

# WiFi CSI Verileri ile İç Mekan Konum Tahmini: Teknik Rapor

## İçindekiler

1. Giriş
2. Veri Seti Analizi
3. Veri Ön İşleme
4. Model Geliştirme
5. Sonuçlar ve Değerlendirme
6. Gelecek Çalışmalar

## 1. Giriş

Bu proje, WiFi Kanal Durum Bilgisi (CSI - Channel State Information) verilerini kullanarak iç mekanlarda hassas konum tespiti yapmayı amaçlamaktadır. CSI verileri, WiFi sinyallerinin iç mekanlarda nasıl yayıldığını gösteren zengin bir bilgi kaynağıdır.

### 1.1 Projenin Amacı

- İç mekan konumlandırma için makine öğrenmesi modelleri geliştirmek
- CSI verilerinden anlamlı özellikler çıkarmak
- Farklı model yaklaşımlarını karşılaştırmak
- 2-3 metre hassasiyetle konum tahmini yapmak

### 1.2 Teknik Altyapı

- Python 3.12 programlama dili
- Scikit-learn, TensorFlow gibi makine öğrenmesi kütüphaneleri
- NumPy, Pandas veri işleme kütüphaneleri
- Matplotlib, Seaborn görselleştirme araçları

## 2. Veri Seti Analizi

### 2.1 Veri Seti Yapısı

- 3 anten/alıcı
- Her anten için 30 alt taşıyıcı
- Her ölçüm için 1500 örnek
- CSI verileri sanal (imaginary) kısım olarak kaydedilmiş

#### 2.1.1 Veri Dizin Yapısı

- `imaginary_part/`: CSI ölçüm verilerinin sanal kısmını içerir
- `coordinate X-Y/`: Gerçek konum etiketlerini (metre cinsinden) içeren dizinler

- X ve Y, dosya aralığını belirtir (örn: coordinate 1-100)
- Her dosya, belirli bir konumdaki ölçümleri içerir
- Koordinat değerleri metre cinsindendir

**2.1.2 Verisetin Gerçek (Real) CSI Parçası** Veri seti analizi sırasında, Meeting Room Dataset, Lab Dataset ve miniLab dizinlerinde gerçek (real) CSI verilerinin bulunmadığı tespit edilmiştir. Tüm CSI ölçümleri sanal (imaginary) kısım olarak kaydedilmiştir. “coordinate” dizinleri, CSI verilerinin gerçek kısmını değil, ölçüm yapılan konumların fiziksel koordinatlarını içermektedir.

## 2.2 Sinyal Özellikleri

Genlik (Amplitude) Özellikleri: - Değer Aralığı: 0 - 55.72 - Ortalama: 19.35 - Standart Sapma: 10.18 - Alt taşıyıcılar arasında belirgin örüntüler

Faz (Phase) Özellikleri: - Değer Aralığı: -3.02 - 3.14 radyan (- - ) - Ortalama: 0.054 - Standart Sapma: 0.89 - Faz sarmalama örüntüleri

## 2.3 Veri Kalitesi

- Eksik veri yok
- Genlik değerleri iyi dağılmış
- Faz değerleri beklenen sarmalama davranışını gösteriyor
- Hem genlik hem faz örüntülerinde net yapı

## 3. Veri Ön İşleme

### 3.1 Faz Düzeltme

```
def phase_correction(phase_data):
    """Faz verilerini alt taşıyıcılar arasında düzeltir."""
    return np.unwrap(phase_data, axis=1)
```

### 3.2 Özellik Çıkarımı

Her anten için çıkarılan özellikler: - Genlik istatistikleri (ortalama, standart sapma, min, max, medyan) - Faz istatistikleri (ortalama, standart sapma, medyan) - Çeyreklik değerleri (Q1, Q3) - Zamansal özellikler (LSTM modeli için)

### 3.3 Veri Normalizasyonu

- StandardScaler kullanılarak özellikler normalize edildi
- Koordinat verileri ölçeklendirildi
- Aykırı değerler temizlendi

## 4. Model Geliştirme

### 4.1 Random Forest Regressor

```
rf_model = RandomForestRegressor(  
    n_estimators=1000,  
    max_depth=20,  
    min_samples_split=5,  
    min_samples_leaf=4,  
    max_features='sqrt',  
    bootstrap=True,  
    oob_score=True,  
    random_state=42,  
    n_jobs=-1  
)
```

Özellikler: - En tutarlı performans - Merkezi alanlarda daha iyi doğruluk - Muhafazakar tahminler

### 4.2 Gradient Boosting

```
gb_model = MultiOutputRegressor(  
    GradientBoostingRegressor(  
        n_estimators=300,  
        learning_rate=0.01,  
        max_depth=6,  
        min_samples_split=5,  
        min_samples_leaf=4,  
        subsample=0.8,  
        validation_fraction=0.2,  
        n_iter_no_change=10,  
        random_state=42  
    )  
)
```

