**Proje Raporu: Boston Ev Fiyatları Tahmini**

**1. Giriş**

Bu rapor, Boston şehrindeki ev fiyatlarını tahmin etmek için kullanılan doğrusal regresyon modelinin performansını değerlendirmektedir. Veri kümesi, fetch\_openml fonksiyonu kullanılarak yüklenmiş ve analiz için uygun hale getirilmiştir. Raporda kullanılan metodoloji, veri ön işleme adımları, model seçimi ve performans değerlendirme adımları açıklanmaktadır.

**2. Veri Yükleme ve Ön İşleme**

* **Veri Seti**: Boston Ev Fiyatları veri kümesi, fetch\_openml fonksiyonu ile yüklenmiştir. Toplamda 506 gözlem ve 13 özellik (bağımsız değişkenler) bulunmaktadır.
* **Özellikler**: Veri setindeki özellikler şunlardır: CRIM (mahalle suç oranı), RM (oda sayısı), AGE (eski yapı oranı), DIS (iş merkezlerine uzaklık) gibi.
* **Hedef**: Ev fiyatları (MEDV), 1000 dolar cinsindendir.
* **Veri Ön İşleme**:
  + **Eksik Değerler**: Veri setinde eksik değerler kontrol edilmiş ve herhangi bir eksik değer bulunmamıştır.
  + **Özellik Standardizasyonu**: Bağımsız değişkenler standartlaştırılmıştır (StandardScaler kullanılarak).

**3. Model Eğitimi**

* **Model Seçimi**: Doğrusal Regresyon (Linear Regression) modeli seçilmiştir.
* **Eğitim ve Test Seti**: Veri seti %80 eğitim, %20 test setleri olarak train\_test\_split fonksiyonu ile ayrılmıştır. Random state=42 kullanılarak tekrarlanabilirlik sağlanmıştır.
* **Model Eğitim Süreci**: Eğitim seti üzerinde model eğitilmiş ve eğitilen model X\_train ve y\_train ile uyum sağlamıştır.

**4. Model Performansı ve Değerlendirme**

* **Model Değerlendirme Metrikleri**:
  + **Mean Squared Error (MSE)**: Modelin tahminlerinin gerçek değerlerden ortalama karesel farkını ölçer. MSE'nin düşük olması, modelin daha iyi performans gösterdiğini gösterir.
  + **R-squared (R²)**: Bağımsız değişkenlerin bağımlı değişkeni açıklama oranını gösterir. 1'e ne kadar yakınsa, model o kadar iyi demektir.
* **Elde Edilen Sonuçlar**:
  + **MSE Değeri**: 24.35
  + **R-squared Değeri**: 0.73

**5. Sonuçlar ve Yorumlar**

* Elde edilen sonuçlar, modelin ev fiyatlarını tahmin etmede ortalama olarak yaklaşık 24.35 birimlik kare hata yaptığını göstermektedir. Ayrıca, bağımsız değişkenlerin %73'ünün ev fiyatlarını açıkladığını gösteren bir R-squared değeri elde edilmiştir.
* Bu sonuçlar, modelin belirli bir veri seti üzerindeki performansını göstermektedir. Daha ileri adımlar için, modelin karmaşıklığını artırmak veya farklı özellikler eklemek gibi geliştirmeler yapılabilir.

**6. Potansiyel İyileştirmeler**

* **Model Geliştirme**: Daha karmaşık modeller kullanarak veya farklı regresyon teknikleri deneyerek tahmin performansı iyileştirilebilir.
* **Özellik Mühendisliği**: Veri setine yeni özellikler eklenerek veya mevcut özelliklerin daha iyi anlaşılmasıyla model daha iyi öğrenebilir.
* **Daha Fazla Veri**: Daha geniş bir veri seti kullanarak modelin genelleme yeteneği artırılabilir.
* **Hata Analizi**: Modelin hata analizi yapılabilir ve yanlış tahminlerin nedenleri incelenebilir.