

PAPER DETAILS

TITLE: RASTGELE ORMAN ALGORITMALARI ILE OTEL ÖZELLİKLERİ ANALIZI

AUTHORS: Sila SIRIN

PAGES: 123-132

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/1162556>

RASTGELE ORMAN ALGORİTMALARI İLE OTEL ÖZELLİKLERİ ANALİZİ

HOTEL FEATURES ANALYSIS WITH RANDOM FOREST ALGORITHMS

DOI: 10.33461/uybisbbd.756276

Sıla ŞİRİN*

Öz

Günümüzde insanlar otel yelpazesinin çok geniş olması nedeniyle otel seçimlerini, kendi tercihleri doğrultusunda filtreleyerek gerçekleştirmek istemektedirler. Farklı yaş grupları ve çocuklu ailelerin otellerden beklentileri değişmektedir. Örneğin, çocuklu aileler çocuklarının da eğlenerek vakit geçirebilecekleri, denize yakın, kumlu plaja sahip olan otelleri tercih etmektedirler. Daha ileri yaş gruplarında sessiz, spa özellikleri olan oteller tercih edilebilmektedir. İnsanların tercihleri mevsimlere göre de değişiklik gösterebilmektedir. Bu nedenle otel rezervasyonu yapan şirketler için; belirli müşteri gruplarına, bütçeye, yerleşim bölgesine ve mevsimlere göre tercihlerin yorumlanabilmesi önem arz etmektedir. Bu çalışmada müşteri tercihlerinin otel özellikleri bakımından yorumlanabilmesi için, ilk olarak otel özellikleri frekans analizi yöntemi ile azaltılmıştır. Kalan özellikler üzerinde Rastgele Orman Algoritmaları çalıştırılarak yaş gruplarına, mevsimlere ve çocuklu ailelere göre önemli otel özellikleri belirlenmiştir.

Anahtar Kelimeler: Rastgele Orman Algoritmaları, Otel Özellikleri, Otel Öneri Sistemi, Akıllı Seyahat.

Abstract

Today people want to choose hotels by filtering them according to their different features, because the hotel range is very wide. The expectations of different age groups and families with children also vary. For example, today's families with children prefer hotels that are close to the sea, have sandy beach and features where their children can have fun. At older age groups, quiet, spa hotels can be preferred. The choices of people can also change according to the seasons. For this reason, hotel features that stand out according to certain customer groups, budget, residential area and seasons are important for companies that book hotels to form an accurate suggestion system. In this study, firstly, hotel features were reduced by frequency analysis method. The important features were determined for age groups, seasons and families with children by Random Forest Algorithms.

Keywords: Random Forest Algorithms, Hotel Features, Hotel Recommendation System, Smart Travel.

* Türkiye, e-posta: silaasirin@gmail.com, ORCID: 0000-0002-4928-4046

1. GİRİŞ

Otel rezervasyonu yapan şirketler için, müşterilerin tercihlerini tahmin ederek otel rezervasyonu yapılma olasılığını artırmak en büyük amaçtır. Bu şirketlerin maksimum kâr elde edebilmeleri için otelleri doğru zamanda doğru müşteriye önermeleri gerekir (Boz ve ark., 2018). Günümüzde var olan rezervasyon uygulamaları, müşterilerin seçtikleri kriterleri kaydedebilmekte, en çok tercih edilen ve en çok memnun kalınan otellere göre otelleri filtreleyip, sıralayabilmektedir. Otellerin hangi özelliklerine göre ön plana çıktığı, farklı müşteri gruplarının hangi otelleri, hangi özelliklerine göre seçtiği analizinin yapılması gerekmektedir. Rezervasyon şirketlerinin müşterilere doğru otelleri önererek, satış sayılarını artırmaları gerekmektedir. Bu nedenle şirketlerin, otel tercihlerinde öne çıkan özellikleri belirleyebilecekleri ve sürekli yeni verilerle eğitebilecekleri bir modele ihtiyaçları vardır. Bu tür analizler için makine öğrenmesi yöntemlerine sıkça başvurulmaktadır. Bu çalışmada da Rastgele Orman Algoritmalarıyla cinsiyet, çocuklu aile, yaş aralıkları gibi müşteri özelliklerine ve mevsimlere göre otellerin öne çıkan özellikleri belirlenmek istenmiştir.