Özellikler: - Random Forest'a benzer performans - Daha yüksek varyans - Yoğun eğitim verisi olan alanlarda daha iyi

### 4.3 Sinir Ağı

Mimarisi: - Giriş katmanı - 3 artık (residual) blok - Batch normalizasyon - Dropout (0.3) - Çıkış katmanı (2 nöron, X,Y koordinatları)

Özellikler: - Doğrusal olmayan örüntüleri yakalama - Değişken performans - Aşırı öğrenme eğilimi

#### 4.4 LSTM Modeli

Zamansal özellikleri kullanarak: - Hareket örüntülerini yakalama - Sıralı veri analizi - Yüksek tahmin varyansı

### 5. Sonuçlar ve Değerlendirme

#### 5.1 Model Performansları

**5.1.1 Hata Ölçümü ve Birimler** RMSE (Metre): Eğer test koordinatı ( $X_{\text{gerçek}}$ ,  $Y_{\text{gerçek}}$ ) ve tahmin edilen koordinat ( $X_{\text{tahmin}}$ ,  $Y_{\text{tahmin}}$ ) ise:

$$\text{RMSE} = \sqrt{\text{mean}((X_{\text{gerçek}} - X_{\text{tahmin}})^2 + (Y_{\text{gerçek}} - Y_{\text{tahmin}})^2)}$$

Koordinatlar metre cinsinden olduğundan, RMSE değeri de metre cinsindendir. Bu, modelin tahmin hatalarının doğrudan fiziksel mesafe olarak yorumlanabileceği anlamına gelir.

**5.1.2 Model Metrikleri** Elde edilen metrikler: - Random Forest: \* RMSE: 0.9811 metre \* MAE: 0.7293 metre \*  $R^2$  Skoru: 0.0392

- Gradient Boosting:
  - RMSE: 0.9761
  - MAE: 0.7310
  - $R^2$  Skoru: 0.0498
- Sinir Ağı:
  - RMSE: 1.1359
  - MAE: 0.8647
  - $R^2$  Skoru: -0.2876

#### 5.2 Model Karşılaştırması

1. Random Forest:
  - Tüm modeller arasında en tutarlı sonuçlar
  - $R^2$  skoru pozitif ancak düşük (0.0392)
  - En düşük MAE değeri (0.7293)
  - Tahminlerde düşük varyans
2. Gradient Boosting:
  - Random Forest'a çok yakın performans
  - En iyi  $R^2$  skoru (0.0498)
  - MAE değeri 0.7310
  - Orta düzey tahmin varyansı
3. Sinir Ağı:
  - En kötü performans gösteren model
  - Negatif  $R^2$  skoru (-0.2876)
  - En yüksek hata değerleri (RMSE: 1.1359)
  - Yüksek tahmin varyansı
  - Aşırı öğrenme belirtileri

#### 4. LSTM:

- Hareket tahmininde başarılı
- En yüksek varyans
- Zamansal ilişkileri yakalama

### 5.3 Hata Analizi

- Tüm modellerde beklenenden düşük performans
- PCA boyut indirgeme sonrası varyans kaybı (%46.41)
- Özellik mühendisliği sürecinde bilgi kaybı
- Koordinat tahminlerinde yüksek sapma
- 2B görselleştirmelerde belirgin tahmin hataları
- Modellerin tahmin yeteneklerinde ciddi sınırlamalar
- Özellikle sinir ağında aşırı öğrenme sorunları

## 6. Gelecek Çalışmalar

### 6.1 Veri Toplama

- Yüksek hatalı bölgelerde ek veri
- Çevresel özelliklerin eklenmesi
- Daha iyi gürültü filtreleme

### 6.2 Model İyileştirmeleri

- Topluluk yöntemleri
- Belirsizlik tahmini
- Hiperparametre optimizasyonu

### 6.3 Özellik Mühendisliği

- Gelişmiş faz düzeltme
- Ek zamansal özellikler
- Sinyal gücü göstergeleri

## Ekler

### Ek-1: Örnek Görselleştirmeler

Model tahminlerinin 2B görselleştirmeleri: - random\_forest\_2d.png: Random Forest modelinin tahmin sonuçları - gradient\_boosting\_2d.png: Gradient Boosting modelinin tahmin sonuçları - neural\_network\_2d.png: Sinir Ağı modelinin tahmin sonuçları

Her görselleştirmede: - Mavi noktalar: Gerçek konumlar - Kırmızı noktalar: Tahmin edilen konumlar - Gri çizgiler: Tahmin hatası mesafesi - Sağ üst köşe: Model performans metrikleri (RMSE, MAE, R<sup>2</sup>)

**Ek-2: Performans Grafikleri**

[Detaylı performans grafikleri docs/figures/ dizininde bulunmaktadır]

**Ek-3: Kod Dokümantasyonu**

Tüm kod tabanı Türkçe ve İngilizce dokümantasyon içermektedir.