

# İNTERNETTEN ALIŞVERİŞ YAPAN HANELERİN RASTGELE ORMAN YÖNTEMİYLE TAHMİN EDİLMESİ<sup>1</sup>



Kafkas Üniversitesi  
İktisadi ve İdari Bilimler  
Fakültesi  
KAÜİBFD  
Cilt, 12, Sayı 24, 2021  
ISSN: 1309 – 4289  
E – ISSN: 2149-9136

Makale Gönderim Tarihi: 22.07.2021

Yayına Kabul Tarihi: 23.11.2021

Uğur ERCAN  
Öğr. Gör. Dr.  
Akdeniz Üniversitesi  
Enformatik Bölüm Başkanlığı,  
Antalya, Türkiye  
ugurercan@akdeniz.edu.tr  
ORCID ID: 0000-0002-9977-2718

**ÖZ** | Gerçekleştirilen çalışmanın amacı, Türkiye’de internetten alışveriş yapan hanelerin tespit edilmesidir. Modelleme aşamasında sınıflandırma problemlerinde sıklıkla tercih edilen Rastgele Orman yöntemi kullanılmıştır. TÜİK 2019 Hanehalkı Bütçe Anketinde yer alan ve 11521 haneden derlenen veri seti kullanılmıştır. Çalışmanın veri seti SMOTE ve Rastgele Yetersiz Örnekleme yöntemleri ile dengeli hale getirilmiştir. Çalışmanın doğruluğunu artırmak için çapraz doğrulama yöntemi kullanılmıştır. Kurulan modellerin performansları karşılaştırılarak yorumlanmış, örnekleme yöntemleri ve çapraz doğrulamanın doğru bir şekilde kullanılması ile sınıflandırıcı performansının artırılabilceği gösterilmiştir. Eğitim veri setinde, SMOTE yöntemi uygulanarak kurulan model, diğer modellere göre tüm ölçütlerin (F, DP, G-Means ve MCC ) sonuçlarına göre daha başarılı olduğu görülmüştür. Test veri setinde ise SMOTE yöntemi uygulanan modelin, F ve MCC ölçütlerinin sonucuna göre daha başarılı olduğu görülürken, Yetersiz Örnekleme yöntemi ile kurulan modelin G-Means ölçütü sonucuna göre, hiçbir yöntem kullanılmadan oluşturulan modelin ise DP ölçütü sonucuna göre başarılı olduğu görülmüştür.

**Anahtar Kelimeler:** Rastgele orman, internetten alışveriş, dengesiz veri seti, SMOTE, rastgele yetersiz örnekleme

**JEL Kodu:** C50, D30, D60

**Alan:** İşletme

**Türü:** Araştırma

**DOI:** 10.36543/kauibfd.2021.030

**Atıfta bulunmak için:** Ercan, U. (2021). İnternette alışveriş yapan hanelerin rastgele orman yöntemiyle tahmin edilmesi. *KAÜİBFD*, 12(24), 728-752.

<sup>1</sup> İlgili çalışmanın etik kurallara uygunluğu beyan edilmiştir.

# ESTIMATING OF HOUSEHOLDS SHOPPING ON THE INTERNET USING RANDOM FOREST METHOD



Kafkas University  
Economics and Administrative  
Sciences Faculty  
KAUEASF  
Vol. 12, Issue 24, 2021  
ISSN: 1309 – 4289  
E – ISSN: 2149-9136

Article Submission Date: 22.07.2021

Accepted Date: 23.11.2021

Uğur ERCAN  
Lecturer Dr.  
Akdeniz University  
Department of Informatics,  
Antalya, Turkey  
ugurercan@akdeniz.edu.tr  
**ORCID ID: 0000-0002-9977-2718**

## ABSTRACT

The aim of the study is to determine the households shopping online in Turkey. During the modeling phase, the Random Forest method, which is frequently preferred in classification problems, was used. The data set in the TÜİK 2019 Household Budget Survey and gathered from 11521 households was used. The data set of the study was balanced with SMOTE and Random Undersampling methods. The cross-validation method was used to increase the accuracy of the study. The performances of the established models were compared and interpreted, and it was shown that the classifier performance could be increased with the correct use of sampling methods and cross-validation. In the training dataset, the model established by applying the SMOTE method was found to be more successful than the results of all criteria (F, DP, G-Means and MCC ) compared to other models. In the test data set, while it was observed that the model with the SMOTE method was more successful than the results of the F and MCC criteria, the model established with the Undersampling method was more successful according to the result of the G-Means criterion, and the model created without using any method was found to be successful according to the result of the DP criterion.

