**Kalori Tahmini Projesi: Teknik Rapor**

**Proje Amacı**

Bu proje, bireylerin fiziksel aktiviteleri sırasında yakılan kalori miktarını tahmin eden bir makine öğrenmesi modeli geliştirmek amacıyla tasarlanmıştır. Model, kullanıcının yaşı, boyu, kilosu gibi demografik özelliklerin yanı sıra, aktivite süreleri ve kalp atış hızı gibi fiziksel özelliklere dayanarak tahmin yapar. Bu çalışma, bireylerin kalori yönetimi ve egzersiz planlamasında rehberlik edebilecek bir aracın geliştirilmesine yöneliktir.

**Veri Seti ve Analizi**

**Veri Seti**

Proje kapsamında "exercise.csv" ve "calories.csv" adında iki ayrı dosya kullanılmıştır. Bu dosyalar birleştirilerek nihai veri seti oluşturulmuştur. Veri seti, 1500 bireyin fiziksel aktivite bilgilerini ve yakılan kalori miktarlarını içerir.

**Verinin Keşfedilmesi (EDA)**

* **Dağılım Analizi:** Tüm sayısal değişkenlerin dağılım grafikleri incelenmiş, veride aykırı değerlerin bulunup bulunmadığı kontrol edilmiştir.
* **Korelasyon Analizi:** Sayısal değişkenler arasındaki ilişkiler bir korelasyon matrisi ve ısı haritası ile görselleştirilmiştir. "Kalp Hızı" ve "Süre" gibi değişkenlerin kalori yakımıyla doğrudan ilişkili olduğu görülmüştür.
* **Kategorik Değişken Analizi:** Cinsiyet gibi kategorik değişkenlerin dağılımları görselleştirilmiş ve veri setindeki dengesi incelenmiştir.

**Veri Temizleme ve Önişleme**

1. **Gereksiz Sütunlar:** “User\_ID” gibi tahmin modeli için gereksiz olan sütunlar kaldırılmıştır.
2. **Eksik Veriler:** Veri setinde eksik bilgi bulunmadığı tespit edilmiştir.
3. **Kategorik Veriler:** "Gender" (cinsiyet) sütunu, 0 (kadın) ve 1 (erkek) şeklinde sayısal olarak kodlanmıştır.
4. **Veri Normalizasyonu:** Sayısal veriler, “StandardScaler” ile ölçeklendirilmiştir.
5. **Aykırı Değerler:** Boxplot analizleriyle tespit edilen aykırı değeler dikkate alınarak modeller eğitilmiştir.

**Modelleme**

**Kullanılan Modeller**

Proje kapsamında aşağıdaki modeller kullanılarak kalori tahmini yapılmıştır:

1. Linear Regression
2. Ridge Regression
3. Decision Tree Regressor
4. Random Forest Regressor
5. K-Nearest Neighbors Regressor
6. Gradient Boosting Regressor

**Performans Ölçütleri**

Modellerin performansı aşağıdaki metrikler kullanılarak karşılaştırılmıştır:

* **R² Score:** Modelin açıklayabildiği varyans oranı.
* **MSE (Ortalama Kare Hatası):** Tahminlerin sapmalarını kare ortalaması.
* **MAE (Ortalama Mutlak Hata):** Tahmin sapmalarının mutlak ortalaması.
* **Cross Validation Score:** Model genelleme performansı.

En iyi performansı “Random Forest Regressor” modeli göstermiştir. “GridSearchCV” ile bu modelin hiperparametre optimizasyonu yapılmış ve en iyi şu parametreler bulunmuştur:

* **n\_estimators:** 100
* **max\_depth:** 20
* **min\_samples\_split:** 5

Modellerin karşılaştırmalı sonuçları aşağıdaki gibi görselleştirilmiştir:

* “Random Forest Regressor” en yüksek R² skoru ile doğru tahminler sağlamıştır.
* Gradient Boosting ve Ridge Regression da yüksek doğruluk sergilemiştir.

**Model Kaydetme ve Kullanma**

Model, Joblib kütüphanesi ile kaydedilmiş ve aşağıdaki adımlar izlenerek kullanılmıştır:

1. Model ve scaler dosyası “calorie\_prediction\_model.pkl” ve “scaler.pkl” olarak kaydedilmiştir.
2. Rastgele bir kullanıcı verisi kullanılarak tahmin işlemi başarıyla gerçekleştirilmiştir.

**Örnek Tahmin**

* **Girdi Verileri:** Erkek, 30 yaşında, 175 cm boyunda, 70 kg ağırlığında, 45 dakika süreyle egzersiz yapmış, kalp atış hızı 120 bpm, vücut sıcaklığı 37°C.
* **Tahmin Sonucu:** Model, yaklaşık olarak 225 kcal kalori yakıldığını öngörmüştür.

**Arayüz Geliştirme**

Kullanıcı dostu bir deneyim sağlamak için Tkinter kullanılarak bir masaüstü arayüz tasarlanmıştır. Özellikleri:

1. Kullanıcıdan cinsiyet, yaş, boy, kilo, süre, kalp hızı ve vücut sıcaklığı gibi verileri toplar.
2. Bu verileri, kayıtlı scaler ve model kullanarak tahmin işlemi yapar.
3. Tahmin sonucu, bir bilgi kutusunda görüntülenir.
4. Kullanıcı girişi sırasında herhangi bir hata durumunda kullanıcıyı uyarır.

**Arayüz Detayları**

* **Girdi Alanları:** Kullanıcıdan alınan veriler arayüzdeki giriş kutuları ile girilmektedir.
* **Tahmin Butonu:** "Kalori Tahmini Yap" butonu tahmin sürecini başlatır ve tahmin sonuçlarını kullanıcıya bir mesaj kutusunda iletir.
* **Hata Yönetimi:** Kullanıcı hatalı veri girişi yaparsa bir hata mesajı gösterilir.

**Sonuç ve Öneriler**

* **Sonuçlar:** Random Forest Regressor, kalori tahmini için başarılı bir performans göstermiştir. Kullanılan arayüz ile model pratik bir şekilde kullanıma sunulmuştur.
* **Öneriler:**
  + Veri seti daha fazla özellik ile genişletilerek modelin doğruluğu artırılabilir.
  + İleri seviye derin öğrenme modelleri ile karşılaştırmalar yapılabilir.
  + Mobil veya web tabanlı bir uygulama geliştirilerek model daha geniş bir kitleye ulaştırılabilir.
  + Kullanıcının farklı egzersiz türleri için kalori tahminlerini görselleştirebileceği bir grafik arayüzü eklenebilir.