



YAPAY ZEKA UZMANLIK PROGRAMI BİTİRME PROJESİ

PROJE ADI: PYTHON'DA MAKİNE ÖĞRENMESİ

PROJE YÜRÜTÜCÜSÜ: HABİB ŞAKO

MENTÖR: BEHLUL KIZILTAŞ

RAPOR TARİHİ: 14.06.2024



Havayolu Rezervasyonları Tamamlama ve Ek Hizmet Talep Analizi

Özet

Havayolu sektörü, müşteri memnuniyetini artırmak ve operasyonel verimliliği maksimize etmek için veri odaklı karar verme süreçlerine giderek daha fazla önem vermektedir. Bu bağlamda, yolcu davranışlarını ve tercihlerini anlamak, havayolu şirketlerinin hem gelirlerini artırmalarına hem de müşteri deneyimini iyileştirmelerine yardımcı olabilir. Özellikle rezervasyon tamamlama, ekstra bagaj talepleri ve uçak içi yemek talepleri gibi kritik hizmetlerin doğru tahmin edilmesi, hem operasyonel hem de stratejik avantajlar sağlamaktadır.

Çalışmanın, havayolu şirketlerinin müşteri davranışlarını doğru bir şekilde tahmin edebilecek modeller geliştirebilmesi ve bu tahminlerin operasyonel ve stratejik karar alma süreçlerinde etkin bir şekilde kullanılmasıdır. Bu doğrultuda, yapılan tahminlerin havayolu şirketlerine nasıl fayda sağlayabileceği ve bu tahminlerin nerede ve nasıl kullanılabileceği üzerinde durulmuştur. Bu çalışma, havayolu sektöründe veri analitiği ve makine öğrenimi uygulamalarının önemini vurgularken, aynı zamanda şirketlerin rekabet avantajı elde etmelerine ve müşteri memnuniyetini artırmalarına yardımcı olmayı amaçlamaktadır.

Ayrıca hangi ek hizmetlerin daha fazla talep gördüğünü belirlemek ve bu hizmetlerin talebi ile gelir arasındaki ilişkiyi analiz etmek için ek hizmetlerden kaynaklanan geliri artırmak için stratejiler önerilecektir.

Veri Seti

Bu veri seti, yolcu sayısı, satış kanalı, seyahat türü, satın alma teslim süresi, konaklama süresi, uçuş saati, uçuş günü, rota, rezervasyon menşei ve müşterinin ekstra bagaj isteyip istemediğini belirten çeşitli bayraklar dahil olmak üzere havayolu rezervasyonları hakkında bilgiler içerir. Tercih edilen koltuk veya uçakta yemek. Veri seti ayrıca toplam uçuş süresini ve rezervasyonun tamamlanıp tamamlanmadığını gösteren bir bayrağı da içerir.

Veri setinin içeriğinde bulunan sütunların adları ve tuttuğu bilgiler:

- num_passengers= seyahat eden yolcu sayısı
- sales_channel= satış kanalı rezervasyonu şu tarihte yapıldı:
- trip_type= Yolculuk Tipi (Gidiş-Dönüş, Tek Yön, Dairesel Yolculuk)
- purchase_lead= seyahat tarihi ile rezervasyon tarihi arasındaki gün sayısı
- length_of_stay= varış yerinde geçirilen gün sayısı
- flight_hour= uçuşun kalkış saati
- flight_day= uçuşun kalkış haftasının günü
- route= menşei-> varış uçuş rotası
- booking_origin= rezervasyonun yapıldığı ülke
- wants_extra_baggage= eğer müşteri rezervasyon sırasında ekstra bagaj istediye
- wants_preferred_seat= eğer müşteri rezervasyon sırasında tercih edilen bir koltuk istiyorsa
- wants_in_flight_meals= müşteri rezervasyon sırasında uçakta yemek istiyorsa
- flight_duration= toplam uçuş süresi (saat cinsinden)
- booking_complete= müşterinin rezervasyonu tamamlayıp tamamlamadığını gösteren bayrak



3.Giriş

Bu bitirme projesinin amacı, havayolu rezervasyonlarının tamamlanma olasılığını tahmin eden bir model geliştirmektir. Projede kullanılacak veri setinde, müşterilerin rezervasyon yapma ve tamamlama davranışları ile ilgili çeşitli değişkenler bulunmaktadır. Bu değişkenler kullanılarak, müşterilerin rezervasyonlarını tamamlama olasılıkları analiz edilecek ve tamamlanmamış rezervasyonları azaltmaya yönelik stratejiler önerilecektir.