**Keywords:** *Random forest, shop on the internet, imbalanced dataset, SMOTE, random undersampling*

**Jel codes:** *D10, R20, D80*

**Scope:** *Business Administration*

**Type:** *Research*

## 1. GİRİŞ

Online ya da internet üzerinden alışveriş, geleneksel pazarlama yöntemlerine alternatif ve doğrudan bir ticaret aracı olarak hızla büyüyen bir alandır. Bilgi ve iletişim teknolojilerinin giderek yaygınlaşması, web sitelerinin çoğalması (Armağan ve Turan, 2014), internetin sadece bilginin paylaşıldığı bir yer değil aynı zamanda ürün ve hizmetlerin ticaretinin yapılmasında sıkça kullanılan bir araç haline gelmesi ile birlikte hızla yaygınlaşmış ve işlem hacmi gün geçtikçe artış göstermiştir. Böylelikle internet ortamının, işletmelerin yok sayamayacağı kadar büyük bir pazar yeri, alternatif bir dağıtım kanalı, rekabet ve verimlilik unsuru olarak görülmesi kaçınılmazdır (Doğrul, 2012). Öyle ki internetten alışveriş, mağazalardan veya alışveriş merkezlerinden satın almayı içeren geleneksel satın alma yöntemlerinin yerini almaktadır (Sreerekha, Saranya, & Prabhu, 2019). Tüketiciler, internet üzerinden alışveriş yaparken ürünlere ait video, resim, bilgi ve özelliklerine rahatça erişebilmektedir. İnternet üzerinden alışveriş yapmak pratiktir, zaman tasarrufu sağlar, fiyat ve ürün karşılaştırmaya imkân tanır, evden dışarı çıkmadan ve sıra beklemeden alışveriş imkânı sağlamaktadır (Çiçek ve Mürütsoy, 2014), işletmelere düşük maliyet sunar (Al-Maghrabi vd., 2011), ürün çeşitliği ve detaylı ürün bilgisi sağlar. Bunun yanı sıra bireyler zaman ve mekândan bağımsız olarak alışverişlerini yapabilmektedir (Cengiz ve Şekerya, 2010). Son on yılda internet üzerinden alışveriş, geleneksel alışverişe göre daha ekonomik ve daha uygun bir satın alma yaklaşımı temsil etmesi nedeniyle hızlı bir büyüme yaşamıştır (Vasić, Kilibarda, & Kaurin, 2019). Bhatti ve Akram (2020) bireylerin ABD’de %68’nin, Malezya’da %68’nin internet üzerinden alışveriş yaptığını belirtmiştir. EUROSTAT verilerine göre Avrupa Birliğine üye 28 ülkede 2012 yılında bireylerin Birleşik Krallık’da %64’ü, Norveç’te %62’si, Danimarka’da %60’ının internetten alışveriş yaptığı görülürken, en düşük oranın %6 ile Bulgaristan ve %3 ile Romanya’da gerçekleştiği görülmüştür. 2019 yılında ise bireylerin alışveriş oranları İngiltere’de %80, İsviçre’de %75, Danimarka’da %74 olarak görülmüştür. En düşük oranın yine %15 ile Romanya ve %14 ile Bulgaristan’da gerçekleştiği görülmüştür (EUROSTAT, 2021). Türkiye’de hanelerin internet erişim oranları 2011 yılında %42.9, 2015 yılında %69.5, 2020 yılında %90.7, 2021 yılında ise %92 olarak ölçülmüştür. Bilgisayar kullanım oranlarının, 2011 yılında %46.4, 2015 yılında %54.8, 2018 yılında ise %59.6 olduğu görülmüştür. İnternet kullanan bireylerin, internet üzerinden mal veya hizmet siparişi verme ya da satın alma oranı 2011 yılında %18.6, 2015 yılında %33.1, 2020 yılında (2019, Nisan-2020, Mart) ise %36.5’dir (TÜİK, 2021a; TÜİK, 2021b; TÜİK, 2021c). Bu bilgiler doğrultusunda, Türkiye’de bilgi ve iletişim teknolojilerinin

kullanımının artması ile beraber internet üzerinden yapılan alışveriş oranlarının da arttığı görülmektedir.

Dünya üzerinde her an milyonlarca insan bilgisayar, telefon ve tablet aracılığıyla internete bağlanmakta, web sitelerinde ve mobil uygulamalarda çeşitli amaçlar için gezinmektedir. Bu insanlar aslında bütün firmalar için potansiyel bir müşteridir (Sreerekha vd., 2019). Gültaş ve Yıldırım (2016) psikolojik ve demografik özellikler gibi tüketicilerin karar verme süreçlerinde etkili olan faktörleri anlamının, internetteki potansiyel tüketicileri kazanmanın başarı faktörü olarak belirtmiştir. Bunun yanı sıra birçok insan günlük yaşantısında youtube, facebook, instagram, twitter gibi çeşitli sosyal medya platformlarını kullanmaktadır. Bu platformlar gerçekleştirdikleri hizmetler esnasında doğrudan ya da dolaylı olarak reklam vererek bireylerin dikkatini çekmekte, insanların satın alma tercihlerine etki edebilmekte ve yönlendirebilmektedir (Jothi ve Gaffoor, 2017; Torun, 2017).

Bireylerin internet üzerinden satın alma davranışını etkileyen birçok değişken bulunmaktadır. Bunlar; psikolojik, sosyokültürel ve sosyoekonomik (Akhter, 2003; Akhter, 2012; Oke ve Ramachandran, 2021; Soler, Christidis, & Vassallo, 2021), demografik (Akhter, 2003; Goldsmith ve Flynn, 2005; Richards ve Shen, 2006; Özgüven, 2011; Özhan ve Altuğ, 2015), sosyal medya (Jothi ve Gaffoor, 2017; Zhang vd., 2017; Ertürk ve Aktepe, 2020) değişkenleri olarak sınıflandırılmaktadır. İnternette alışveriş davranışlarının makine öğrenmesi yöntemleri ile tahmini için gerçekleştirilen çalışmalar incelendiğinde; Maheswari ve Priya (2017) müşterilerin çevrimiçi satın alma davranışlarını Destek Vektör Makineleri ile tahmin etmiştir. Kumar vd. (2019) müşterilerin internette tekrar satın alma niyetlerini makine öğrenmesi ve yapay arı kolonisi algoritmasının kombinasyonundan oluşan hibrit bir yöntemle tahmin etmiştir. AdaBoost yönteminin C5.0, Yapay Sinir Ağları, Destek Vektör Makineleri ve Rastgele Orman yöntemlerine göre sınıflandırma performansının daha iyi olduğu görülmüştür. Koehn, Lessmann, & Schaal (2020) müşterilerin internette alışveriş davranışını Yinelemeli Sinir Ağları yöntemi ile incelemiştir. Yinelemeli Sinir Ağları tabanlı sınıflandırıcıların diğer yöntemlere göre daha başarılı olduğu görülmüştür. Rausch, Derra, & Wolf (2020) internette alışveriş yapan bireylerden alışveriş sepetini terk edenleri makine öğrenmesi yöntemleri ile belirlemiştir. Gradient Boost (with regularization) yöntemi ile kurulan modelin en iyi doğruluk değeri elde ettiği görülmüştür. Barua, Zou, & Liu (2021) hanehalkının internette alışveriş talebini makine öğrenmesi yöntemlerinden Gradient Boost yöntemi ile tahmin etmiştir.

