Fabrikalar ve Motorlu Taşıtlardan Kaynaklanan Hava Kirliliği Görüntü Verisi Toplama ve Ön İşleme

Adı Soyadı: Mehmet Oğuz Ergin   
Kurum: *Kocaeli Üniversitesi*Adres: İzmit/Kocaeli  
Mail: oguzergin377@gmail.com

# Giriş

Hava kirliliği, insan sağlığı ve çevre üzerinde önemli etkilere sahiptir. Bu çalışmada, fabrikalar ve motorlu taşıtlardan kaynaklanan kirlilik olmak üzere iki ana kaynağa odaklanılmaktadır. Toplanan görüntü verisi, bu kirlilik kaynaklarını tanımlayabilen ve sınıflandırabilen yapay zekâ modellerini eğitmek için kullanılacaktır. Veri ön işleme aşamaları, makine öğrenmesi için uygun kalite ve formatta görüntülerin hazırlanmasını sağlar.

# Veri Toplama Yöntemi

## Araçlar ve Teknolojiler

* **Programlama Dili**: Python
* **Kullanılan Kütüphaneler**: Selenium, PIL (Python Imaging Library), urllib
* **Web Kazıma Platformu**: Google Görseller

## Uygulama

Hava kirliliği görüntülerini toplama sürecini otomatikleştirmek için bir Python kodu geliştirilmiştir. Kod, aşağıdaki adımları gerçekleştirir:

1. **WebDriver Başlatma**: Google Görseller ile etkileşim için ChromeDriver kullanılmıştır.
2. **Görsel Arama**: İlgili arama terimleri girilmiş ve daha fazla görsel yüklemek için sayfa kaydırılmıştır.
3. **Görsel İndirme ve Yeniden Boyutlandırma**: Her görsel yerel olarak kaydedilmiş, 1024×768 çözünürlüğünde yeniden boyutlandırılmış ve işlenmeye uygun hale getirilmiştir.

# Ön İşleme Adımları

1. **Görsel Yeniden Boyutlandırma**:  
   Tüm görseller PIL kütüphanesi kullanılarak 1024×768 piksel standart çözünürlüğüne getirilmiştir.
2. **Kalite Artırımı**:  
   Yüksek kaliteli LANCZOS filtresi uygulanarak net ve tutarlı görüntü çıkışı sağlanmıştır.
3. **Dosya Organizasyonu**:  
   Görseller, kolay kategorize edilmesi için yapılandırılmış dosya adlarıyla kaydedilmiştir (örneğin, air\_pollution\_1.jpg).

# Sonuçlar

* **Toplam Toplanan Görsel Sayısı**: 6000
* **Kategoriler**: Fabrikalardan kaynaklanan hava kirliliği, Motorlu taşıtlardan kaynaklanan hava kirliliği.
* **Çözünürlük**: Tüm görseller 1024×768 piksel çözünürlüğünde yeniden boyutlandırılmıştır.  
    
  Veri seti, her bir kirlilik kaynağına karşılık gelen görselleri içeren dizinler şeklinde düzenlenmiştir. Ön işleme adımları, görsellerin makine öğrenmesi modellerini eğitmeye hazır hale gelmesini sağlamıştır.

# Tartışma

Otomatik veri toplama yöntemi, önemli ölçüde zaman ve çaba tasarrufu sağlamıştır. Dinamik olarak yüklenen içeriklerle uğraşmak ve yüksek çözünürlüklü görsellerin indirilmesini sağlamak, karşılaşılan başlıca zorluklardandır. Bu sorunlar Selenium’un bekleme işlevi ve başarısız indirmeler için tekrar deneme mekanizmaları ile aşılmıştır.

# Sonuç

Bu proje, hava kirliliği görsellerinden oluşan bir veri setini başarıyla toplamış ve ön işlemlerini gerçekleştirmiştir. Veri seti, kirlilik kaynaklarını sınıflandırabilen yapay zekâ modellerinin eğitiminde önemli bir rol oynayacaktır ve çevre izleme ve yönetimine katkıda bulunacaktır.

# Kaynaklar

* Selenium Dokümantasyonu. [Çevrimiçi]. Erişim: <https://selenium.dev/documentation/>
* Python Imaging Library (PIL). [Çevrimiçi]. Erişim: <https://pillow.readthedocs.io/>

 **Veri Seti Bağlantısı**: [Google Drive Linki](https://drive.google.com/drive/folders/1lq3v8Lyv2_YxNMw9sAWGxDME18I1epj3?usp=sharing).

 **Kod Deposu**: [GitHub Linki](https://github.com/moguzergin/web_scraping_project).