Temel Sorun: Rezervasyon Tamamlama Olasılığı

Havayolu şirketleri, müşteri rezervasyonlarının tamamlanmaması durumunda ciddi gelir kayıpları yaşayabilirler. Tamamlanmamış rezervasyonlar hem koltukların boş kalmasına hem de müşteri memnuniyetsizliğine yol açabilir. Bu projenin temel amacı, veri setindeki bilgilerden yararlanarak müşterilerin rezervasyonlarını tamamlama olasılığını tahmin etmektir. Bu olasılığı doğru bir şekilde tahmin edebilmek, havayolu şirketlerinin müşteri ilişkileri yönetimini iyileştirmelerine ve gelirlerini artırmalarına yardımcı olacaktır.

Ek Sorun: Ek Hizmetlerin Talep ve Gelir Üzerindeki Etkisi

Projenin ikinci aşamasında, havayolu şirketinin sunduğu ek hizmetlerin (örneğin, bagaj hizmetleri, yiyecek ve içecek hizmetleri) hangi derecede talep gördüğü ve bu hizmetlerin şirketin gelirleri üzerindeki etkisi analiz edilecektir. Müşterilerin hangi ek hizmetlere daha fazla ilgi gösterdiği ve bu hizmetlerin talebi ile gelir arasındaki ilişkinin belirlenmesi, havayolu şirketlerine ek gelir kaynaklarını optimize etme fırsatı sunacaktır.

Veri Setindeki Bilgiler ve Çözölmek İstenen Sorunlar

Rezervasyon Tamamlama Olasılığı: Müşterilerin rezervasyonlarını tamamlama olasılığını hangi faktörler etkiliyor? Bu faktörler arasında demografik bilgiler, rezervasyon kanalı, uçuş detayları ve ek hizmetlerin rolü nedir?

Ek Hizmetlerin Talebi: Hangi ek hizmetler daha fazla talep ediliyor? Bu hizmetlerin talebi, müşteri segmentlerine göre nasıl değişiyor?

Gelir Analizi: Ek hizmetlerin talebi ile havayolu şirketinin gelirleri arasındaki ilişkiyi optimize etmek için hangi stratejiler geliştirilebilir?

Hedef ve Stratejiler

Bu projede geliştirilecek model ve analizler sonucunda, havayolu şirketleri için şu stratejiler önerilecektir:

Rezervasyon Tamamlama Stratejileri: Müşterilerin rezervasyonlarını tamamlama olasılığını artırmak için hedefli pazarlama kampanyaları ve teşvik programları oluşturulabilir. Örneğin, rezervasyonlarını tamamlama olasılığı düşük olan müşteri segmentlerine özel indirimler veya esnek rezervasyon seçenekleri sunulabilir.

Ek Hizmet Stratejileri: Talebi yüksek olan ek hizmetlerin tanıtımına yönelik kampanyalar düzenlenebilir ve bu hizmetlerin fiyatlandırma stratejileri gözden geçirilebilir. Ayrıca, ek hizmetlerin paket halinde sunulması ve bu paketlerin farklı müşteri segmentlerine özel olarak uyarlanması, gelirlerin artırılmasına katkıda bulunabilir.

Bu şekilde, havayolu şirketleri hem rezervasyon tamamlama oranlarını artırabilir hem de ek hizmetlerden elde edilen gelirleri maksimize edebilirler. Projenin nihai amacı, havayolu şirketlerinin operasyonel verimliliğini ve müşteri memnuniyetini artırarak rekabet avantajı sağlamalarına yardımcı olmaktır.



4. Materyal & Metot

Bu bölümde, projenin gerçekleştirilmesi için kullanılan yöntemler ve adımlar detaylı olarak açıklanmaktadır.

Veri Ön İşleme:

- Veri setindeki eksik ve hatalı verilerin temizlenmesi.
- Kategorik verilerin kodlanması ve gerekirse ölçeklendirilmesi.