Hane, fert ve tüketim karakteristiklerine göre internette alışveriş yapan hanelerin tahmin edilmesi, bu alanda faaliyet gösteren üretim, dağıtım,

pazarlama, satış ve tedarik yapan bütün işletmeler açısından doğrudan ya da dolaylı olarak büyük önem arz etmektedir. Çalışmanın amacı, hanelere ait karakteristikler kullanılarak internetten alışveriş yapan hanelerin makine öğrenmesi yöntemleri ile tahmin edilmesidir. Makine öğrenmesi yöntemlerinden, sınıflandırma problemlerinde yüksek doğruluk başarısı gösteren Rastgele Orman yöntemi kullanılmıştır.

## 2. VERİ SETİ ve YÖNTEM

### 2.1. Veri Seti

Çalışmada, Türkiye İstatistik Kurumu tarafından derlenen, 11521 hanehalkına ait 2019 yılı Hanehalkı Bütçe Anketi (HBA) ait verileri kullanılmıştır (TÜİK, 2019). İnternette alışveriş yapmayı etkileyeceği düşünülen hanehalkı karakteristikleri öznelik seçimi (feature selection) yöntemi belirlenmiştir. RO modelinde kullanılan karakteristikler ve bu karakteristiklere ait tanımlayıcı istatistikler Tablo 1’de belirtilmiştir. Tablo 1 aynı zamanda, internetten sipariş veren hanelere ilişkin tanımlayıcı istatistikleri de göstermektedir. Veri setindeki tüm gelir ve harcama verileri 2019-Aralık ayına göre enflasyondan arındırılmıştır.

**Tablo 1.** Çalışmada Kullanılan Değişkenler ve İnternette Alışveriş Yapan Hanelere Ait Tanımlayıcı İstatistikler

Değişkenler	Kategoriler	Çalışmada Kullanılan Değişkenler ve Tanımlayıcı İstatistikler		İnternette Alışveriş Yapan Hanelere Ait Tanımlayıcı İstatistikler	
		Frekans	Yüzde	Frekans	Yüzde
Kredi kartı sahiplik durumu	Yok	5546	48.14	156	9.73
	Var	5975	51.86	1448	90.27
Hanehalkı reisinin medeni durumu	Bekâr	2446	21.23	371	23.13
	Evli	9075	78.77	1233	76.87
Hanehalkı reisinin sigorta durumu	Sigortasız	489	4.24	40	2.49
	Sigortalı	11032	95.76	1564	97.51
Hanehalkı reisinin çalışma durumu	Çalışmıyor	4605	39.97	353	22.01
	Çalışıyor	6916	60.03	1251	77.99
Hanehalkı reisinin cinsiyeti	Kadın	2680	23.26	350	21.82
	Erkek	8841	76.74	1254	78.18
Hanehalkı reisinin yaşı	15-29 yaş arası	705	6.12	275	17.14
	30-44 yaş arası	3398	29.49	700	43.64
	45-59 yaş arası	3837	33.30	487	30.36
	60 yaş ve üzeri	3581	31.08	142	8.85
Hanehalkı reisinin eğitim düzeyi	Eğitimsiz	1521	13.20	30	1.87
	İlköğretim	6111	53.04	363	22.63
	Lise	1877	16.29	346	21.57
	Üniversite ve üzeri	2012	17.46	865	53.93

