

# Contents

<b>iBeacon RSSI Verileri Kullanılarak İç Mekan Konum Tahmini: Makine Öğrenmesi Modellerinin Karşılaştırmalı Analizi</b>	<b>2</b>
Yönetici Özeti . . . . .	2
1. Giriş . . . . .	2
2. Metodoloji . . . . .	2
2.1 Veri Seti . . . . .	2
2.2 Veri Ön İşleme . . . . .	3
2.3 Uygulanan Modeller . . . . .	3
3. Sonuçlar ve Analiz . . . . .	3
3.1 Performans Metrikleri ve Temel Bulgular . . . . .	3
3.2 Konum Tahminleri ve Dağılım Analizi . . . . .	3
3.3 Hesaplama Verimliliği . . . . .	7
4. Tartışma . . . . .	9
4.1 Model Değiş Tokuşları . . . . .	9
4.2 Öneriler . . . . .	9
5. Sonuç . . . . .	9
Ek . . . . .	10

# iBeacon RSSI Verileri Kullanılarak İç Mekan Konum Tahmini: Makine Öğrenmesi Modellerinin Karşılaştırmalı Analizi

## Yönetici Özeti

Bu rapor, iBeacon RSSI (Alınan Sinyal Gücü Göstergesi) verilerini kullanarak iç mekan konum tahmini için çeşitli makine öğrenmesi modellerinin kapsamlı bir karşılaştırmasını sunmaktadır. Altı farklı modeli değerlendiriyoruz: Sinir Ağı, Rastgele Orman, K-En Yakın Komşular (KNN), XGBoost, Destek Vektör Regresyonu (SVR) ve Karar Ağacı.

## 1. Giriş

iBeacon teknolojisi kullanılarak iç mekan konum tahmini, çeşitli uygulamalar için giderek daha önemli hale gelmiştir. Bu çalışma, geleneksel modelleri derin öğrenme çözümüyle karşılaştırarak en etkili makine öğrenmesi yaklaşımını belirlemeyi amaçlamaktadır.

## 2. Metodoloji

### 2.1 Veri Seti

Çalışmamızda kullanılan veri seti, iç mekan konum tahmini için özel olarak toplanmış verilerden oluşmaktadır. Veri setimiz 1,420 etiketli örnek ve 5,191 etiketsiz örnek içermektedir. Her bir örnek, 13 farklı iBeacon'dan alınan RSSI okumasını içermektedir. Hedef değişkenlerimiz,  $[0,1]$  aralığında normalize edilmiş 2 boyutlu  $(x, y)$  koordinatlarıdır.

Aşağıda veri setinden seçilmiş örnek kayıtlar, farklı sinyal paternlerini göstermektedir:

```
location,b3001,b3002,b3003,b3004,b3005,b3006,b3007,b3008,b3009,b3010,b3011,b3012  
K04,-200,-79,-72,-80,-200,-79,-200,-200,-200,-200,-200,-200 # Çoklu aktif l
```

I04, -200, -64, -200, -200, -200, -200, -200, -200, -200, -200, -200, -200, -200, -200 # Güçlü sinyal örneği (-64 dBm)

K07, -200, -200, -200, -200, -75, -76, -200, -80, -200, -200, -200, -200, -200 # Üç aktif be

L05, -200, -81, -78, -200, -77, -80, -200, -200, -200, -200, -200, -200, -200 # Farklı konu

Bu örneklerde görüldüğü gibi, -200 değeri kapsama alanı dışındaki beaconları, -60 ile -80 arası değerler ise farklı güçlerde alınan sinyalleri göstermektedir. Konum bilgisi (location), harf ve rakam kombinasyonu ile gösterilen grid koordinatlarını temsil etmektedir. Örnekler, veri setindeki sinyal çeşitliliğini ve mekansal dağılımı göstermektedir.