Keşifsel Veri Analizi (EDA):

- Değişkenlerin dağılımlarının incelenmesi.
- Rezervasyon tamamlama durumuyla ilişkili olabilecek değişkenlerin tespit edilmesi.

Model Geliştirme ve Değerlendirme:

- Farklı makine öğrenimi algoritmalarının (lojistik regresyon, karar ağaçları, random forest, vb.) kullanılarak modellerin geliştirilmesi.
- Modellerin performanslarının doğruluk, hassasiyet, özgüllük gibi metriklerle değerlendirilmesi.

İlişki Analizi ve Strateji Önerileri:

- Rezervasyon tamamlama durumu ile diğer değişkenler arasındaki ilişkilerin korelasyon ve regresyon analizleri ile incelenmesi.
- Tamamlanmamış rezervasyonları azaltmaya yönelik, örneğin belirli ülkeler veya belirli ek hizmetlerin talebi gibi faktörlere dayalı stratejiler geliştirilmesi.

Ek Hizmet Talebi ve Gelir Analizi:

- Ek hizmet taleplerinin analiz edilmesi ve hangi ek hizmetlerin daha fazla talep gördüğünün belirlenmesi.
- Ek hizmet talebi ile gelir arasındaki ilişkinin analiz edilmesi.
- Ek hizmetlerden kaynaklanan geliri artırmak için stratejiler önerilmesi.



5. Makine Öğrenimi Modeli Seçimi

1. Rezervasyon Tamamlama Olasılığı: Karar Ağaçları, Lojistik Regresyon veya Random Forest.

Rezervasyonların tamamlanma olasılığını tahmin etmek için karar ağaçları, lojistik regresyon ve random forest modelleri kullanılabilir. Karar ağaçları ve random forest modelleri, veri

özelliklerine dayalı olarak güçlü sınıflandırma yetenekleri sunarken, lojistik regresyon, rezervasyon tamamlama olasılığını doğrudan tahmin etmek için etkili bir yöntem sağlar.

Hedef değişken: 'booking_complete '

2. Ek Hizmetlerin Talebi: K-Means veya GMM.

Yolcuların ek hizmet taleplerini değerlendirmek için K-Means ve Gaussian Mixture Models (GMM) kullanılabilir. Bu modeller, yolcuları benzer talep davranışlarına göre kümelere ayırarak ek hizmetlerin (ekstra bagaj, tercih edilen koltuk, uçak içi yemekler) talebini anlamaya ve optimize etmeye yardımcı olabilir.

Hedef değişken: 'wants_extra_baggage ', 'wants_preferred_seat ', 'wants_in_flight_meals '

6. Yapılan Çalışmalar

Veri Setinin Yüklenmesi:

- Veri seti yüklenirken, Python'ın varsayılan karakter kodlaması olan UTF-8 ile veri setinin Latin1 karakter kodlaması arasında uyumsuzluk yaşandı. Bu nedenle, encoding='latin1' kullanılarak veri seti başarıyla yüklendi.

Boş Değerlerin Kontrolü:

- Veri setinde boş değerlerin olup olmadığı kontrol edildi ve boş değer bulunmadı.

Kategorik Verilerin Analizi:

- Kategorik sütunların unique değerleri incelendi. booking_origin sütununda istenmeyen 'R\&2union' ve '(not set)' değerleri tespit edildi. 'R\&2union' değeri Reunion olarak değiştirildi.

Kategorik Değerlerin Sayısal Değerlere Çevirimi:

- trip_type sütunu one-hot encoding yöntemi ile dönüştürüldü.
- Diğer dört sütun label encoder ile sayısal değerlere dönüştürüldü.
- One-hot encoding sonucunda veri setine üç yeni sütun eklendi.
- Bu işlemler sonucunda, boş değeri bulunmayan ve kategorik değerlerin sayısal değerlere çevrildiği temiz bir veri seti elde edildi.

Aykırı Değerler (Outliers) Tespit Edilmesi, Analizinin Yapılması ve Kaldırılması:

- Tespit edilen aykırı değerler incelenerek, veri setine olan etkileri değerlendirilecek ve gerekli kısımları veri setinden çıkarıldı.