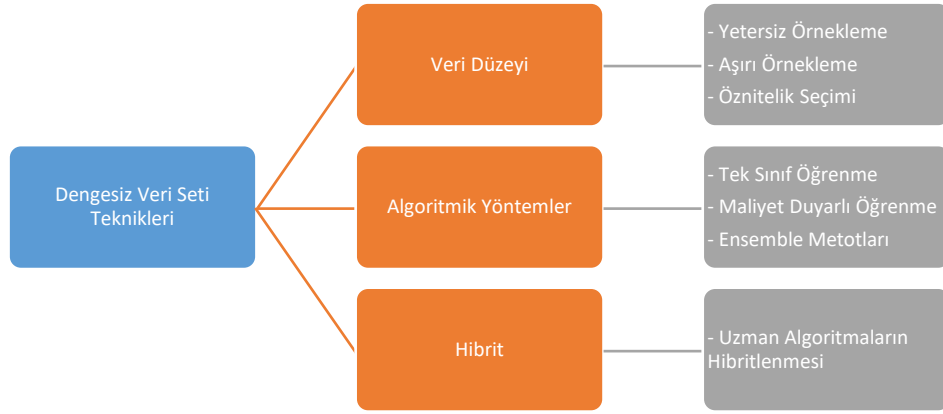
Değişkenler	Kategoriler	Çalışmada Kullanılan Değişkenler ve Tanımlayıcı İstatistikler		İnternette Alışveriş Yapan Hanelere Ait Tanımlayıcı İstatistikler	
		Frekans	Yüzde	Frekans	Yüzde
Hanehalkı kompozisyonu	Tek yaşayan	1224	10.62	185	11.53
	Aile	8634	74.94	1237	77.12
	Geniş aile	1422	12.34	100	6.23
	Bir arada yaşayan kişiler	241	2.09	82	5.11
	Ev sahibi	7093	61.57	763	47.57
Mülkiyet durumu	Kiracı	2578	22.37	596	37.16
	Lojman	206	1.79	66	4.11
	Diğer (ev sahibi değil ama kira ödemiyor)	1644	14.27	179	11.16
Hanede internet bağlantı durumu	Yok	3566	30.95	83	5.17
	Var	7955	69.05	1521	94.83
Hanede şans oyunu alışkanlığı	Yok	11029	95.73	1477	92.08
	Var	492	4.27	127	7.92
Hanede pazar alışkanlığı	Yok	4502	39.08	642	40.02
	Var	7019	60.92	962	59.98
Hanede sinema alışkanlığı	Yok	5815	50.47	302	18.83
	Var	5706	49.43	1302	81.17
Hanedeki bilgisayar sayısı (ortalama)		0.46 adet		1.06 adet	
Hanedeki cep telefonu sayısı (ortalama)		2.36 adet		2.50 adet	
Hanedeki öğrenci sayısı (ortalama)		0.46 kişi		0.81 kişi	
Ortalama hanehalkı büyüklüğü (hane)		3.36 kişi		3.14 kişi	
Ortalama yıllık kullanılabilir gelir (hane)		67026.29 TL		109685.54 TL	
Ortalama aylık ortalama gelir (hane)		5585.52 TL		9140.46 TL	
Ortalama aylık toplam gıda harcaması (hane)		1084.05 TL		1182.86 TL	
Ortalama aylık toplam harcama (hane)		4945.82 TL		8054.67 TL	

Tablo 1 incelendiğinde, internette alışveriş yapan hanelerin yüksek gelire sahip, toplam harcaması fazla, kredi kartı sahipliği oranı yüksek, hanehalkı reisinin eğitim düzeyinin yüksek ve yaşının düşük olduğu, internet bağlantısına sahip, bilgisayar ve telefon sayısının fazla olduğu haneler olarak görülmektedir.

## 2.2. Yöntem

2019 yılı HBA verilerine göre internette alışveriş yapan hane sayısı 1604, alışveriş yapmayan hane sayısı ise 9917'dir. Elde edilen frekanslara göre, internette alışveriş yapan ve yapmayan hane sayılarının dengesiz olduğu görülmektedir. Dengesiz veri setlerinden bilgi çıkarmak oldukça zordur, bunlarla

İlgili temel sorun standart öğrenme algoritmalarının ve sınıflandırma tekniklerinin performansını düşürme yeteneğine sahip olmasıdır (Ramyachitra ve Manikandan, 2014; Spelmen ve Porkodi, 2018). Dengesiz veri seti kullanılarak gerçekleştirilen sınıflandırma işlemlerinde, sınıflandırma algoritmaları çoğunlukla aşırı yayılım göstermeleri nedeniyle baskın sınıf lehine yanlı (taraf) sonuçlar vermektedir (Lessmann, 2004). Güvenilmez telekomünikasyon müşterilerinin tespiti, uydu radar görüntülerinde petrol sızıntılarının tespiti, kelime telaffuzlarını öğrenme, metin sınıflandırma, sahte telefon aramalarının tespiti, bilgi alma ve filtreleme görevleri gibi birçok gerçek problemde dengesiz veri kümeleri mevcuttur (Kotsiantis, Kanellopoulos, & Pintelas, 2006). Bu sorunun üstesinden gelmek için farklı yöntemler önerilmiştir. Veri düzeyinde, algoritmik yöntemler ve hibrit metotlar sıklıkla kullanılan yöntemlerdir (Spelmen ve Porkodi, 2018). Şekil 1, dengesiz veri setlerinde karşılaşılan problemlerin üstesinden gelmek için ele alınan yöntemleri göstermektedir.