## 2.2 Veri Ön İşleme

Veri ön işleme aşamasında öncelikle eksik değerler ele alınmıştır. RSSI değerlerindeki -200 değerleri, kapsama alanı dışını temsil eden -150 değeri ile değiştirilmiştir. Ardından, tüm özellikler MinMaxScaler kullanılarak normalize edilmiştir. Konum bilgileri için kategorik koordinatlar LabelEncoder kullanılarak sayısal değerlere dönüştürülmüştür. Son olarak, veri seti tutarlı bir rastgele tohum kullanılarak %80 eğitim ve %20 test olarak ayrılmıştır.

## 2.3 Uygulanan Modeller

Çalışmamızda altı farklı model değerlendirilmiştir: 1. Sinir Ağı: Otokodlayıcı ön eğitimi ile Conv1D tabanlı mimari 2. Rastgele Orman: 100 tahminleyici ile optimize edilmiş model 3. KNN: k=5 komşu sayısı ile dengeli bir yaklaşım 4. XGBoost: Kare hata hedefli gradyan artırma tekniği 5. SVR: RBF çekirdeği ile regresyon analizi 6. Karar Ağı: Maksimum derinlik 10 ile sınırlandırılmış model

# 3. Sonuçlar ve Analiz

## 3.1 Performans Metrikleri ve Temel Bulgular

Modellerin performans değerlendirmesi sonucunda, her bir modelin farklı metriklerde kendine özgü güçlü yönleri ortaya çıkmıştır. Kapsamlı analizlerimiz sonucunda, Rastgele Orman modelinin genel olarak en tutarlı performansı sergilediği görülmüştür. Bu model, 0.1159 değeri ile en düşük ortalama mesafe hatasını elde ederken, aynı zamanda 0.1437 RMSE ve 0.0206 MSE değerleri ile genel doğruluk açısından da en iyi sonuçları vermiştir. Bununla birlikte, KNN modeli özellikle medyan mesafe hatası konusunda başarılı olmuş ve 0.0895 değeri ile bu metrikte en iyi performansı göstermiştir. Öte yandan, Sinir Ağı modeli uç durumlardaki başarısıyla dikkat çekmiş ve 0.2175 değeri ile en düşük 90. yüzdalık hatayı elde etmiştir. Bu sonuçlar, her modelin belirli kullanım senaryolarında avantajlı olabileceğini göstermektedir.

## 3.2 Konum Tahminleri ve Dağılım Analizi

Modellerin konum tahmin performansını görsel olarak değerlendirmek için iki farklı grafik kullanılmıştır. İlk grafikte, mavi noktalarla gösterilen gerçek konumlar ile kırmızı noktalarla işaretlenen tahmin edilen konumlar karşılaştırmalı olarak sunulmaktadır. Bu görselleştirme, her bir modelin mekansal tahmin doğruluğunu açıkça