Feature Engineering :

- Özellik seçimi, makine öğrenme algoritmalarını eğitmek için en uygun özelliklerin seçilmesi planlanmaktadır. Bu işlem, model performansını artırmak, hesaplama maliyetini azaltmak ve modelin anlaşılabilirliğini iyileştirmek amacıyla yapılacaktır. Özellik çıkarımı yeni özellikler



oluşturulacak ve mevcut özelliklerin daha anlamlı temsilleri elde edilecektir. Özellik inşası var olan özelliklerden yeni özellikler türetilerek modelin tahmin yeteneği artırılacaktır.

- **Çarpan Özellik (Interaction Feature):**

- route_booking_origin: route ve booking_origin sütunlarının çarpımı. Bu, iki özelliğin etkileşiminden yeni bir özellik yaratır.
- flight_hour_day: flight_hour ve flight_day sütunlarının çarpımı. Bu, uçuş saatinin ve uçuş gününün etkileşiminden yeni bir özellik oluşturur.

- **Logaritmik Dönüşüm (Log Transformation):**

log_purchase_lead: purchase_lead sütununun logaritmasını alarak yeni bir özellik oluşturur. Bu, değerlerin daha küçük bir aralığa sıkıştırılmasını sağlar ve büyük değişkenlik gösteren değerlerin etkisini azaltır. 1 eklenmesinin sebebi logaritma işleminin 0 veya negatif değerler için tanımlı olmamasıdır.

- **Oran Özelliği (Ratio Feature):**

lead_stay_ratio: purchase_lead sütununun length_of_stay sütununa bölünmesiyle oluşturulan bir oran. Bu oran, satın alma öncesi geçen sürenin konaklama süresine oranını ifade eder. 1 eklenmesinin sebebi bölme işleminde 0 hatasını önlemektir.

Veri Setinin Hazırlanması: Hedef Değişkenlerin ve Özelliklerin Ayrılması:

- Orijinal Veri Seti
 - Orijinal veri setinde, hedef değişkenler ve özellikler aşağıdaki gibi ayrıştırılmıştır:
 - X: Özellikler. Bu değişken, booking_complete, wants_in_flight_meals ve wants_extra_baggage sütunları dışındaki tüm sütunları içerir.
 - y_booking_complete: booking_complete hedef değişkeni.
 - y_flight_meals: wants_in_flight_meals hedef değişkeni.
 - y_extra_baggage: wants_extra_baggage hedef değişkeni.
- Aykırı Değerleri Çıkarılmış Veri Seti
 - Aykırı değerlerin veri setinden çıkarılması, model performansını artırmak için kritik bir adımdır. Aykırı değerleri çıkarılmış veri setinde, hedef değişkenler ve özellikler şu şekilde belirlenmiştir:
 - X_no_outliers: Aykırı değerleri çıkarılmış veri setinde, özellikler.
 - y_booking_complete_no_outliers: Aykırı değerleri çıkarılmış veri setinde booking_complete hedef değişkeni.
 - y_flight_meals_no_outliers: Aykırı değerleri çıkarılmış veri setinde wants_in_flight_meals hedef değişkeni.
 - y_extra_baggage_no_outliers: Aykırı değerleri çıkarılmış veri setinde wants_extra_baggage hedef değişkeni.
- Özellik Mühendisliği Uygulanmış Veri Seti
 - Özellik mühendisliği, veri setindeki önemli değişkenlerin ortaya çıkarılması ve modelin tahmin performansının artırılması için yapılan bir işlemdir. Özellik mühendisliği uygulanmış veri setinde, hedef değişkenler ve özellikler şu şekilde belirlenmiştir:
 - X_fe: Özellik mühendisliği uygulanmış veri setinde, özellikler.
 - y_booking_complete_fe: Özellik mühendisliği uygulanmış veri setinde booking_complete hedef değişkeni.