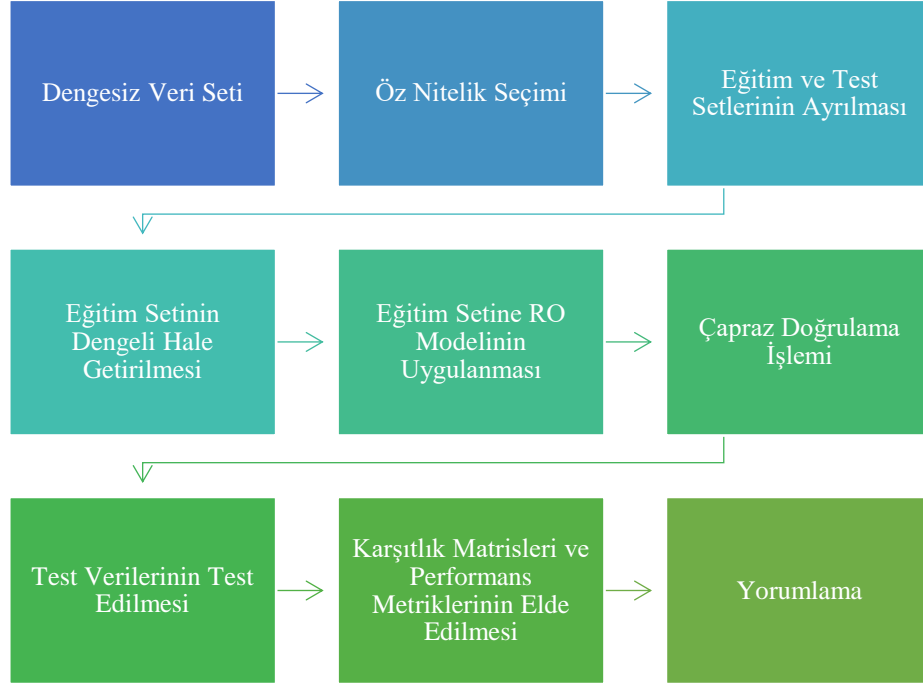
**Şekil 1.** Dengesiz Veri Setleri İle Mücadele Yöntemleri

**Kaynak:** (Spelmen ve Porkodi, 2018)

Aşırı Örneklem yönteminde, dengeli bir veri seti elde etmek için, frekansı az olan azınlık sınıfı frekansı çok olan baskın sınıfa yaklaştırılmaktadır. Yetersiz Örneklem yönteminde ise tam tersi bir işlem yapılmaktadır (Arafat, Hoque, & Farid, 2017). Çoğunluk sınıfında yapılan azaltma rastgele yapılırsa Rastgele Yetersiz Örneklem (Random Undersampling), istatistiksel bilgi kullanılarak yapılırsa Bilgili Yetersiz Örneklem (Informed Undersampling)

olarak isimlendirilir. Rastgele Aşırı Örneklem yönteminde mevcut azınlık örnekleri bir azınlık sınıfının boyutunu arttırmak için çoğaltılırken, Sentetik Aşırı Örneklem yönteminde ise azınlık sınıfı örnekleri için suni örnekler üretilmektedir (Kartal ve Özen, 2017).

Çalışmanın adım adım gösterimi Şekil 2’de gösterilmiştir.



**Şekil 2.** Çalışma Akış Şeması

Sentetik azınlık örneklem artırma yöntemi, kısaca SMOTE (synthetic minority oversampling technique) Chawla vd. (2002) tarafından geliştirilmiştir. Veri düzeyinde kullanılan en popüler yöntem olan SMOTE algoritmasının SMOTEboost (Chawla vd, 2003), Borderline-SMOTE (Han, Wang, & Mao, 2005), RAMOBoost (Chen, He, & Garcia, 2010), SMOTE-D (Torres, Carrasco-Ochoa, & Martínez-Trinidad, 2016), SMOTE-ENC (Mukherjee ve Khushi, 2021) gibi birçok türevi bulunmaktadır. Rastal örneklem yöntemlerinden farklı olarak



azınlık sınıfı verilerini kopyalamak yerine, incelenen örneklerden en yakın “k” komşusunu baz alarak yapay örnekler üretme prensibine dayanır (Yavaş, Güran, & Uysal, 2020). Gerçekleştirilen çalışmada veri setinin dengelenmesi için iki farklı yöntem kullanılmıştır. Sentetik Aşırı Örneklemeye Yöntemi (SMOTE) ve Rastgele Yetersiz Örneklemeye Yöntemi kullanılarak veri seti dengelenmiş, elde edilen her bir veri setine Rastgele Orman modeli uygulanmıştır.

Çalışmanın modelleme aşamasında, sınıflandırma problemlerinde sıklıkla tercih edilen (Dogru ve Subasi, 2018; Ecer, Yetgin, & Celik, 2018) RO yöntemi kullanılmıştır. RO yönteminin kullanıldığı araştırmalara bakıldığında; sağlık (Daş ve Türkoğlu, 2014; Gökteş ve Yağanoğlu, 2020; Subudhi, Dash, & Sabut, 2020; Buyrukoğlu, 2021; Jackins vd., 2021; Yang vd., 2021), ziraat (Zhang vd., 2020; Magidi vd., 2021), mimari (Wang vd., 2019), mühendislik (Dogru ve Subasi, 2018; Marins vd., 2021; Wei vd., 2021), elektronik ve bilgisayar (Alam ve Vuong, 2013; Farnaaz ve Jabbar, 2016; Saki vd., 2016; Sujithra ve Padmavathi, 2016; Ecer vd., 2018;), uzaktan algılama (Gislason, Benediktsson, & Sveinsson, 2004) olmak üzere birçok alanda kullanıldığı görülmektedir.

Aslında literatürde Rastgele Orman (RO) kavramı belirsiz bir kavramdır. Bazı yazarlar ağaçlar nasıl elde edilirse edilsin rastgele karar ağaçlarının toplanması olarak ifade ederken, diğer yazarlar ise Breiman'ın (2001) orijinal algoritmasına atıfta bulunmaktadır (Biau ve Scornet, 2016). Hem sınıflandırma hem de regresyon işlemlerinde kullanılabilen, çok yönlü, akıllı bir makine öğrenme yöntemi olarak tanımlanabilir. Boyutsal indirgeme yapabilme, aykırı/eksik (Sulliwana, 2017), sürekli ve kesikli değerlerle çalışabilme (Akın ve Terzi, 2020), büyük veri kümeleri üzerinde hızlı işlem yapabilme ve hesaplama açısından etkili bir teknik olan RO (Oshiro, Perez, & Baranauskas, 2012) birçok problemde kullanılabilen uzman bir yöntemdir (Sulliwana, 2017). Toplu öğrenme son zamanlarda, makine öğrenme araştırmalarında büyük ilgi duyulan bir alan haline gelmiştir. Toplu öğrenmede amaç, aynı problemin birçok sınıflandırıcı tarafından çözülerek oluşturulan sonuçların birleştirilmesine dayanmaktadır. En yaygın kullanılan toplu öğrenme metotları; Boosting, Bagging ve Rastgele Orman'dır (Oshiro vd., 2012).