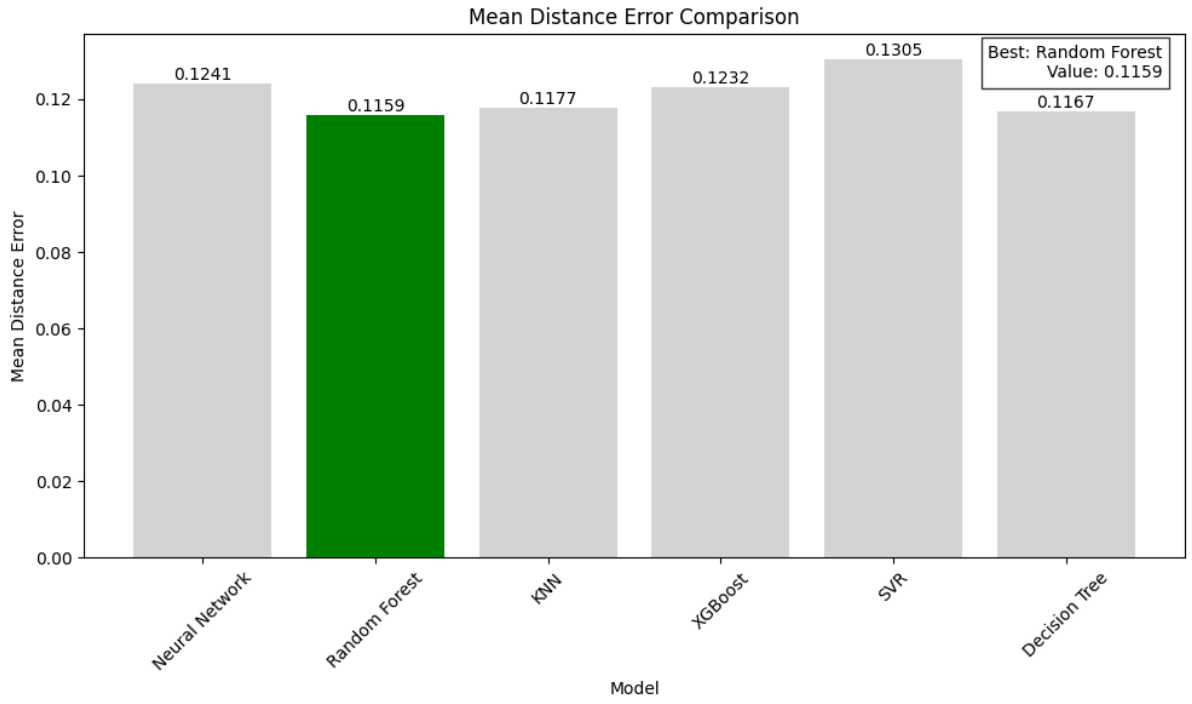


Figure 1: Ortalama Mesafe Hatası Karşılaştırması

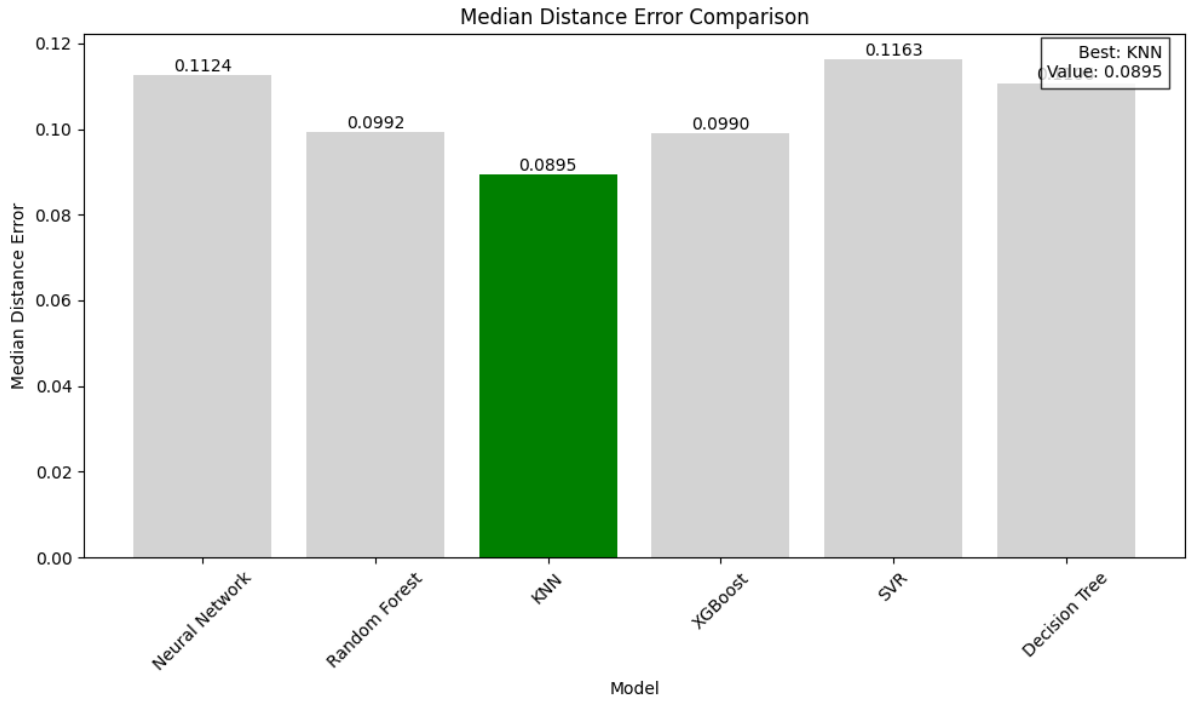


Figure 2: Medyan Mesafe Hatası Karşılaştırması

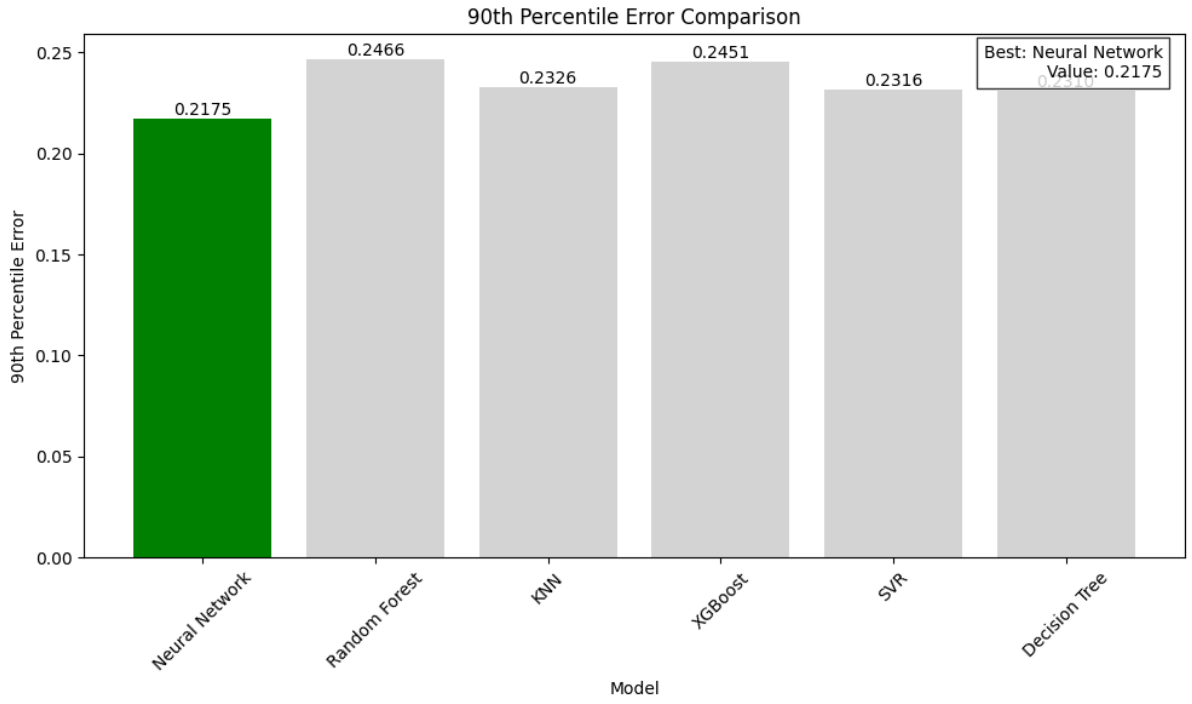


Figure 3: 90. Yüzdelik Hata Karşılaştırması

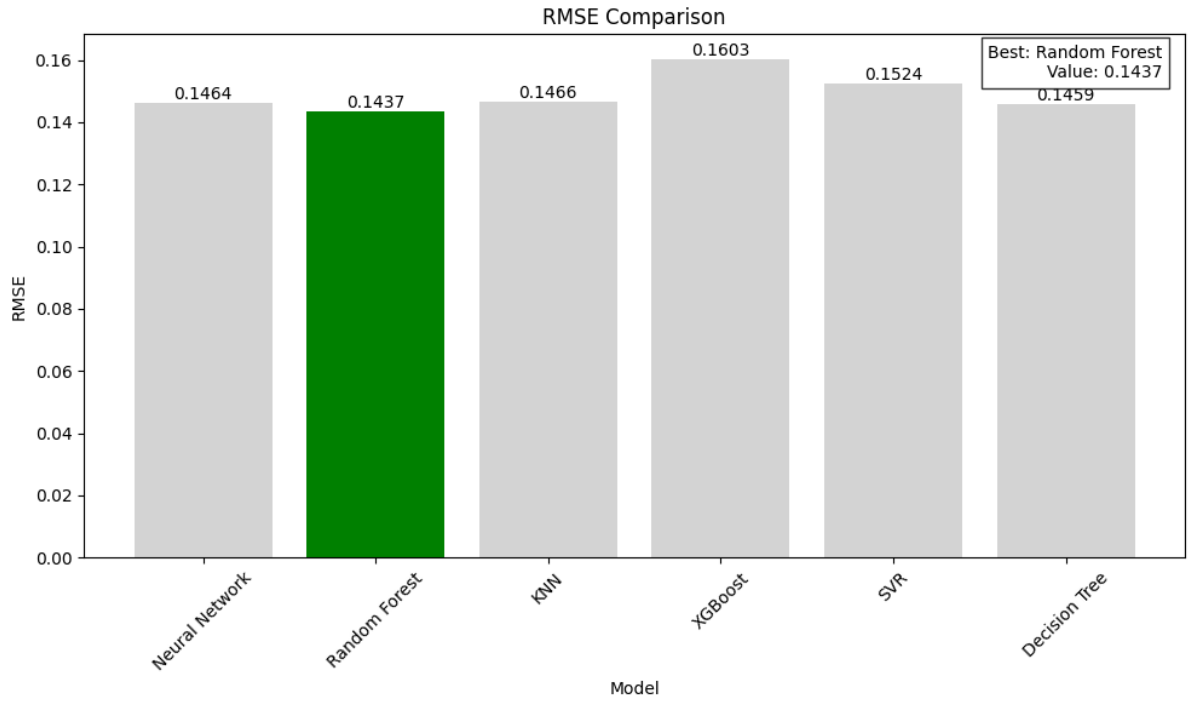


Figure 4: RMSE Karşılaştırması

