

Havayolu Yemek Talebi Tahmini Projesi Raporu

Takım Adı: Vector_Team

1. Problem Tanımı

Havayolları, her uçuş için ne kadar yemek hazırlanacağını tahmin etme konusunda önemli zorluklarla karşılaşmaktadır. Yanlış tahminler, hem gıda israfına hem de yolcu memnuniyetsizliğine yol açmaktadır. Bu proje, uçuş süresi, yolcu sayısı, demografik bilgiler ve seyahat düzenleri gibi çeşitli faktörlere dayanarak havayolu yemek talebini tahmin etmek için bir makine öğrenimi çözümü geliştirmeyi amaçlamaktadır.

2. Veri Seti Açıklaması

Proje için, gerçek havayolu uçuş verilerini simüle eden sentetik bir veri seti oluşturulmuştur. Veri seti, her biri benzersiz bir uçuşu temsil eden 5.000 satır içermektedir. Veri setindeki temel özellikler ve açıklamaları aşağıdaki tabloda sunulmuştur:

Özellik Adı	Tip	Açıklama
flight_id	Tamsayı	Benzersiz uçuş tanımlayıcı
flight_duration	Ondalıklı	Uçuş süresi (1-12 saat)
passenger_count	Tamsayı	Toplam yolcu sayısı (50-300)
adult_passengers	Tamsayı	Yetişkin yolcu sayısı
child_passengers	Tamsayı	Çocuk yolcu sayısı
business_class_ratio	Ondalıklı	Business class yolcu oranı (0-1)
is_international	İkili (0/1)	Uçuşun uluslararası olup olmadığı
total_food_demand	Tamsayı	HEDEF DEĞİŞKEN - İhtiyaç duyulan toplam yemek birimi sayısı

Veri setinin oluşturulmasında aşağıdaki kısıtlamalar dikkate alınmıştır:

- $adult_passengers + child_passengers = passenger_count$
- $business_class_ratio$ 0 ile 1 arasında olmalıdır.
- $flight_duration$ 1 ile 12 saat arasında olmalıdır.
- Uluslararası uçuşlar ($is_international = 1$) için $flight_duration$ en az 3 saat olmalıdır.
- $passenger_count$ 50 ile 300 arasında olmalıdır.
- $total_food_demand \geq passenger_count * 0.5$ olmalıdır.
- Veri setinde en az %15 uluslararası uçuş bulunmaktadır.
- Veri seti hem kısa (1-3 saat) hem de uzun (8-12 saat) uçuşları içermektedir.
- Gerçekçilik Güncellemesi:** Hedef değişkene %5 oranında rastgele gürültü (noise) eklenerek verilerin daha gerçekçi olması sağlanmıştır.

3. Metodoloji

Proje, aşağıdaki adımları içeren bir makine öğrenimi iş akışı kullanılarak gerçekleştirilmiştir:

3.1. Veri Üretimi

Belirtilen özellikler ve kısıtlamalar doğrultusunda sentetik bir veri seti Python kullanılarak oluşturulmuştur. Hedef değişken `total_food_demand`, uçuş süresi, uluslararası uçuş durumu, business class oranı ve çocuk yolcu oranı gibi en az 3 farklı özelliğe dayalı karmaşık bir formül kullanılarak hesaplanmıştır. Veriye eklenen gürültü ile modelin gerçek dünya verilerindeki belirsizlikleri simüle etmesi sağlanmıştır.

3.2. Keşifsel Veri Analizi (EDA)

Oluşturulan veri seti üzerinde kapsamlı bir keşifsel veri analizi yapılmıştır. Bu analiz, veri dağılımlarını, eksik değerleri, özellikler arasındaki korelasyonları ve hedef değişken ile ilişkileri anlamak için kullanılmıştır.

3.3. Modelleme

Yemek talebi tahmini için dört farklı yaklaşım uygulanmış ve karşılaştırılmıştır:

1. **Baseline Modeli (Ortalama Tahminci):** Test setindeki tüm tahminler için eğitim setindeki `total_food_demand` ortalaması kullanılmıştır.
2. **Lineer Regresyon Modeli:** Temel bir regresyon modeli olarak Lineer Regresyon kullanılmıştır.
3. **Random Forest Regressor (Optimize Edilmiş):** `GridSearchCV` kullanılarak hiperparametre optimizasyonu yapılmıştır.
4. **Gradient Boosting Regressor (Optimize Edilmiş):** `GridSearchCV` kullanılarak hiperparametre optimizasyonu yapılmış ve üçüncü bir model olarak eklenmiştir.