Bir RO topluluğu, karar ağacının her düğümündeki bölünmeyi rastgele hale getirerek oluşturulan çok sayıda, tek ve budanmamış karar ağacı kullanır. Her ağacın, kesin bölünmelerle oluşturulan bir ağaçtan daha az doğru olması muhtemeldir. Buna karşın RO'da amaç, yaklaşık ağaçlardan birkaçını bir toplulukta birleştirerek doğruluğu artırmaya dayanmaktadır. Bu da çoğu zaman kesin bölmelere sahip tek bir ağaçtan daha iyi sonuç vermektedir (Rokach ve Maimon, 2014). RO yöntemi formal olarak Eşitlik 1'deki gibi tanımlanır,

$$\{h_k(x, T_k)\}, k = 1, 2, \dots, L \quad (1)$$

“ $T_k$ ” bağımsız ve özdeş şekilde dağıtılmış rastgele örneklerdir ve her ağaç, “ $x$ ” girdisindeki en popüler sınıf için bir birim oy vermektedir (Oshiro vd., 2012).

RO yönteminde eğitim veri setinden, kullanıcı tarafından belirlenen oluşturulacak ağaç sayısı kadar önyükleme örnekleme yapılmaktadır. Oluşturulan örneklemin 2/3’ü ormandaki ağaçları oluşturmak için InBag verisi olarak, 1/3’ü ise kurulan modelin hata oranını hesaplamak için Out Of Bag verisi olarak ayrılmaktadır. Bu hata oranı modelin kendi kendini test etmesi sonucu oluşmuş bir model içi hata oranıdır (Ayas, 2014). RO yöntemi, her bir Rastgele Ağaca ait rastgele eğitim seti örneklerini üretmek için Bagging yönteminde kullanılan yöntemin bir benzerini kullanır. Her yeni eğitim seti, orijinal eğitim setinden değiştirilerek oluşturulur. Böylece ağaç, yeni alt küme ve rastgele bir öznitelik seçimi kullanılarak oluşturulur. Seçilen rastgele nitelikler üzerindeki en iyi bölünme, düğümü bölmek için kullanılır. Bununla beraber oluşturulan ağaçlar budanmaz (Oshiro vd., 2012).

Bootstrap toplaması olarak da bilinen Bagging yöntemi Breiman tarafından geliştirilmiş, istatistiksel öğrenme yönteminin hata varyansını azaltmayı amaçlayan bir yöntemdir (Ekelik ve Altaş, 2019). Bagging yöntemi, boosting yöntemi gibi farklı sınıflandırıcılar üretmek için kullanılan bir tekniktir (Rokach ve Maimon, 2014). Öğrenme yöntemleri karşılaştırıldığı zaman Boosting yöntemi, RO yöntemine göre eğitim aşamasında daha yavaştır. Boosting yöntemi hesaplama yükü fazlalığından dolayı Bagging yönteminden daha yavaş olmasına karşın, çoğu durumda daha doğru sonuçlar vermektedir (Özdarıcı Ok, Akar, & Güngör, 2011). Bagging yönteminde her örnek eşit olasılıkla seçilirken Boosting yönteminde örnekler ağırlıklarıyla orantılı bir olasılıkla seçilir (Rokach ve Maimon, 2014). Boosting yönteminin yavaşlığı, gürültüye karşı duyarlılığı, tekrarlı eğitim gibi dezavantajlarına karşın RO, hesapsal olarak Boosting yöntemine göre daha basit ve gürültüye karşı daha sağlamdır (Özdarıcı Ok vd., 2011). RO yöntemi, Bagging ve Boosting yöntemlerine göre daha hızlı, performansları Boosting yöntemi kadar iyidir ve veriye aşırı uyum sağlamazlar (Oshiro vd., 2012). Karar ağaçlarının induksiyonunda kullanılan özniteliklerin seçimine yönelik birçok yaklaşım vardır ve çoğu yaklaşım doğrudan öznitelige bir kalite ölçüsü atamaktadır. Karar ağacı induksiyonunda en sık kullanılan öznitelik seçim ölçütleri Bilgi Kazanım Oranı (Information Gain Ratio) ve Gini İndeksidir. Rastgele Orman sınıflandırıcısı,

seçim ölçütü olarak Gini İndeksini kullanır (Pal, 2005). İtalyan ekonomist ve istatistikçi Corrado Gini'den adını alan Gini indeksi, popüler bir bölünme ölçütüdür (Berry ve Linoff, 2004). Hedef öznitelik değerlerinin olasılık dağılımları arasındaki farklılıkları ölçen Gini İndeksi Eşitlik 2'deki gibi tanımlanır;

$$Gini(y, S) = 1 - \sum_{c_j \in dom(y)} \left( \frac{|\sigma_{y=c_j} S|}{|S|} \right)^2 \quad (2)$$