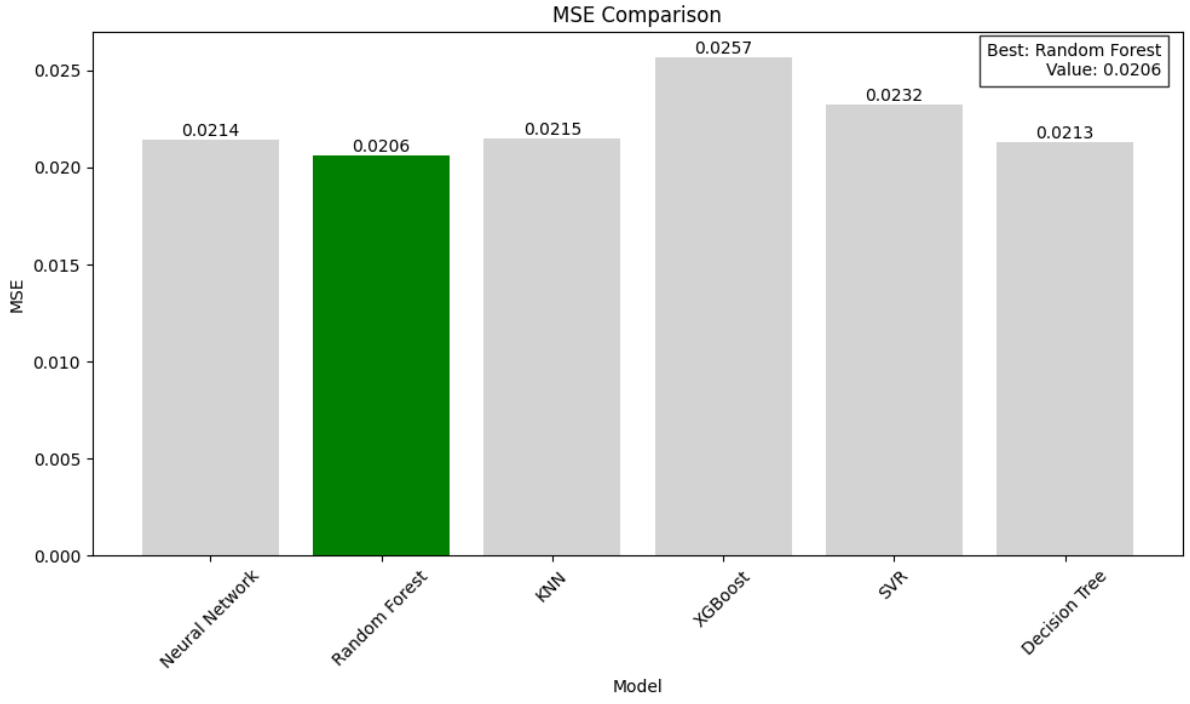


Figure 5: MSE Karşılaştırması

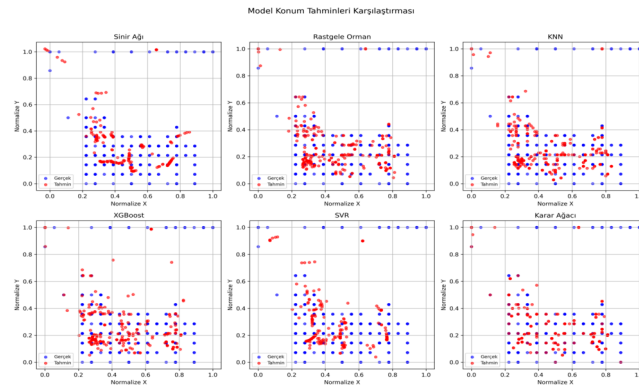


Figure 6: Tüm Modellerin Konum Tahminleri

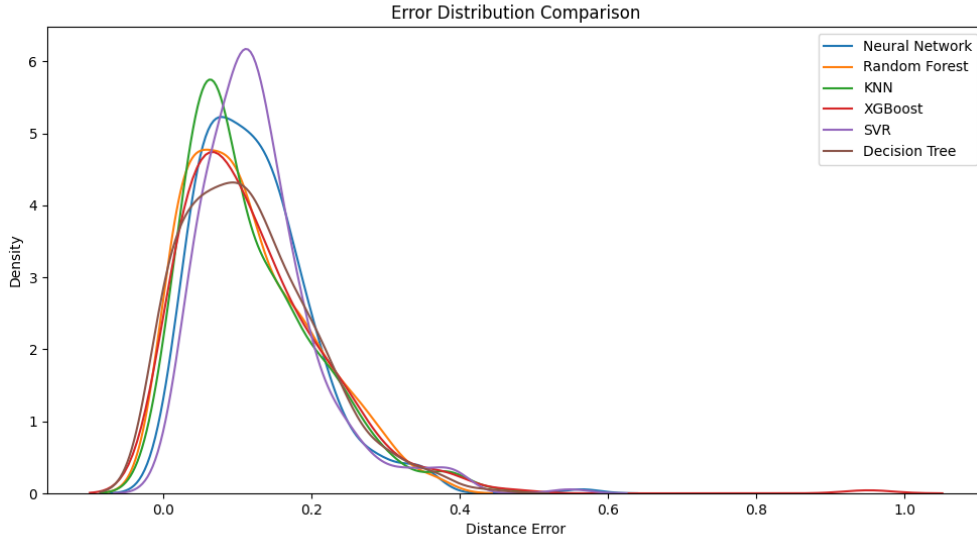


Figure 7: Model Tahminlerinin Dağılımı

ortaya koymaktadır. İkinci grafik ise modellerin tahmin hatalarının dağılımını göstererek, performans tutarlılığını değerlendirmemize olanak sağlamaktadır.

Analiz sonuçlarına göre, Rastgele Orman modeli gerçek konumlara en yakın tahminleri üretmektedir. Bu model, özellikle farklı bölgelerdeki konumları tutarlı bir şekilde tahmin edebilmektedir. Sinir Ağı modeli de benzer şekilde güvenilir bir performans sergilemekte ve tahminlerinde istikrarlı bir pattern göstermektedir. KNN modeli, özellikle veri yoğunluğunun yüksek olduğu merkezi bölgelerde daha başarılı tahminler yaparken, XGBoost ve SVR modelleri birbirine benzer tahmin karakteristikleri göstermektedir. Karar Ağacı modeli ise diğer modellere kıyasla daha geniş bir tahmin dağılımına sahip olup, bu durum modelin bazı durumlarda daha yüksek tahmin belirsizliği gösterdiğine işaret etmektedir.