Çalışmayı yürütürken hangi otel özelliklerinin ön plana çıktığı vurgulanarak otel rezervasyonu yapan şirketler için satış sayılarını artırmak amaçlanmıştır. Elde edilen bilgiler ışığında çalışma farklı şekillerde detaylandırılarak; örneğin otel özelliklerine göre benzer otellerin belirlenip müşteriye sunulması gibi çalışmalar ile sürdürülebilir. Bu alandaki diğer çalışmalar incelendiğinde otel satışlarını artırmak için daha önceki müşteri yorumları üzerinde duygu analizleri yürütüldüğü gözlenmiştir. Bu çalışmada farklı olarak müşteri yorumları kullanılmadan satış ve otel verileri üzerinden satışın gerçekleşmesini sağlayan otel özellikleri belirlenmek istenmiştir. Çalışmayı sunarken ilk olarak veri özellikleri ve kullanılan metot anlatılacaktır, daha sonra ise elde edilen sonuçlar verilecektir.

2. LİTERATÜR TARAMASI

Tatil ayarlarken en önemli adımlardan birisi tatilcinin tercihlerine göre en uygun oteli bulabilmektir. Bu nedenle otel rezervasyonu yapan şirketler için uygun otelleri müşterilere önererek; müşterilerin konaklayacakları otellere bu şirketler üzerinden ulaşabiliyor olması önem arz etmektedir. Bu şirketler için başarılı bir gelir yönetim sisteminin önemli göstergelerinden biri, ne kadar çok rezervasyon ayarlayabildiğidir. Bu çalışmada bazı kriterlere göre otellerin öne çıkan özellikleri belirlendi ancak benzer çalışmalara bakıldığında genel olarak otel öneri sistemleri üzerinde çalışıldığı görüldü. Bizim çalışmamızdan farklı olarak otellerin ön plana çıkan özellikleri yerine, benzer otel grupları oluşturulması, satış tahminleri, maliyet hesaplamaları ve rezervasyon iptallerinin önlenmesi, müşteri yorumları üzerinden duygu analizi gibi konulara yoğunlaşan ve farklı metotların karşılaştırıldığı çalışmalar yapılmıştır. Otel öneri sistemlerinin çoğu otel puanlama ile ilişkilidir ve bu nedenle araştırmaların çoğunda otel sıralaması yapmak amaçlanmıştır (Sayar & Turdaliev, 2018).

Mavalankar ve ark. (2019)'nın ortak çalışmasında Expedia'nın veri seti kullanılmıştır. Çalışmada 100 farklı otel grubundan müşterinin hangisinde kalacağını tahmin etmek amaçlanmıştır. Bunun yanında ek bir çalışma ile her kullanıcının arama sorgusu için en olası ilk beş otelin sıralanması sağlanmıştır. Bizim çalışmamızdan farklı olarak bu çalışmada Stokastik Gradyan İniş sınıflandırıcısı, Rastgele Orman, XG Boost ve Naive Bayes metotları karşılaştırılmıştır. Stokastik Gradyan İniş, dışbükey kayıp fonksiyonları altındaki ayırt edici doğrusal sınıflandırıcılar için basit ama etkili bir yaklaşımdır. Naive Bayes, Bayes teoremini her özellik çifti arasında saf bağımsızlık varsayımıyla uygulamaya dayanan bir dizi denetimli öğrenme algoritmasıdır.

XGBoost, bir tahmin yapmak için zayıf öğrenici topluluklarını kullanmayı ifade eden Extreme Gradient Boosting'in kısaltmasıdır. Bu yöntem de Naive Bayes gibi bir denetimli öğrenme algoritmasıdır. XGBoost ve Rastgele Orman Algoritmaları, model olarak ağaç topluluklarını

kullanmaları bakımından benzerdir (Mavalankar ve ark., 2019). Çalışmanın sonucuna bakacak olursak; en iyi sonuca Rastgele Orman Algoritmaları ile ulaştıkları görülmektedir. Bizim çalışmamızda da Rastgele Orman Algoritmalarının otel verisi benzeri verilerde daha iyi sonuçlar elde edeceği öngörülmüştür; çünkü Rastgele Orman Algoritması aykırı değerlere karşı dirençlidir ve doğrusal olmayan verilerle iyi çalışır, aynı zamanda bizimki gibi büyük veri kümelerinde daha iyi sonuçlar elde edilir. Rastgele Orman Algoritmaları daha detaylı olarak Bölüm 3'te anlatılmıştır.

Aras ve ark. (2019) çalışmasında rezervasyon, otel özellikleri ve online seyahat acentesi verilerini kullanarak satış tahmini yapılması amaçlanmıştır. Yine bu çalışmada da bizden farklı, Mavalankar ve ark. (2019) çalışmasına benzer olarak 4 farklı algoritma sonuçları karşılaştırılmıştır. Yine aynı tür veri için XGBoost, Rastgele Orman Algoritmaları, Gradyan Artırma ve Genelleştirilmiş Doğrusal Model algoritmaları denenmiştir. Gradyan artırma Rastgele Orman Algoritmalarına benzer şekilde zayıf tahmin modellerinin bir araya gelmesiyle tipik olarak karar ağaçlarının oluşturduğu bir model oluşturur ve bunun üzerinde çalışır (Aras ve ark., 2019). Yine bu çalışmada da ağaç yapısı üzerine kurulu modellerin daha doğru sonuçlar verdiği gözlenmiştir. Çalışma sonucunda bizim çalışmamızdan farklı olarak geçmişteki satış verilerinden yararlanarak gelecekteki net toplam maliyet öngörülmüştür; bunun için bizim çalışmaya benzer olarak otel özelliklerinden ve satış verilerinden yararlanmışlardır.

