük veri kaynakları, çeşitli projeler için ham veri sağlamaktadır. Ses tanıma, zaman serisi analizi, görüntü işleme ve doğal dil işleme gibi çeşitli alanlarda projeler geliştirilirken bu verilere ihtiyaç duyulmaktadır. Bu projede, size atanan veri tipiyle ilgili büyük ölçekli bir veri seti oluşturmanız ve bu veriyi ilgili veri ön işleme adımlarına tabi tutmanız gerekmektedir.

İSTERLER

- Toplanan veriler ile sınıflandırma işlemi yapılacaksa, sınıf başına minimum 5000 kayıt toplamanız gerekmektedir.
- Toplanan veriler için yapılması gereken asgari veri işleme adımları listelenmiştir. Bu ön işleme adımlarını artırmalısınız:
 - **Ses verisi:** Gürültü temizleme, ses normalizasyonu, Mel frekans cepstrum katsayıları (MFCC) çıkarma
 - **Zaman serisi:** Eksik verilerin doldurulması, Zaman serilerinin yeniden örneklenmesi, Trend ve mevsimsellik bileşenlerinin ayrıştırılması.
 - **Görüntü verisi:** Görüntü yeniden boyutlandırma, parlaklık ve kontrast ayarları, veri artırma (augmentation) işlemleri.
 - **Metin verisi:** Tokenizasyon, stopwords temizleme, lemmatizasyon ve kök bulma.

NOTLAR

- **Platform:** Veri toplama için örnek platformlar aşağıda verilmiştir:
 - **Ses verisi:** Librivox, YouTube
 - **Zaman serisi:** Kaggle, Yahoo Finance API
 - **Görüntü verisi:** Reddit, Flickr, Shutterstock
 - **Metin verisi:** Vikipedi, Reddit, Sözlük siteleri (Türkçe için).
- Öğrenciler grupları ikiyeşerli olarak kendi aralarında oluşturacaktır.
 - Her öğrenci, yalnızca aynı konuya sahip bir diğer öğrenci ile grup olabilir.
 - Son tarihten sonra grup oluşturmayan öğrencilerin grupları rastgele oluşturulacaktır.
 - **Grup oluşturmak için son tarih: 4 Ekim 2024 Cuma**
 - Grup oluşturma bağlantısı [için tıklayınız.](#)
- Geliştirdiğiniz projeyi **son teslim tarihinden önce** GitHub platformuna da yüklemelisiniz.
 - GitHub reponuzda **README** (beni oku) dokümanını oluşturmalsınız.
 - README dokümanı; proje özeti, geliştirme ortamı, projenin yüklenmesi ve çalışır hale getirilmesi, geliştirilen arayüzün örnek görseli gibi bilgileri içeren bir dokümandır. Bu doküman maddeler halinde, kısa ve net cümlelerden oluşur.
 - Topladığınız verileri de proje son teslim tarihinde teslim etmelisiniz. Bu verilerin boyutu büyük olabileceğinden Google Drive, Dropbox gibi platformlara yükleyerek bu platformlardan link olarak paylaşmanız gerekmektedir. Bu paylaşacağınız bağlantı, GitHub reponuzda da olmalıdır.
- Proje kapsamında hazırladığınız rapor, GitHub reposuna yüklenmelidir. **Raporun çıktısını almayınız.**
- Rapor, IEEE konferans şablonuna ([tıklayınız](#)) uygun hazırlanmalıdır. Bu şablona uymayan raporlar değerlendirmeye alınmaz.
- Proje kapsamında kullanılacak teknoloji ve yazılım dili kısıtlaması yoktur.
- Toplanan verilerin, bir sonraki Yazılım Geliştirme Laboratuvarı projesinde kullanılması zorunludur.
- **Toplantı:** Proje ile ilgili bilgilendirme toplantısı 30 Eylül 2024 Pazartesi Saat 12.30'da Microsoft Teams üzerinden yapılacaktır. Toplantı bağlantısı [için tıklayınız.](#)
 - **Toplanacak veri, Yaz. Gel. Lab. I – II. projenin konusuna uygun olarak seçilmelidir. Bu sebeple toplantıya katılımınız önemlidir.**

PROJE TESLİMİ ve SUNUMU

- Proje sunumu sırasında algoritmanız, geliştirdiğiniz kodun çeşitli kısımlarının ne amaçla yazıldığı ve geliştirme ortamı hakkında sorular sorulabilir. Kullandığınız herhangi bir satır kodu açıklamamız istenebilir.
 - Kodun her satırından tüm grup elemanları sorumludur.
 - Sunum esnasında projeye ilgili tüm grup üyelerine veri toplama ve ön işleme adımları ile ilgili sorular yöneltilecektir.
- **Proje son teslim tarihi: 15 Kasım 2024 Cuma Saat 17.00.**
- **Sunum tarihi: 18-22 Kasım 2024.** Sunum listeleri son teslim tarihinden sonra duyurulacaktır.
- Sunum süresi: 20 dk.
- **İNTİHAL:** İnternette alman kod parçacıkları mutlaka kod içerisinde belirtilecek ve açıklama satırı ile kaynak gösterilecektir. Aksi durumda kopya olarak değerlendirilecektir. Kopya çektiği ya da kopya verdiği tespit edilen öğrenciler sunuma alınmayacaktır.