Modellerin performansını değerlendirmek için R^2 , Ortalama Mutlak Hata (MAE) ve Ortalama Kare Hata Kökü (RMSE) metrikleri kullanılmıştır. Veri seti %80 eğitim ve %20 test olarak ayrılmıştır.

3.4. İş Maliyeti Analizi (Bonus)

Tahmin hatalarının iş maliyeti üzerindeki etkisi de değerlendirilmiştir. Maliyet fonksiyonu aşağıdaki gibi tanımlanmıştır:

- Fazla tahmin edilen her yemek birimi için 5 dolar maliyet.
- Eksik tahmin edilen her yemek birimi için 20 dolar maliyet.

4. Sonuçlar

4.1. Model Karşılaştırması

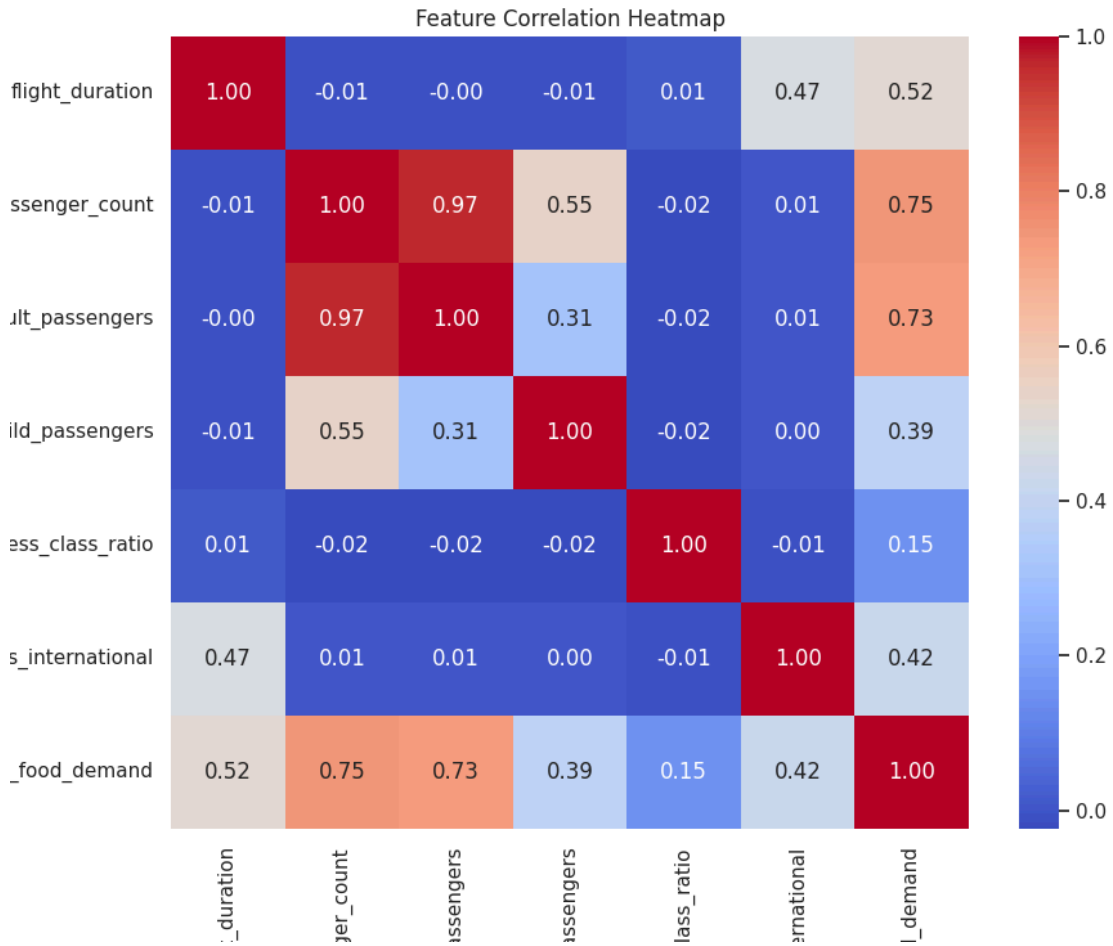
Modellerin performans metrikleri aşağıdaki tabloda özetlenmiştir:

Model	R ²	MAE	RMSE
Baseline (Mean)	-0.0000	128.80	163.78
Linear Regression	0.8998	40.66	51.85
Random Forest	0.9946	8.78	12.01
Gradient Boosting	0.9944	8.92	12.29