- y_flight_meals_fe: Özellik mühendisliği uygulanmış veri setinde wants_in_flight_meals hedef değişkeni.
- y_extra_baggage_fe: Özellik mühendisliği uygulanmış veri setinde wants_extra_baggage hedef değişkeni.
- Aykırı Değerleri Çıkarılmış ve Özellik Mühendisliği Uygulanmış Veri Seti
 - Bu veri setinde, hem aykırı değerler çıkarılmış hem de özellik mühendisliği uygulanmıştır. Bu sayede, verinin temizlenmesi ve önemli özelliklerin ortaya çıkarılması sağlanmıştır. Bu veri setinde hedef değişkenler ve özellikler aşağıdaki gibi ayrıştırılmıştır:
 - X_no_outliers_fe: Aykırı değerleri çıkarılmış ve özellik mühendisliği uygulanmış veri setinde, özellikler.
 - y_booking_complete_no_outliers_fe: Aykırı değerleri çıkarılmış ve özellik mühendisliği uygulanmış veri setinde booking_complete hedef değişkeni.
 - y_flight_meals_no_outliers_fe: Aykırı değerleri çıkarılmış ve özellik mühendisliği uygulanmış veri setinde wants_in_flight_meals hedef değişkeni.
 - y_extra_baggage_no_outliers_fe: Aykırı değerleri çıkarılmış ve özellik mühendisliği uygulanmış veri setinde wants_extra_baggage hedef değişkeni.
- Bu işlemler sonucunda, dört farklı veri seti oluşturulmuş ve her biri model eğitimi ve değerlendirmesi için hazır hale getirilmiştir. Bu veri setleri, orijinal veri seti, aykırı değerler çıkarılmış veri seti, özellik mühendisliği uygulanmış veri seti ve hem aykırı değerler çıkarılmış hem de özellik mühendisliği uygulanmış veri seti olarak adlandırılmıştır. Bu veri setleri, model performanslarının karşılaştırılmasına ve en iyi sonuç veren modelin seçilmesine olanak tanır.

Model Eğitim ve Değerlendirme Süreci: En İyi Modelin Belirlenmesi:

- En İyi Model ve Doğruluk Değerlerini Saklamak İçin Değişkenler
 - Her veri seti için en iyi model ve doğruluk değerini saklamak amacıyla bir sözlük yapısı kullanılmıştır. Bu sözlük, original, no_outliers, feature_engineered ve no_outliers_fe veri setlerini ve bu veri setleri için en iyi modelleri ve doğruluk değerlerini içerir.
- Eğitim ve Değerlendirme Fonksiyonları
 - Model eğitimi ve değerlendirmesi için iki ana fonksiyon geliştirilmiştir: train_and_evaluate ve train_and_evaluate_clustering. Bu fonksiyonlar sırasıyla denetimli (supervised) ve denetimsiz (unsupervised) öğrenme modellerinin eğitim ve değerlendirme işlemlerini gerçekleştirir.
- Modellerin Eğitim ve Değerlendirilmesi
 - Her veri seti için modellerin eğitimi ve değerlendirilmesi evaluate_models fonksiyonu ile yapılmıştır. Bu fonksiyon, her bir hedef değişken (booking_complete, wants_in_flight_meals, wants_extra_baggage) için modelleri eğitir ve değerlendirir. Ayrıca, en iyi modeli ve doğruluk değerini belirler.
- Orijinal Veri Seti Sonuçları
 - Orijinal veri seti üzerinde yapılan model eğitimi ve değerlendirme sonuçları elde edilmiştir. Sonuçlar, her bir modelin doğruluk, Hamming kaybı, Hausdorff mesafesi ve eğitim-test süreleri gibi metriklerle değerlendirilmiştir. En iyi model, doğruluk değerine göre belirlenmiştir.