Bir diğer çalışmada Kasper ve Vela (2012) farklı sitelerden müşteri yorumlarını toplayan otel yöneticileri için; kullanıcı yorumlarını analiz edip, sınıflandırmayı planlamışlardır. Müşteri yorumlarını pozitif, negatif, tarafsız ve birden çok konu olacak şekilde sınıflandırmışlardır. Bunun için genel olarak dilbilim analizleri yapılmış ve sınıflandırma algoritmaları kullanılmıştır. Bu çalışmada otel özelliklerinden hiç yararlanılmamış ve oteller için yapılan müşteri yorumlarının analizi üzerine durulmuştur.

3. MATERYAL VE YÖNTEM

Elimizde Setur grubunun kişisel veri olmayacak şekilde maskelenmiş, otel ve müşteri satış verileri bulunmaktadır. Bu bölümde elimizdeki tabloların öznitelik özellikleri ve veri üzerinde yapılan ön işlemler anlatılmıştır.

3.1. Veri Özellikleri ve Ön İşleme

Yapılan çalışmada otel özellikleri, satış ve müşteri bilgilerinin tutulduğu veri setlerinden yararlanılmıştır. TABLO 1, TABLO 2 ve TABLO 3'te verilerin hangi sütunlarda tutuldukları, veri tip ve nitelik bilgileri verilmiştir. (Varchar: farklı uzunluklarda veri girişi yapılacağı zaman kullanılan karakteristik veri tipi, Int: Tam sayı tutan nümerik bir veri tipi, Numeric: Ondalık ve tam sayı türünde veri saklayabilen nümerik bir veri tipi, Datetime: Tarih ve saat verilerini tutan tarih veri tipi (Anonim, 2017).)

TABLO 1 otel özelliklerinin tutulduğu tablo bilgisini içermektedir. Bu tabloda her bir otel için sahip olduğu özellikler verilmiştir. Otel özellikleri veri setinden Otel Numarası ve Özellik Adı nitelikleri kullanılmıştır. Analize başlamadan önce veri seti, her bir otel özelinde özelliklerin tek bir satırda gösterilecek şekilde değiştirilmiştir. Yani; otel özellikleri yeni birer sütun olarak tanımlanarak otel özelinde var/yok(1/0) değerleriyle tek satırda tutulmuştur.

TABLO 1. Otel Özellikleri Veri Seti

Nitelik	Veri Tipi
Otel Özellik Numarası	int
Özellik Numarası	int
Otel Numarası	int
Özellik Adı	varchar
Özellik Kategorisi	varchar

TABLO 2 satış verilerinin tutulduğu tablo bilgilerini içermektedir. Analizler sırasında Müşteri Numarası, Otel Numarası, Otel Satışındaki Toplam Çocuk Sayısı, Otel Satışındaki Toplam Yetişkin Sayısı, Hizmet Başlangıç Tarihi kolonlarından yararlanılmıştır. Satış sayısı verisi her otel özelinde verinin otel numarasına göre gruplanmasıyla elde edilmiştir ve bu bilgi yeni bir Satış Sayısı kolonu yaratılarak bu kolonda tutulmuştur. Bu şekilde otel bazında satış sayılarına ulaşılmıştır.

Ailelerin tercih ettiği otellerin belirlenebilmesi için Otel Satışındaki Toplam Çocuk Sayısı sütunu kullanılmıştır. Otel Satışındaki Toplam Çocuk Sayısı sütunu dolu olan satışlar çocuklu aile olarak sınıflandırılmıştır.

TABLO 2. Satış Veri Seti

Nitelik	Veri Tipi
Satış Numarası	int
Satış Tipi Numarası	int
Müşteri Numarası	int
Satış Yapan Firma Numarası	int
Satılan Otel Numarası	int
Satılan Cruise Numarası	int
Satılan Ekstra Servis Numarası	int
Satılan Tur Numarası	int
Kişi Adedi	int
Otel Satışındaki Toplam Yetişkin Sayısı	int
Otel Satışındaki Toplam Çocuk Sayısı	int
Tur Satışındaki Toplam Yetişkin Sayısı	int
Tur Satışındaki Toplam Çocuk Sayısı	int
Cruise Satışındaki Toplam Yetişkin Sayısı	int
Cruise Satışındaki Toplam Çocuk Sayısı	int
Verilen Hizmetin Toplamda Kaç Gece Sürdüğü Bilgisi	int
Verilen Hizmetteki Toplam Oda Sayısı	int
Satış Tarihi	datetime
Hizmet Başlangıç Tarihi	datetime
Hizmet Bitiş Tarihi	datetime
Kar	numeric
Satış Tutarı	numeric

TABLO 3 müşteri verilerinin tutulduğu tablodur. Müşteri Numarası ve Müşterinin Yaşı kolonları müşterilerin yaş aralığına göre otel özellikleri tercihlerinin belirlenmesi için kullanılmıştır.

