



**1. Model Performansı ve RMSE Analizi**

Kod, dört farklı makine öğrenimi modelini kullanarak otel fiyatlarını tahmin eder ve modellerin hata oranlarını değerlendirir. Root Mean Squared Error (RMSE) değerleri aşağıda verilmiştir:

**Test Seti RMSE Sonuçları (Gerçekçi Model Başarısı)**

* **SVR:** 11,767.01
* **Ridge:** 10,788.14
* **Lasso:** 10,796.77
* **RandomForest:** **10,564.98 (En iyi test seti başarımı)**

**Eğitim Seti RMSE Sonuçları (Aşırı Öğrenme Durumu Analizi)**

* **SVR:** 11,587.15
* **Ridge:** 10,141.81
* **Lasso:** 10,141.77
* **RandomForest:** **4,666.95 (Önemli derecede düşük, potansiyel aşırı öğrenme belirtisi)**

**Analiz:**

* RandomForest modeli eğitim setinde **4,666.95** gibi oldukça düşük bir RMSE elde ederken, test setinde **10,564.98** RMSE'ye çıkmıştır. Bu durum, modelin eğitildiği veriye aşırı derecede uyum sağladığını ve test setinde biraz daha fazla hata yaptığını gösterir. **Overfitting (aşırı öğrenme) belirtileri mevcut olabilir.**
* Ridge ve Lasso modelleri, düzenleme (regularization) sayesinde hem eğitim hem de test setinde daha dengeli sonuçlar vermektedir.
* SVR modeli, Ridge ve Lasso'ya kıyasla en yüksek hatayı üretmiştir, bu da doğrusal modellemenin veriye tam olarak uygun olmayabileceğini gösterir.

**Model Training**

* **SVR (Support Vector Regression – Linear Kernel):**  
  Bu model, verideki doğrusal ilişkileri modellemek amacıyla kullanılmaktadır. SVR, daha az parametrik yapısı nedeniyle klasik lineer regresyona benzer bir yaklaşım sergilemekte, aynı zamanda veri üzerinde optimal sınırları belirleyerek hata toleransını minimize etmeye çalışmaktadır.
  + Avantaj: Sınırlı sayıda veriyi destek vektörleri üzerinden genelleme yaparak modellemeye çalışır.
  + Dezavantaj: Büyük veri setlerinde yavaş çalışabilir ve karmaşık ilişkileri yeterince yakalayamayabilir.
* **Ridge Regression:**  
  Ridge modeli, L2 norm düzenlemesi (regularization) uygulayarak model katsayılarının büyüklüğünü sınırlar. Bu yaklaşım, modelin aşırı uyum (overfitting) yapmasını engelleyerek genelleme performansını artırmaya yönelik önemli bir strateji sunar.
  + Avantaj: Model katsayılarını kontrol altında tutarak overfitting'i önler.
  + Dezavantaj: Doğrusal modelleme yapması nedeniyle bazı karmaşık ilişkileri yakalayamayabilir.
* **Lasso Regression:**  
  Lasso, L1 düzenlemesi kullanarak model katsayılarının bazılarını sıfıra indirgeme özelliğine sahiptir. Bu özellik, yalnızca önemli değişkenlerin modele dahil edilmesini sağlayarak otomatik özellik seçimi (feature selection) yapmaya imkan tanır.
  + Avantaj: Ridge ile benzer olup, aynı zamanda önemsiz değişkenleri sıfıra indirerek otomatik özellik seçimi yapar.
  + Dezavantaj: Küçük veri setlerinde bazı değişkenleri tamamen sıfıra indirebilir, bu da bilgi kaybına yol açabilir.
* **RandomForestRegressor:**  
  Bu model, karar ağaçlarının ensemble yöntemine dayanmaktadır. RandomForest, veri içindeki non-lineer ilişkileri yakalamada oldukça etkilidir; model, çeşitli ağaçların oluşturduğu kolektif sonuçlardan yararlanarak yüksek doğrulukta tahminler üretir.
  + Avantaj: Doğrusal olmayan ilişkileri yakalamakta oldukça başarılıdır. Özellikle outlier’lara karşı daha dayanıklıdır.
  + Dezavantaj: Ağaç tabanlı modellerin fazla büyümesi, aşırı öğrenmeye sebep olabilir (Overfitting riski).

**Rationale for Model Selection**

* **Diversity (Çeşitlilik):**  
  Kullanılan modeller arasında hem doğrusal (SVR, Ridge, Lasso) hem de non-lineer (RandomForest) yaklaşımların bulunması, veri içerisindeki farklı ilişkilerin ve yapısal özelliklerin daha kapsamlı şekilde yakalanmasına olanak sağlar.
* **Regularization and Feature Selection (Düzenleme ve Özellik Seçimi):**  
  Ridge ve Lasso modelleri, düzenleme (regularization) teknikleri sayesinde model karmaşıklığını kontrol altına alır. Bu yöntemler, aşırı uyumu önleyerek daha stabil ve genelleyici modeller elde edilmesine yardımcı olurken, Lasso'nun sıfır katsayıya indirme özelliği, veri içerisindeki önemli değişkenlerin belirlenmesinde kritik rol oynar.
* **Ensemble Approach (Ensemble Yaklaşımı):**  
  Farklı modellerden elde edilen tahminlerin ortalaması alınarak oluşturulan ensemble yöntemi, tek bir modelin hatalarından kaynaklanabilecek yanlılıkları minimize eder. Bu, genel tahmin performansını artırarak daha güvenilir ve sağlam sonuçlar elde edilmesini sağlar.

Bu yöntemlerin kombinasyonu, hem model genellemesi hem de özelliklerin önem derecelerinin daha objektif bir biçimde değerlendirilmesi açısından kapsamlı bir analiz sağlamaktadır.

**Ensemble Yaklaşımı ile Daha Güçlü Tahminler**

* Dört modelin tahminleri alınarak **ortalama tahmin değeri** hesaplanmıştır.
* Bu ensemble yöntemi, **her modelin güçlü yanlarını birleştirerek** hata oranını düşürmeyi amaçlamaktadır.
* Otel fiyatlarının gerçeğe en yakın tahmini bu yöntemle elde edilmiştir.