- **Aykırı Değerler Çıkarılmış Veri Seti Sonuçları**
 - Aykırı değerlerin çıkarıldığı veri seti üzerinde yapılan model eğitimi ve değerlendirme sonuçları elde edilmiştir. Bu veri seti üzerinde de aynı metrikler kullanılarak modeller değerlendirilmiş ve en iyi model belirlenmiştir.
- **Özellik Mühendisliği Uygulanmış Veri Seti Sonuçları**
 - Özellik mühendisliği uygulanmış veri seti üzerinde yapılan model eğitimi ve değerlendirme sonuçları elde edilmiştir. Bu süreçte, özellik mühendisliği ile veri setine eklenen yeni özelliklerin modellerin performansına etkisi incelenmiştir.
- **Aykırı Değerler Çıkarılmış ve Özellik Mühendisliği Uygulanmış Veri Seti Sonuçları**
 - Aykırı değerler çıkarılmış ve özellik mühendisliği uygulanmış veri seti üzerinde yapılan model eğitimi ve değerlendirme sonuçları elde edilmiştir. Bu veri seti üzerinde yapılan değerlendirmeler, her iki veri hazırlama adımının birleştirilmesinin modellerin performansına etkisini ortaya koymuştur.
- **Sonuçların Yazdırılması**
 - Sonuçlar, her veri seti için ayrı ayrı yazdırılmış ve değerlendirilmiştir. Örneğin, orijinal veri seti için elde edilen sonuçlar, her bir modelin doğruluk, Hamming kaybı vb. metrikler üzerinden ayrıntılı bir şekilde sunulmuştur. Benzer şekilde, diğer veri setleri için de sonuçlar ayrıntılı olarak değerlendirilmiştir.
- **En İyi Modeller ve Doğruluk Değerleri**
 - Her veri seti için en iyi model ve doğruluk değerleri belirlenmiş ve raporlanmıştır. Bu adımlar sonucunda, farklı veri setleri üzerinde yapılan model eğitim ve değerlendirme işlemleri başarıyla tamamlanmış ve en iyi modeller belirlenmiştir. Bu çalışma, veri hazırlama ve model seçimi süreçlerinin etkili bir şekilde yönetilmesi ve optimize edilmesi konusunda önemli bulgular sunmaktadır.



7. Sonuçlar

Bu çalışmada, bir havayolu firmasının müşteri davranışlarını tahmin etmek amacıyla Kaggle'dan elde edilen bir veri seti kullanılarak booking_complete, wants_extra_baggage ve wants_in_flight_meals sütunları için çeşitli makine öğrenimi modelleriyle tahminler yapılmıştır. Logistic Regression ve Random Forest modelleri, genellikle en iyi performansı göstermiştir. Çeşitli veri ön işleme teknikleri ve model değerlendirmeleri sonucunda elde edilen bulgular ve bu bulguların havayolu şirketleri için potansiyel kullanım alanları aşağıda detaylandırılmıştır.

Orijinal Veri Seti Sonuçları

Booking Complete Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.758	0.242	0.000
LogisticRegression	0.850	0.150	1.000
RandomForest	0.850	0.150	0.000
KMeans	0.707	0.293	0.014
GMM	0.839	0.161	0.019

Wants In-Flight Meals Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.591	0.409	0.000
LogisticRegression	0.674	0.326	0.000
RandomForest	0.679	0.321	0.000
KMeans	0.544	0.456	0.011
GMM	0.571	0.429	0.017

Wants Extra Baggage Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.636	0.364	0.000
LogisticRegression	0.723	0.277	0.000
RandomForest	0.724	0.276	0.000
KMeans	0.389	0.611	0.012
GMM	0.339	0.661	0.031

Aykırı Değerler Çıkarılmış Veri Seti Sonuçları

Booking Complete Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.760	0.240	0.000
LogisticRegression	0.850	0.150	1.000
RandomForest	0.850	0.150	0.000
KMeans	0.692	0.308	0.011



T.C.
SANAYİ VE TEKNOLOJİ BAKANLIĞI
MİLLİ TEKNOLOJİ GENEL MÜDÜRLÜĞÜ



GMM	0.826	0.174	0.022
-----	-------	-------	-------

Wants In-Flight Meals Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.591	0.409	0.000
LogisticRegression	0.671	0.329	0.000
RandomForest	0.673	0.326	0.000
KMeans	0.543	0.457	0.031
GMM	0.574	0.426	0.020

Wants Extra Baggage Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.623	0.378	0.000
LogisticRegression	0.711	0.289	0.000
RandomForest	0.712	0.288	0.000
KMeans	0.413	0.586	0.028
GMM	0.347	0.654	0.023

Özellik Mühendisliği Uygulanmış Veri Seti Sonuçları

Booking Complete Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.758	0.242	0.000
LogisticRegression	0.850	0.150	1.000
RandomForest	0.850	0.150	0.000
KMeans	0.709	0.291	0.016
GMM	0.753	0.247	0.022