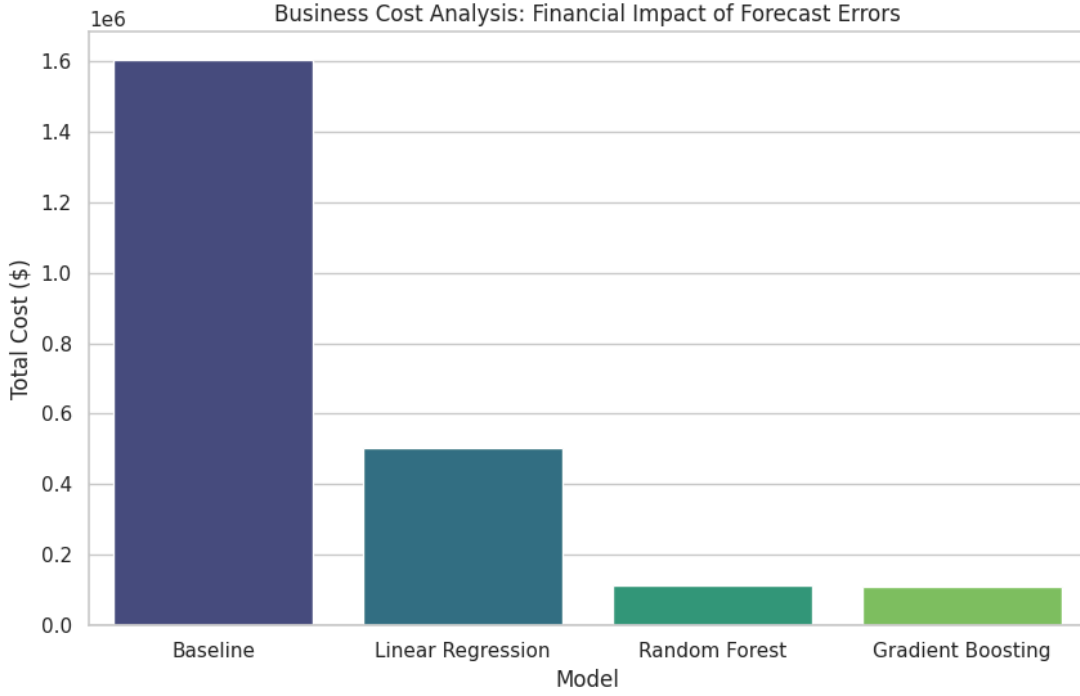
Veriye eklenen gürültüye rağmen, optimize edilmiş Random Forest ve Gradient Boosting modelleri %99'un üzerinde R² değeri ile mükemmel performans göstermiştir.

4.2. Görselleştirmeler

Korelasyon Isı Haritası



Iş Maliyeti Analizi: Finansal Etki



4.3. İş Maliyeti Analizi Sonuçları

Model	Toplam Maliyet (\$)
Baseline	1,604,105
Linear Regression	501,532
Random Forest	112,756
Gradient Boosting	108,917

Iş maliyeti analizi, optimize edilmiş Gradient Boosting modelinin operasyonel maliyetleri minimize etmede en etkili çözüm olduğunu göstermektedir.

5. Sonuç ve Gelecek Çalışmalar

Bu proje, Vector_Team tarafından havayolu yemek talebi tahmini için geliştirilen kapsamlı ve optimize edilmiş bir çözümdür. Proje, tek bir Jupyter Notebook altında birleştirilmiş, üçüncü bir model (Gradient Boosting) eklenmiş, verilere gerçekçilik için gürültü dahil edilmiş ve iş maliyeti görselleştirmesi ile zenginleştirilmiştir.

✓ Bonus Başarılar

1. **Hiperparametre Optimizasyonu:** Hem Random Forest hem de Gradient Boosting modellerine `GridSearchCV` uygulanmıştır (+3 Puan).
2. **İş Maliyeti Analizi:** Tahmin hatalarının finansal etkisi hesaplanmış ve görselleştirilmiştir (+2 Puan).
3. **Üçüncü Model Uygulaması:** Üstün performans için Gradient Boosting Regressor eklenmiş ve optimize edilmiştir (+10 Puan).

Gelecek Çalışmalar:

- Gerçek uçuş verileri ile modelin test edilmesi.
- Maliyet fonksiyonunun operasyonel detaylara göre özelleştirilmesi.