Sonuç olarak  $a_i$  kriterini seçmek için değerlendirme kriteri Eşitlik 3'deki gibi tanımlanır.

$$GiniGain(a_i, S) = Gini(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot Gini(y, \sigma_{a_i=v_{i,j}} S) \quad (3)$$

(Rokach ve Maimon, 2014). C&RT'yi geliştirenlerin, Bilgi Kazancı yerine Gini İndeksini tercih etmelerinin nedeni Gini İndeksinin simetrikleştirilmiş maliyetleri içerecek şekilde genişletilebilir olması ve Bilgi Kazancından daha hızlı hesaplanmasıdır (Kantardzic, 2019). Bilgi Kazanım Oranı ise, Bilgi Kazancını Eşitlik 5'deki gibi normalleştiren bir ölçüttür.

$$Entropy(y, S) = \sum_{c_j \in dom(y)} \frac{|\sigma_{y=c_j} S|}{|S|} \cdot \log_2 \frac{|\sigma_{y=c_j} S|}{|S|} \quad (4)$$

$$InformationGainRatio(a_i, S) = \frac{InformationGain(a_i, S)}{Entropy(a_i, S)} \quad (5)$$

İki aşamalı hesaplanan bu kriterde ilk olarak Bilgi Kazancı, bütün öznitelikler için hesaplanır. En az, ortalama Bilgi Kazancı kadar performans gösteren öznitelikler dikkate alınır ve en iyi Kazanç Oranını elde eden öznitelik seçilir. Hem doğruluk hem de sınıflandırıcı karmaşıklığı açısından Bilgi Kazanım Oranının, basit Bilgi Kazancı (Information Gain) ölçütüne göre daha iyi performans gösterme eğiliminde olduğu gösterilmiştir (Rokach ve Maimon, 2014).

### 2.3. Araştırmanın etik izinleri

Gerçekleştirilen bu çalışmada Türkiye İstatistik Kurumu tarafından derlenen 2019 Yılı Hanehalkı Bütçe Anketi verileri kullanılmıştır. Veriler ikincil veri olup herhangi bir etik izin gerektirmemektedir.

### 3. UYGULAMA

Gerçekleştirilen çalışmanın amacı, hanelere ait karakteristikler kullanılarak Türkiye’de internetten alışveriş yapan hanelerin makine öğrenmesi yöntemlerinden Rastgele Orman yöntemi ile tahmin edilmesidir. Çalışmanın uygulama adımında hiçbir yöntem kullanılmadan, SMOTE ve Rasgele Yetersiz Örnekleme olmak üzere üç adet RO modeli kurulmuştur. Veri seti, test ve eğitim olarak ayrılırken, %60-%40, %70-%30, %80-%20 olmak üzere üç farklı şekilde denemiştir. Elde edilen sonuçlar arasında ciddi farklar bulunmamakla birlikte en iyi sonuç, veri setinin %60-eğitim ve %40-test olarak ayrıldığı durumda alınmıştır. Kurulan modelde en iyi sonuç 120 ağacın kullanıldığı modeldir.

Birinci modelde herhangi bir örnekleme yöntemi kullanılmamıştır. Bu modelde, 11521 haneden rastgele seçilen 6912 hane ( $\approx$ %60) eğitim, 4609 hane ( $\approx$ %40) ise test için kullanılmıştır. Modelin eğitimi için ayrılan hanelerin 981 tanesi internetten alışveriş yaparken 5931 hane internet alışverişini yapmamaktadır. Eğitim veri setinde 10’lu çapraz doğrulama yöntemi kullanılarak hedef sınıflar tahmin edilmiştir. Daha sonra eğitim aşamasında elde edilen bilgiler ve deneyimlerle model test edilmiştir. Pozitif sınıflar (internetten alışveriş yapan) için kurulan RO modeli sonucu oluşan karşıtlık matrisi Tablo 3’de gösterilmiştir.

İkinci model Sentetik Aşırı Örnekleme Yöntemi-SMOTE kullanılarak gerçekleştirilmiştir. Bu modelde, 11521 haneden rastgele seçilen 6912 hane ( $\approx$ %60) eğitim, 4609 hane ( $\approx$ %40) ise test için kullanılmıştır. Modelin eğitimi için ayrılan hanelerin 981 tanesi internetten alışveriş yaparken 5931 hane internet alışverişini yapmamaktadır. SMOTE algoritması ile alışveriş yapan hane sayısı 5931’e çıkartılarak eğitim aşamasında veri seti dengelenmiştir. Toplam 11862 haneli veri seti ile eğitim gerçekleştirilmiştir. Elde edilen yeni veri setinde 10’lu çapraz doğrulama yöntemi kullanılarak hedef sınıflar tahmin edilmiştir. Daha sonra eğitim aşamasında elde edilen bilgiler ve deneyimlerle model test edilmiştir. Pozitif sınıflar (internetten alışveriş yapan) için kurulan RO modeli sonucu oluşan karşıtlık matrisi Tablo 3’de gösterilmiştir.