TABLO 3. Müşteri Veri Seti

Nitelik	Veri Tipi
Müşteri Numarası	int
Müşteri Cinsiyet Bilgisi	varchar
Müşterinin Yaşı	int
Ülke Bilgisi	int
Şehir Bilgisi	int

Frekans analizi yöntemiyle frekansı 0.00050'ten düşük olan otel özellikleri çıkarılarak analiz yapılacak veri küçültülmüştür. Frekans analizi, otel özelliklerinin sıklığı incelenerek yürütülmüştür. Frekans analizinden önce veri satır sayısı 39559 iken frekans analizi ile 2263'e; aynı şekilde özellik sayısı da 207'den 122'ye düşürülmüştür.

Satış verileri içerisindeki mükerrer kayıtlar çıkarılmıştır. Mükerrer kayıtlar sonuçlarda sapmalara neden olacağı için veri ön işleme için önemli bir adımdır. Mükerrer kayıtların çıkarılması ile veri satır sayısı 300571'den 284329'a düşmüştür.

3.2. Rastgele Orman Algoritmaları

Rastgele Orman Algoritması, amaca uygun olarak; belirlenen ağaç sayısından meydana gelen sınıflandırma veya regresyon ağaç topluluklarından oluşmaktadır. Rastgele orman, her girdi vektöründen bağımsız olarak örneklenen rastgele bir vektör kullanılarak oluşturulan ağaçların kombinasyonundan oluşur. Bu çalışma için kullanılan rastgele orman, bir ağacı büyütmek için rastgele seçilen özelliklerin birleşiminden oluşturulmuştur. Bir sınıflandırıcı olarak, rastgele orman sınıflandırma için bir "güçlü değişkenler" alt kümesi kullanarak örtük bir özellik seçimi gerçekleştirir ve bu durum yüksek boyutlu veriler üzerinde üstün performansa yol açar (Breiman, 2001).

Rastgele orman algoritmaları hızlı, esnek ve yüksek boyutlu verileri analiz etmek için yeterince güçlüdür. Binlerce değişkeni silmeden veya doğrulukta bozulma olmadan işleyebilecek olağandışı yeteneğe sahip bir algoritmadır (Breiman, 2004). Algoritmanın, sınıflandırma ağaçlarına göre eksikliği, çıktı olarak bir ağaç vermemesidir (Korkmaz ve ark., 2018). Alternatif makine öğrenme algoritmalarına göre bir avantajı, ilgili özellikleri tanımlamak veya değişken seçimi yapmak için kullanılabilen değişken önem ölçütleridir. Gini endeksi gibi bölünmelerdeki kirliliğin azaltılmasına dayalı ölçütler yaygın olarak kullanılır, çünkü bu ölçütler basit ve hızlı sonuçlar verir. Gini endeksi, sınıflandırma ağaçlarında bölme kriteri olarak yaygın bir şekilde kullanıldığından, karşılık gelen safsızlık önemine genellikle Gini önemi denir.

Karar ağacının tasarımı için öznelik seçim ölçüsü ve budama yönteminin seçilmesi gerekir. Bu çalışmada Sklearn kütüphanesinin tanımladığı Rastgele Orman Algoritması kullanılmıştır. Sklearn kütüphanesi Rastgele Orman Algoritması için öznelik önemi(feature_importances_) özelliğini sunmaktadır. Öznelik önemi bir öznelik seçim ölçütüdür ve değerinin büyük olması öznelik öneminin de fazla olması demektir. Öznelik önemi, tahmin yaparken; her bir özelliğin göreceli önemini belirten bir tahmin modeli oluşturmak için, her bir özelliğe belli bir skor belirler. Bu skorlar sayısal bir değeri tahmin etmeyi amaçlayan regresyon veya sınıflandırma problemleri için hesaplanabilir. Rastgele Orman Algoritmasında öznelik önemi, ağacın bölünme noktalarını seçmek için kullanılan kriterdeki (Gini endeksi) indirgeme duruma göre bu skorları belirler. Yani bu