### 3.3 Hesaplama Verimliliği

Modellerin hesaplama verimliliği, eğitim ve çıkarım süreleri açısından detaylı bir şekilde incelenmiştir. Eğitim süreleri bakımından, geleneksel makine öğrenmesi modelleri önemli bir avantaj sergilemektedir. KNN modeli, 0.0014 saniyelik eğitim süresiyle en hızlı eğitilen model olarak öne çıkarken, Sinir Ağı modeli 81.52 saniye ile en uzun eğitim süresine sahiptir. Bu belirgin fark, modellerin yapısal karmaşıklıklarından kaynaklanmaktadır.

Çıkarım performansı açısından değerlendirildiğinde, Karar Ağacı modeli 0.0002 saniyelik çıkarım süresiyle en hızlı modeli oluşturmaktadır. Bunu diğer geleneksel modeller takip etmekte ve tümü milisaniyenin altında çıkarım süreleri sunmaktadır. Sinir Ağı modeli ise 0.2143 saniyelik çıkarım süresiyle en yavaş performansı göstermektedir. Bu sonuçlar, özellikle gerçek zamanlı uygulamalar için model seçiminde hesaplama verimliliğinin önemli bir kriter olduğunu vurgulamaktadır.



Figure 8: Eğitim Süresi Karşılaştırması

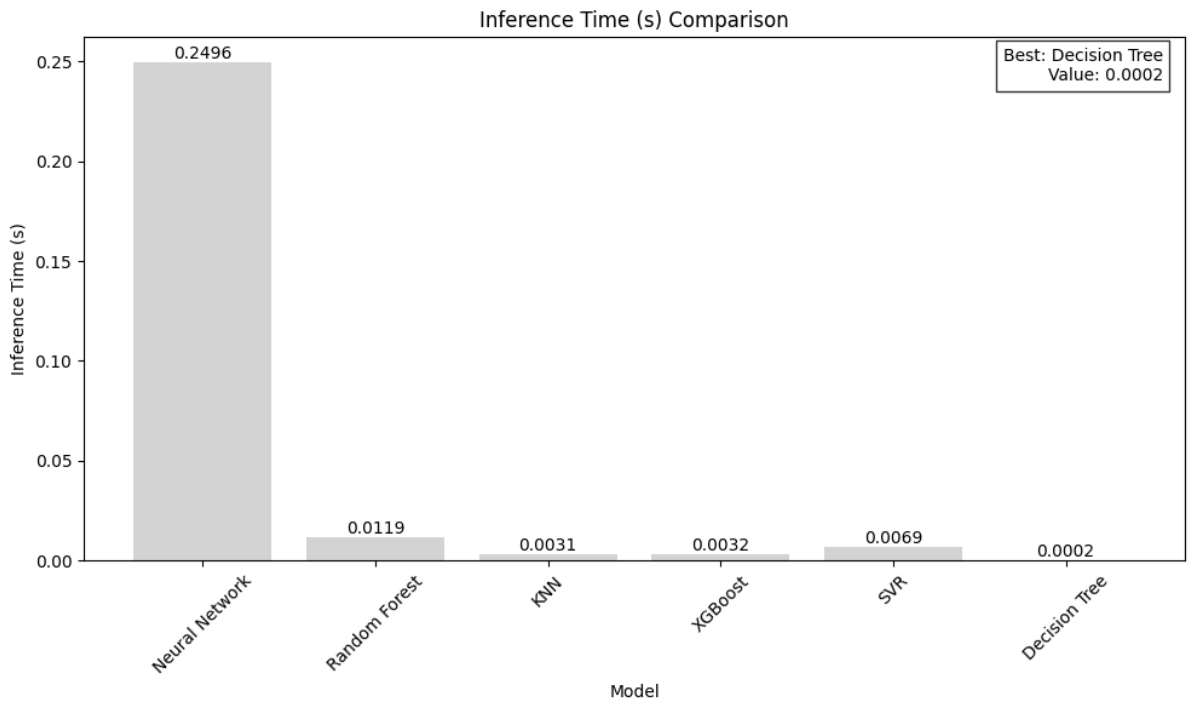


Figure 9: Çıkarım Süresi Karşılaştırması



## 4. Tartışma

### 4.1 Model Değiş Tokuşları

Çalışmamızda incelenen modeller arasında önemli performans ve kaynak kullanımı farklılıkları gözlemlenmiştir. Sinir Ağı modeli, karmaşık yapısı ve derin öğrenme yaklaşımı sayesinde özellikle uç durumlarda tutarlı sonuçlar vermektedir. Ancak bu başarısı, yüksek hesaplama maliyeti ile gelmektedir. Model, 81.52 saniyelik eğitim süresi ve 0.2143 saniyelik çıkarım süresiyle en yavaş seçenek olarak öne çıkmaktadır.

Geleneksel modeller ise hesaplama verimliliği konusunda önemli avantajlar sunmaktadır. Bu modeller arasında Rastgele Orman, 0.1437 RMSE ve 0.0206 MSE değerleriyle en yüksek doğruluk seviyesine ulaşmıştır. Bununla birlikte, geleneksel modellerin bazı uç durumlarda hata oranlarının yükseldiği gözlemlenmiştir. Bu durum, model seçiminde uygulama gereksinimlerinin dikkatle değerlendirilmesi gerektiğini göstermektedir.