Wants In-Flight Meals Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.586	0.414	0.000
LogisticRegression	0.673	0.327	0.000
RandomForest	0.678	0.322	0.000
KMeans	0.544	0.457	0.013
GMM	0.546	0.454	0.025

Wants Extra Baggage Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.635	0.365	0.000
LogisticRegression	0.722	0.278	0.000
RandomForest	0.725	0.275	0.000
KMeans	0.388	0.612	0.013
GMM	0.354	0.654	0.023

Aykırı Değerler Çıkarılmış ve Özellik Mühendisliği Uygulanmış Veri Seti Sonuçları

Booking Complete Tahmini Sonuçları:



Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.759	0.241	0.000
LogisticRegression	0.850	0.150	1.000
RandomForest	0.850	0.150	0.000
KMeans	0.691	0.309	0.011
GMM	0.826	0.174	0.022

Wants In-Flight Meals Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.595	0.405	0.000
LogisticRegression	0.671	0.329	0.000
RandomForest	0.674	0.326	0.000
KMeans	0.543	0.457	0.031
GMM	0.574	0.426	0.020

Wants Extra Baggage Tahmini Sonuçları:

Model	Accuracy	Hamming Loss	Hausdorff Mesafesi
DecisionTree	0.622	0.378	0.000
LogisticRegression	0.711	0.289	0.000
RandomForest	0.712	0.288	0.000
KMeans	0.413	0.587	0.028
GMM	0.347	0.653	0.023

Booking Complete Tahminleri

Sonuçlar:

Logistic Regression ve Random Forest modelleri en yüksek doğruluk oranlarına (yaklaşık %85) sahip olup, Hamming Loss değerleri düşüktür (yaklaşık 0.15).

Diğer modellerin performansı görece düşük olup, özellikle KMeans ve GMM modelleri beklenenin altında kalmıştır.

Öneriler:

Pazarlama Stratejileri: Yüksek doğruluk oranına sahip modeller, müşteri rezervasyonlarının tamamlanma olasılığını doğru bir şekilde tahmin edebilir. Bu tahminler, pazarlama kampanyalarının zamanlaması ve hedeflemesinde kullanılabilir. Örneğin, rezervasyon yapma olasılığı düşük müşterilere yönelik özel indirimler veya teşvikler sunulabilir.

Kapasite Planlaması: Rezervasyon tamamlama olasılıkları, havayolu şirketlerinin uçuş doluluk oranlarını daha iyi tahmin etmesine yardımcı olabilir. Bu bilgi, uçuş kapasitesinin optimize edilmesi, daha fazla talep gören hatlarda ek seferlerin planlanması veya düşük talep gören hatlarda uçuşların yeniden düzenlenmesi gibi operasyonel kararların alınmasına yardımcı olabilir.



Müşteri Deneyimi Yönetimi: Müşteri rezervasyonlarının tamamlama olasılığını anlamak, müşteri memnuniyetini artırmak için önemli bir araçtır. Yüksek tamamlanma olasılığı olan müşterilere yönelik daha kişiselleştirilmiş hizmetler sunulabilir.

Wants Extra Baggage Tahminleri

Sonuçlar:

Logistic Regression ve Random Forest modelleri, ekstra bagaj talebini tahmin etmede en iyi performansı göstermiştir (doğruluk oranı yaklaşık %72, Hamming Loss yaklaşık 0.28). Diğer modellerin performansı, özellikle KMeans ve GMM, oldukça düşüktür.

Öneriler:

Gelir Yönetimi: Ekstra bagaj talebini doğru bir şekilde tahmin etmek, havayolu şirketlerinin bagaj ücretlendirme politikalarını optimize etmelerine yardımcı olabilir. Talebin yüksek olduğu dönemlerde ek bagaj ücretleri artırılabilir veya özel paketler sunulabilir.

Müşteri Memnuniyeti: Ekstra bagaj talep eden müşterilere yönelik özel hizmetler veya indirimler sunarak müşteri memnuniyeti artırılabilir. Örneğin, sık seyahat eden müşterilere veya premium müşterilere ekstra bagaj indirimleri sağlanabilir.