özelliği kullanan ağaç düğümlerinin, ormandaki tüm ağaçlar arasındaki kirliliği ne kadar azalttığına bakarak, bir özelliği ön planda tutan bir araç sağlar ve her bir özellik için bu skoru otomatik olarak eğitir ve sonuçları ölçeklendirir, böylece tüm önemlerin toplamı 1'e eşit olur (Pedregosa, 2011). Rastgele Orman Algoritması öznitelik seçimi için Gini endeksi kullanmaktadır. Gini endeksi ağaçtaki şekilde hesaplanır. (Yılmaz, 2014)

$$I_G = 1 - \sum_{j=1}^c p_j^2 \quad (1)$$

T : Tüm veri seti p_j : Veri setindeki her bir verinin, kendisinden küçük ve kendisinden büyük eleman sayılarına bölüm karesi. c : Seçilen veri

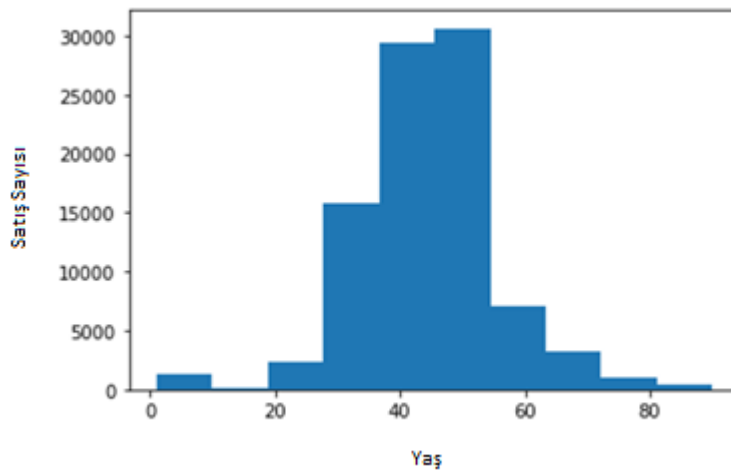
Bir karar ağacı, özelliklerin kombinasyonunu kullanarak yeni eğitim verileri üzerinde maksimum derinliğe kadar büyür. Bu çalışmada maksimum derinlik belirlenmemiştir, bu nedenle ağaçlar üzerinde budama yapılmamıştır. Daha doğru sonuçlar elde edebilmek için oluşturulacak ağaç sayısı 1000 olarak belirlenmiştir.

4. BULGULAR

Veri seti üzerinde istenilen amaca uygun olarak Rastgele Orman Algoritması kullanılmıştır. Öne çıkan özellikler; yaş aralığı, çocuklu aile ve mevsimlere göre ayrı ayrı belirlenmiştir.

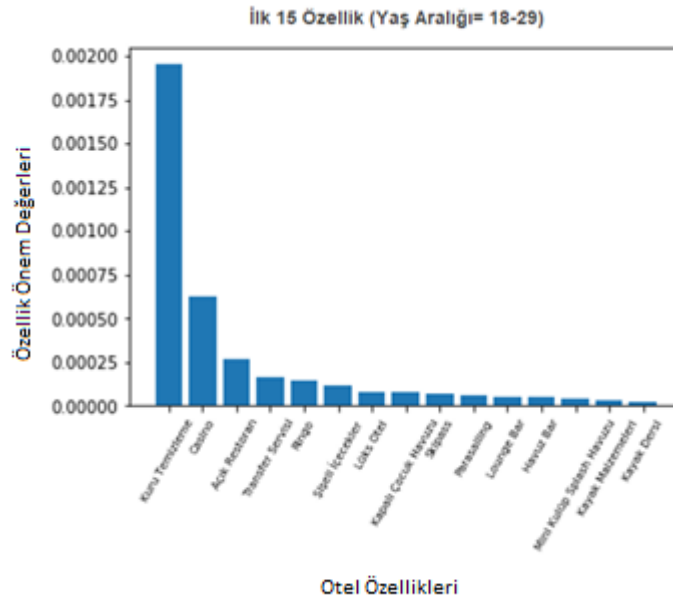
4.1. Yaş Aralığına Göre Öne Çıkan Otel Özellikleri

Bölüm 3'te bahsedildiği gibi müşterilerin verilerinin tutulduğu Müşteri veri seti ve otel satışlarının yer aldığı Satış veri setindeki verilerden yararlanılmıştır. Satış veri setindeki otel satışı olan verilerin müşteri numaraları ile müşteri veri setindeki müşteri numaraları kullanılarak müşterilerin yaş bilgilerine ulaşılmıştır. Bu verilerden yararlanılarak öncelikle Şekil 1'de görüldüğü gibi satış verisinin yaşlara göre dağılımı gösterilmiştir.