### 4.2 Öneriler

Farklı uygulama senaryoları için model seçimi yaparken, performans ve kaynak gereksinimleri arasında optimal denge kurulmalıdır. Gerçek zamanlı uygulamalar için Karar Ağacı ve KNN modelleri ideal seçenekler olarak öne çıkmaktadır. Bu modeller, hızlı çıkarım süreleri ve kabul edilebilir doğruluk seviyeleri sayesinde anlık konum tahmini gerektiren senaryolarda etkili çözümler sunmaktadır.

Hassas konum tahmini gerektiren ve işlem süresinin kritik olmadığı uygulamalarda Rastgele Orman modeli tercih edilmelidir. Bu model, en düşük RMSE ve MSE skorları ile tutarlı ve güvenilir sonuçlar üretmektedir. Özellikle yüksek doğruluk gerektiren endüstriyel uygulamalar ve hassas konumlandırma sistemleri için ideal bir seçenektir.

Sınırlı donanım kaynaklarına sahip sistemler için Karar Ağacı modeli önerilmektedir. Bu model, minimum hesaplama gereksinimleri ve makul doğruluk seviyesi ile kaynak kısıtlı ortamlarda optimal bir denge sağlamaktadır. Özellikle gömülü sistemler ve mobil cihazlar gibi sınırlı kaynaklara sahip platformlarda tercih edilmelidir.

## 5. Sonuç

Çalışmamız, iç mekan konum tahmini için farklı makine öğrenmesi yaklaşımlarının kapsamlı bir karşılaştırmasını sunmaktadır. Yapılan analizler sonucunda, her modelin kendine özgü güçlü yönleri ve kısıtlamaları olduğu görülmüştür. Sinir ağı yaklaşımı uç durumlarda daha başarılı sonuçlar verirken, geleneksel makine öğrenmesi modelleri yüksek hesaplama verimliliği ve rekabetçi doğruluk seviyeleri sunmaktadır. Bu nedenle, model seçimi yapılırken uygulamanın spesifik gereksinimleri, doğruluk beklentileri, hesaplama kaynakları ve çıkarım hızı gibi faktörler bir bütün olarak değerlendirilmelidir. Sonuç olarak, başarılı bir iç mekan konumlandırma sistemi için doğru model seçimi, uygulama özelindeki ihtiyaçlar ve kısıtlamalar göz önünde bulundurularak yapılmalıdır.

## **Ek**

Tam performans metrikleri ekteki model\_comparison\_results.csv dosyasında mevcuttur.