Operasyonel Planlama: Ekstra bagaj talebini öngörmek, uçuş planlamasında ve yer hizmetlerinde etkin rol oynayabilir. Talebin yüksek olduğu uçuşlar için ek bagaj kapasitesi planlanabilir, böylece operasyonel aksaklıklar önlenir.

Wants In-Flight Meals Tahminleri

Sonuçlar:

Logistic Regression ve Random Forest modelleri, uçak içi yemek talebini tahmin etmede en iyi performansı göstermiştir (doğruluk oranı yaklaşık %67, Hamming Loss yaklaşık 0.33). Diğer modellerin performansı, özellikle KMeans ve GMM, daha düşüktür.

Öneriler:

Envanter Yönetimi: Uçak içi yemek talebini doğru tahmin etmek, yemek israfını azaltabilir ve maliyet tasarrufu sağlayabilir. Talebin doğru tahmin edilmesi, uygun miktarda yemeğin hazırlanmasını ve taşınmasını sağlar.

Müşteri Memnuniyeti: Yemek talep eden yolculara yönelik özel menüler veya ek hizmetler sunularak müşteri memnuniyeti artırılabilir. Örneğin, özel diyet veya tercihleri olan yolcular için kişiselleştirilmiş menüler sunulabilir.

Kapsamlı Hizmet Planlaması: Yemek taleplerini tahmin ederek, havayolu şirketleri uçuş sırasında sunacakları hizmetleri optimize edebilirler. Bu, hem maliyetleri düşürür hem de yolcuların daha iyi bir deneyim yaşamalarını sağlar.



Veri Seti	En İyi Model	Accuracy Değeri
Original	LogisticRegression	0.8505
No Outliers	RandomForest	0.8497
Feature Engineered	LogisticRegression	0.8505
No Outliers + FE	RandomForest	0.8497

Her bir veri seti için en iyi model ve bu modelle elde edilen en yüksek accuracy değeri listelenmiştir. "Original" veri setinde ve "Feature Engineered" veri setinde LogisticRegression modelinin en yüksek accuracy değerine ulaştığı görülmektedir, bu durum ilgili veri setlerinin bu model için daha uygun olduğunu göstermektedir. "No Outliers" ve "No Outliers + FE" veri setlerinde ise RandomForest modelinin en yüksek accuracy değerine sahip olduğu gözlemlenmektedir, bu da aykırı değerlerin çıkarılmasının ve özellik mühendisliği işlemlerinin RandomForest modeli için daha faydalı olduğunu işaret etmektedir.

Genel Değerlendirme

Bu çalışmada elde edilen bulgular, havayolu firmaları için önemli stratejik ve operasyonel avantajlar sağlayabilir. Müşteri davranışlarını tahmin eden modellerin doğru ve etkin kullanımı, firmanın gelirlerini artırırken müşteri memnuniyetini de artırabilir. Havayolu firmalarının bu tahminleri nasıl kullanabileceğine dair öneriler, veri odaklı karar verme süreçlerinin geliştirilmesine önemli katkılar sağlayacaktır.

Sonuç olarak, Logistic Regression ve Random Forest modelleri, özellikle müşteri davranışlarını tahmin etmede güvenilir ve etkili araçlar olarak öne çıkmaktadır. Havayolu şirketlerinin bu modelleri benimsemesi, daha stratejik ve verimli operasyonlar yürütmelerine olanak tanıyacaktır. Aykırı değerlerin çıkarılması ve özellik mühendisliği gibi veri ön işleme teknikleri, model performansını optimize etmek için değerlendirilmelidir. Bu yaklaşımlar, havayolu şirketlerinin rekabet avantajını artırmalarına ve müşteri sadakatini güçlendirmelerine yardımcı olacaktır.



Kaynakça

1. Shaw, A. (2023). Airlines Booking Dataset. Kaggle. Erişim Adresi: <https://www.kaggle.com/datasets/anandshaw2001/airlines-booking-csv>
2. Revolution Analytics. (2013). Extending RevoScaleR for Mining Big Data: Discretization. Revolution Analytics Blog. Erişim Adresi: <https://blog.revolutionanalytics.com/2013/04/extending-revoscaler-for-mining-big-data-discretization.html>
3. Güzel, İsmail (2023). YZUP Programı Ders Notları.