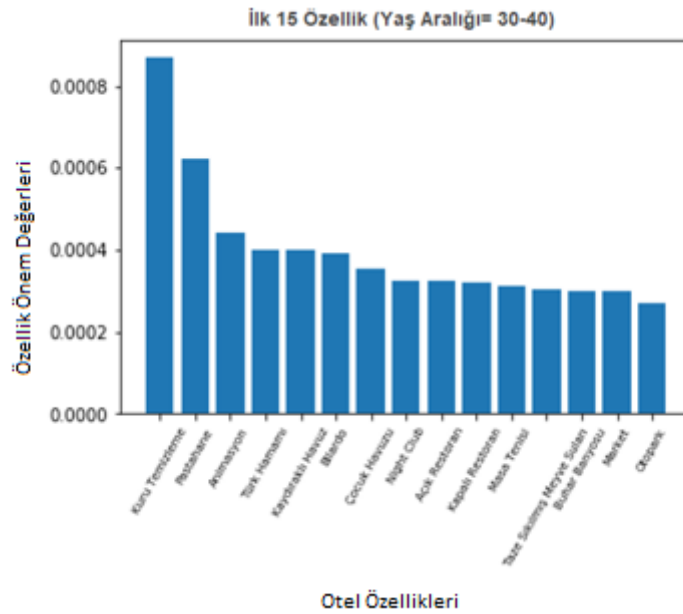


Şekil. 1. Satış verisinin yaşlara göre dağılımı

Şekil 1'deki dağılıma bakılarak müşteri yaş bilgilerinden 18-29, 30-40, 41-55, 56-74, 75+ yaş aralıkları belirlenmiştir. Her bir yaş aralığı sınıf olarak belirlenerek Rastgele Orman Algoritması çalıştırılmıştır. Özelliklerin önemini belirlemek için Bölüm 3'te bahsedilen Rastgele Orman Algoritmasının öznitelik öneminden yararlanılmıştır, daha sonra her bir sınıf için özellikler ağırlıklandırılarak her sınıf için öne çıkan özellikler belirlenmiştir. Şekil 2 ve Şekil 3 üzerinde örnek olarak 18-29 ve 30-40 yaş aralıklarına göre belirlenen 15 en önemli özellik verilmiştir. İki yaş grubu için öne çıkan otel özelliklerinin değiştiği grafiklerde gözlemlenmiştir.



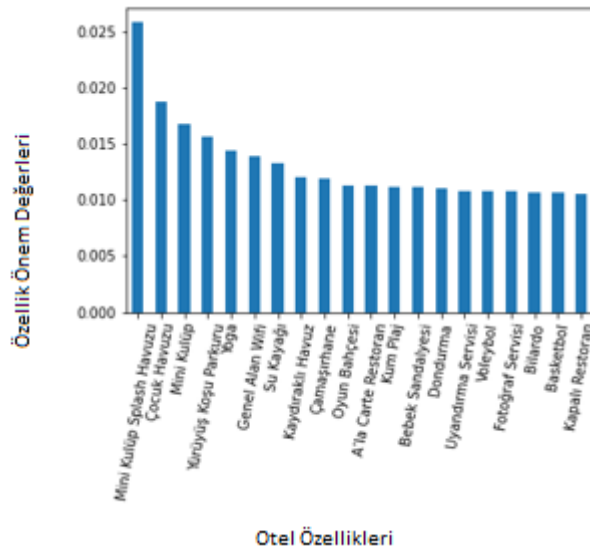
Şekil. 2. 18-29 yaş aralığı için öne çıkan 15 özellik



Şekil. 3. 30-40 yaş aralığı için öne çıkan 15 özellik

4.2. Çocuklu Ailelere Göre Öne Çıkan Otel Özellikleri

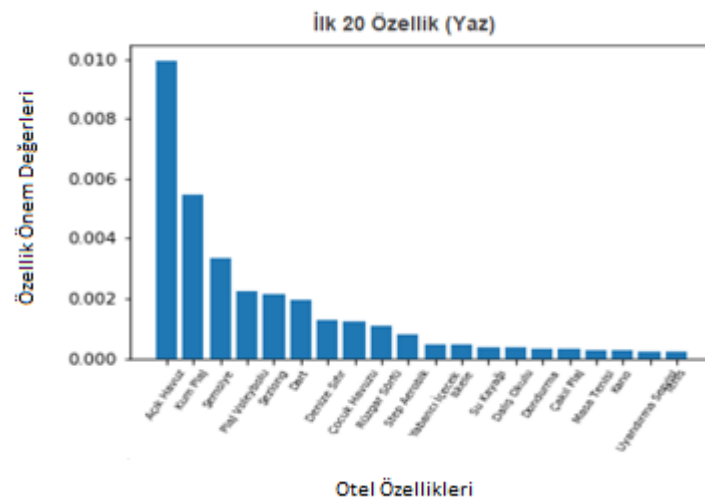
Satış verisi üzerinden Otel Satışındaki Toplam Çocuk Sayısı alanı kullanılarak çocuklu rezervasyonlar ve çocuksuz rezervasyonlar olarak 2 ayrı sınıf belirlenmiştir. Bu sınıflar için Rastgele Orman Algoritması çalıştırılarak yine her bir sınıf için öznelik önemi sonuçları sınıf üzerinde ağırlıklandırılarak çocuklu ailelerin otel seçimlerinde hangi özellikleri tercih ettikleri çıkarılmıştır. Şekil 4 üzerinde çocuklu aileler için öne çıkan 20 özellik gösterilmiştir.



