

Teknik Rapor – Kira Tahminleme Sistemi

Bu çalışmanın amacı, **Hepsiemlak** platformundan alınan bir veri seti kullanılarak Türkiye’deki konut kira fiyatlarını tahmin edebilen bir regresyon modeli geliştirmektir. Gerçek dünya verileri üzerinde yapılan bu analiz, veri ön işleme, modelleme ve değerlendirme adımlarını içermektedir.

Kullanılan Veri Seti

Kaynak: Hepsiemlak (CSV dosyası)

Yöntem: Google Colab üzerinden Google Drive bağlantısı ile yüklenmiştir.

Kullanılan Dil: Python

Veri setimde bulunan özellikler şunlardır:

- Şehir
- İlçe
- Mahalle
- Oda sayısı
- Metrekare
- Bina yaşı
- Bulunduğu kat
- Isıtma türü
- Fiyat — hedef değişken

Veri ön işleme ve özellik mühendisliği kapsamında, kategorik değişkenler LabelEncoder kullanılarak dönüştürülmüş, böylece makine öğrenmesi algoritmaları için uygun hale getirilmiştir. Ayrıca, analiz sürecini daha verimli hale getirmek amacıyla, gerekli olmayan veya fazla aykırı olan bazı satırlar ve sütunlar veri setinden çıkarılmıştır. Sayısal veriler, ölçekleme işlemiyle normalleştirilerek farklı özellikler arasındaki değerlerin dengelenmesi sağlanmıştır. Son olarak, veri train_test_split yöntemiyle eğitim ve test olarak bölünerek modelin doğruluğunu değerlendirmek için uygun bir yapı oluşturulmuştur.

Yapılan analizler sonucunda, metrekare ile fiyat arasında pozitif bir korelasyon olduğu belirlenmiştir, yani alan büyüdükçe fiyatın da arttığı gözlemlenmiştir. Benzer şekilde, oda sayısındaki artışın kira fiyatı üzerinde yükseltici bir etkisi olduğu tespit edilmiştir. Ayrıca, kategorik değişkenlerin dağılımları detaylı olarak incelenmiş; özellikle ısıtma türü ve bulunduğu kat gibi faktörlerin nasıl dağıldığına dair önemli bilgiler elde edilmiştir. Son olarak, analizlerin görsel olarak desteklenmesi amacıyla matplotlib ve seaborn kütüphaneleri kullanılarak çeşitli grafikler oluşturulmuştur, böylece verilerin daha anlaşılır bir biçimde yorumlanması sağlanmıştır.

Modelleme ve Performans Analizi kapsamında, iki farklı regresyon modeli kullanılarak verilerin tahmin gücü incelenmiştir. İlk olarak, temel regresyon modellerinden Lineer Regresyon tercih edilmiştir. Bu model, kolay uygulanabilir ve yorumlanabilir olmasıyla öne çıkmaktadır. Performans değerlendirmesi sonucunda, RMSE (Root Mean Square Error) değeri 2900.15, MAPE (Mean Absolute Percentage Error) ise %13.7 olarak hesaplanmıştır.

Daha gelişmiş ve güçlü bir alternatif olarak XGBoost Regressor kullanılmıştır. Karar ağacı tabanlı bu model, hiperparametre ayarları ile optimize edilerek daha başarılı sonuçlar elde edilmiştir. Performans ölçümlerine bakıldığında, RMSE değeri 2154.42, MAPE ise %9.1 olarak bulunmuş, böylece önceki modele kıyasla daha düşük hata oranlarıyla daha yüksek tahmin doğruluğuna ulaşılmıştır.

Bu sonuçlar, veri setinin ve problem yapısının daha gelişmiş bir modelle işlendiğinde daha iyi performans gösterdiğini ortaya koymaktadır.

Performans Karşılaştırması sonucunda, kullanılan modellerin hata oranları ve doğruluk seviyeleri değerlendirilmiştir. Linear Regression modeli, RMSE değeri 2900.15 ve MAPE oranı %13.7 ile temel bir performans göstermiştir. Buna karşılık, daha gelişmiş bir karar ağacı tabanlı model olan XGBoost Regressor, RMSE değerini 2154.42 ve MAPE oranını %9.1 seviyelerine düşürerek belirgin bir üstünlük sağlamıştır.

Bu sonuçlar, XGBoost modelinin doğrusal olmayan ilişkileri daha iyi yakalayarak Lineer Regresyon'a kıyasla daha başarılı olduğunu ortaya koymaktadır. Lineer Regresyon'un basit ve yorumlanabilir yapısına rağmen, karmaşık veri setlerinde yeterli esnekliği sağlayamayabileceği anlaşılmıştır. Özellikle değişkenler arasındaki karmaşık ilişkiler söz konusu olduğunda, XGBoost'un hiperparametre optimizasyonlarıyla çok daha güçlü tahminler sunduğu gözlemlenmiştir.

Bu analiz, veri setindeki yapıya uygun model seçiminin kritik önem taşıdığını gösteriyor.

Sonuç olarak, gerçek dünya verileri kullanılarak kira tahminleme modeli başarılı bir şekilde oluşturulmuştur. Veri ön işleme adımları, model performansında belirleyici bir rol oynayarak, hataların minimize edilmesine ve daha doğru tahminlerin yapılmasına olanak sağlamıştır. Yapılan performans değerlendirmeleri sonucunda, XGBoost modeli, hem RMSE hem de MAPE açısından daha düşük hata oranı sunarak, en iyi performansı sergileyen algoritma olarak öne çıkmıştır.

Bu başarılı modelleme süreci, emlak platformları için fiyat öneri sistemlerinin temelini oluşturabilecek sağlam bir altyapı sağlamaktadır. Böylece, doğru ve güvenilir tahminler ile kullanıcıların karar alma süreçleri daha verimli hale getirilebilir.