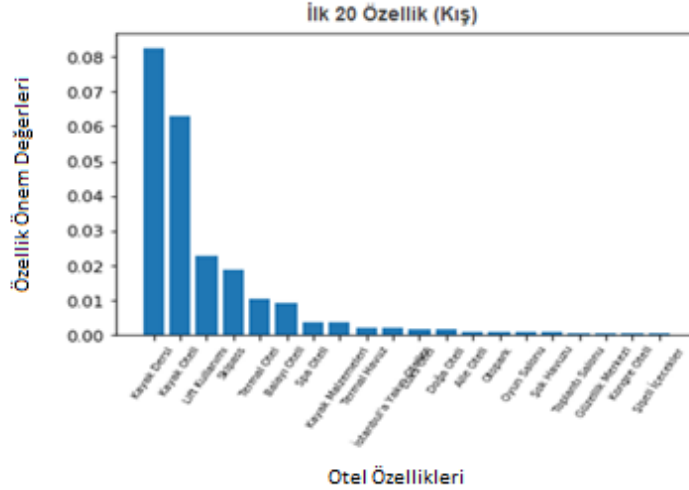
Şekil. 4. Çocuklu ailelere göre öne çıkan 20 özellik

4.3. Mevsimlere Göre Öne Çıkan Otel Özellikleri

Satış verisi üzerinden Hizmet Başlangıç Tarihi alanı kullanılarak satışların hangi tarihler için yapıldığı bilgisi elde edilmiştir. Nisan, Mayıs, Haziran, Temmuz, Ağustos, Eylül ayları yaz ayları olarak; Ekim, Kasım, Aralık, Ocak, Şubat, Mart ayları ise kış ayları olarak belirlenmiştir. Kış ve yaz mevsimleri için iki ayrı sınıf oluşturulmuştur. Satış verisi üzerinde kış ve yaz sınıfları için Rastgele Orman Algoritması çalıştırılarak, her iki sınıf için de öne çıkan özellikler belirlenmiştir. Bunun için yine Rastgele Orman Algoritmasının öznelilik öneminden yararlanılmıştır. Şekil 5 ve 6'da hem yaz hem de kış sınıfları için öne çıkan özellikler ayrı ayrı gösterilmiştir.



Şekil. 5. Yaz ayları için öne çıkan 20 özellik



Şekil. 6. Kış ayları için öne çıkan 20 özellik

5. TARTIŞMA

Otel rezervasyonu yapan şirketler için farklı müşteri gruplarına ve farklı zaman dilimlerine göre hangi otel özelliklerinin rezervasyonun tamamlanması için belirleyici olduğu üzerine çalışma yürütmek önem arz etmektedir. Bu çalışmada elde edilen bulgular üzerinden farklı analizler yürütülerek otel rezervasyonları artırılabilir.

Bu çalışmada veri, üç ana başlık için ayrı ayrı incelenmiştir. Yaş grupları, mevsim ve çocuklu aileler ayrı başlıklar olarak ele alınıp; her bir başlık için öne çıkan otel özellikleri Bölüm 4'te gösterilmiştir. Sonuçlar incelendiğinde; başlıklara göre belirleyici özelliklerin genel olarak makul olduğu görülmektedir. Örneğin; mevsimlere göre öne çıkan özellikler incelendiğinde kış ayları için kayak, termal, spa gibi kış sporları veya iç mekan olanaklarının, yaz ayları içinse havuz, kum plaj, plaj voleybolu gibi dış mekan olanaklarının otel rezervasyonu için daha belirleyici olduğu gözlenmiştir; ancak bunun yanında bazı durumlarda belirleyici olmaması beklenen özelliklerin de ön sıralarda listelendiği görülmüştür. Örneğin, yaş grupları için yapılan çalışma incelendiğinde “Kuru Temizleme” özelliğinin farklı yaş grupları için bile en ön sırada geldiği görülmüştür; ancak “Kuru Temizleme” özelliğinin otel seçiminde en önemli kriter olması genellikle beklenmez. Bu nedenle bu tür sonuçların incelenmesi ve daha net sonuçlar elde edilmesi için çalışma farklı makine öğrenmesi yöntemleri kullanılarak ilerleyen zamanlarda devam ettirilebilir.

Literatürde yapılan çalışmalardan farklı olarak; bu çalışmada otel özelliklerinin farklı müşteri grupları ve farklı mevsimlere göre sıralanması sağlanmıştır. Literatürdeki diğer çalışmalar incelendiğinde, Mavalankar ve ark. (2019) otelleri gruplandırarak müşterinin hangi otelde kalacağını tahmin etmeye çalışırken; Aras ve ark. (2019)'da satış tahmini üzerine yoğunlaşmaktadır. Her iki çalışmada da farklı algoritmalar aynı veri seti üzerinde çalıştırılıp sonuçlar kıyaslanır. Bu çalışmada farklı olarak tek bir algoritma üzerinden öne çıkan otel özelliklerinin belirlenmesi sağlanmıştır. Literatürdeki diğer çalışmalar incelendiğinde benzer konuya eğilen çalışmalar bulunmamaktadır, bu nedenle bu çalışmanın sonuçları diğer çalışmalardan ayrılmaktadır.

6. SONUÇ

Konaklama hizmeti veren otellerin maksimum kâr elde edebilmesi için doluluk oranlarının yüksek olması gerekmektedir. Bu sebeple oteller rezervasyon sistemleri aracılığıyla sınırlı sayıdaki odalarını doğru zamanda, doğru müşteriye tahsis etmelidir. Bu araştırma makine öğrenmesi yöntemlerini kullanarak belirli hedef kitleleri ve belirli bölgeler için hangi otel özelliklerinin ön

plana çıktığını ve bu elde edilen sonuçlar ile daha iyi otel önerileri geliştirebilmek için otel rezervasyonu yapan şirketler ve aynı zamanda müşteriler tarafından kullanılabilecek daha başarılı öneri sistemleri geliştirmek hedeflenmiştir. Bu sayede otel rezervasyonu yapabilen şirketler tur şirketleri vb. her otel için sahip oldukları özelliklerden en çok tercih edilenleri ön plana çıkarıp satışlarını artırabilirler aynı zamanda müşteriler için de ihtiyaçlarına göre otel seçebilmeleri daha kolaylaştırılabilir. Diğer bir yandan şirketler için böyle bir bilgi, yönetsel olarak da stratejilerini geliştirmelerine ve yönetim kalitesini artırmalarına imkân sağlar. Bu sayede şirketlerin elinde bulunan otel özellikleri, müşteri bilgileri, satış verileri arttıkça geliştirilen model ile çok daha kesin ve doğru analizlere ulaşılabilir. Farklı makine öğrenmesi yöntemleriyle bulunan sonuçlar test edilebilir.

KAYNAKÇA

- Anonim, (2017). “Data types (Transact-SQL)”. <https://docs.microsoft.com/en-us/sql/t-sql/data-types/data-types-transact-sql?view=sql-server-2017>, Erişim T: 13.12.2020.
- Aras G., Ayhan G., Sarıkaya M., Tokuç A., Sakar C., (2019). “Forecasting Hotel Room Sales within Online Travel Agencies by Combining Multiple Feature Sets”. 8th International Conference on Pattern Recognition Applications and Methods, 565-573.
- Boz M., Canbazoğlu E., Özen Z., Gülseçen S., (2018). “Otel Rezervasyon İptallerinin Makine Öğrenmesi Yöntemleri ile Tahmin Edilmesi”. Veri Bilimi. 1(1): 7-14.
- Breiman L., (2004). “Consistency for a simple model of random forests”. Technical Report 670, Technical report, Department of Statistics, University of California, Berkeley, USA.
- Breiman L., (2001). “Random Forests”. Machine Learning.
- Kasper W., Vela M., (2012). “Sentiment Analysis for Hotel Reviews”. Speech Technology, 4. 96-109.
- Korkmaz D., Çelik E. H., Kapar M., (2018). “Sınıflandırma ve Regresyon Ağaçları ile Rastgele Orman Algoritması Kullanarak Botnet Tespiti”. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 23 (3): 297-307.
- Mavalankar, A., Gupta, A., Gandotra, C., Misra, R., (2019). “Hotel Recommendation System”. 10.13140/RG.2.2.27394.22728/1.
- Pedregosa F., (2011). “Scikit-learn: Machine Learning in Python”, JMLR 12, pp. 2825-2830.
- Sayar A., Turdaliev N., (2018). “Makine Öğrenmesi ile Adaptif Otel Öneri Sistemi”. 12th Turkish National Software Engineering Symposium, Istanbul, Türkiye.
- Yılmaz H., (2014). “Random Forests Yönteminde Kayıp Veri Probleminin İncelenmesi ve Sağlık Alanında Bir Uygulama”. (Yüksek Lisans Tezi). Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı, Eskişehir, Türkiye.