Üçüncü model ise Yetersiz Örnekleme yöntemi kullanılarak gerçekleştirilmiştir. Bu modelde, 11521 haneden rastgele seçilen 6912 hane ( $\approx$ %60) eğitim, 4609 hane ( $\approx$ %40) ise test için kullanılmıştır. Modelin eğitimi için ayrılan hanelerin 981 tanesi internetten alışveriş yaparken 5931 hane internet alışverişini yapmamaktadır. Rastgele Yetersiz Örnekleme Yöntemi ile alışveriş yapan hane sayısı 981’e düşürülerek eğitim aşamasında veri seti dengelenmiştir. Toplam 1962 haneli veri seti ile eğitim gerçekleştirilmiştir. Elde edilen yeni veri setinde 10’lu çapraz doğrulama yöntemi kullanılarak hedef sınıflar tahmin edilmiştir. Daha sonra eğitim aşamasında elde edilen bilgiler ve deneyimlerle

model test edilmiştir. Pozitif sınıflar (internetten alışveriş yapan) için kurulan RO modeli sonucu oluşan karşıtlık matrisi Tablo 3’de gösterilmiştir.

Çalışmada kullanılan k-katlı çapraz doğrulama (k-fold cross validation) yönteminde veriler rastgele olarak gruplanır ve “k” alt kümeye ayrılır. Bunlardan birisi test, kalan “k-1” tanesi ise eğitim için kullanılır. Bu işlem “k” kez tekrarlanır. Sonuçların ortalaması yöntemin doğruluğunu belirler (Kahramanli ve Allahverdi, 2008). Çalışmalarda “k” sayısı için en uygun değer 10 olduğu görülmüştür (Kahramanli ve Allahverdi, 2008; Aydilek, 2018).

Tablo 2, ikili sınıflandırma problemlerinin değerlendirilmesinde kullanılan karşıtlık matrisini göstermektedir. Tablo 2’de sütunlar sınıflandırıcının tahminlerini gösterirken, satırlar ise gerçek sınıfları göstermektedir. Buna göre, doğru şekilde sınıflandırılan pozitif durumların sayısı TP, hatalı şekilde negatif olarak sınıflandırılan pozitif durumların sayısı FN, hatalı şekilde pozitif olarak tanımlanan negatif durumların sayısı FP, doğru şekilde sınıflandırılan negatif durumların sayısı ise TN olarak belirtilmiştir (Chicco ve Jurman, 2020). Açıklayıcı bir örnek verilecek olursa, gerçekte internetten alışveriş yapan ve kurulan model sonucu internetten alışveriş yaptığı belirlenen hane sayısı TP olarak belirtilmiştir.

**Tablo 2.** İkili Sınıflar İçin Karşıtlık Matrisi

		Tahmini Değer	
		1	0
Gerçek	1	TP	FN
Değer	0	FP	TN

Sınıflandırma analizinde bir sınıflandırıcının başarısı doğruluk metriği ile belirlenir, buna ek olarak kesinlik, özgüllük, duyarlık, pozitif tahmin oranı (PTO) ve negatif tahmin oranı (NTO) metrikleri de kullanılarak sınıflandırıcının başarısı değerlendirilir (Ding, 2011; Chicco ve Jurman, 2020). Eşitlik 6-11’de gösterilen bu metrikler karşıtlık matrisindeki değerler kullanılarak hesaplanır.

$$Doğruluk = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Kesinlik = \frac{TP}{TP+FP} \quad (7)$$

$$Duyarlık = \frac{TP}{TP+FN} \quad (8)$$

$$Özgüllük = \frac{TN}{TN+FP} \quad (9)$$

$$\text{Pozitif Tahmin Oranı (PTO)} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{Negatif Tahmin Oranı (NTO)} = \frac{TN}{TN+FN} \quad (11)$$

Dengesiz veri setlerinde uygulanan metodun sınıflandırma başarısının değerlendirilmesinde yanlıtıcı olabileceğinden dolayı doğruluk değeri genellikle tercih edilmez. Bunun nedeni, bu gibi durumlarda azınlık sınıfına göre çoğunluk sınıfına daha fazla ağırlık verilmesidir, bu da bir sınıflandırıcının azınlık sınıfı üzerinde iyi performans göstermesini zorlaştırır. Genelde analistler hem yanlış pozitif hem de yanlış negatif oranlar arasında denge kurmak isterler (Akosa, 2017). Bunun için dengesiz veri setlerinde Bekkar, Djemaa, & Alitouche, (2013) G-Means (geometric mean), DP (discriminant power), LR (likelihood ratios), MCC (Matthews correlation coefficient), Youden Index, Akosa (2017) Cohen's Kappa (kappa), Dengelenmiş Doğruluk (balanced accuracy) ve F ölçütü gibi birleştirilmiş ölçütler kullanılabilceğini belirtmiştir.

Gerçekleştirilen bu çalışmada modellerin değerlendirilmesinde, doğruluk (accuracy), kesinlik (precision), özgüllük (specifity), duyarlılık (sensitivity), pozitif (positive predictive value) ve negatif tahmin oranının (negative predictive value) yanı sıra G-Means, DP, MCC ve F ölçütleri kullanılmıştır.

G-Means (Geometric-Mean) hem çoğunluk hem de azınlık sınıflarında, sınıflandırma performansları arasındaki dengeyi gösterir. Düşük bir G-Means'de negatif durumlar doğru şekilde sınıflandırılmış olsa bile pozitif durumların sınıflandırılmasında zayıf bir performansın göstergesidir. G-Means, negatif sınıfa fazla ve pozitif sınıfa eksik uyumun önlenmesi açısından önemlidir (Akosa, 2017). G-Means ölçütü Eşitlik 12'de gösterilmiştir (Su ve Hsiao, 2007).

$$G - \text{Means} = \sqrt{\text{Duyarlık} \times \text{Özgüllük}} \quad (12)$$

DP (Discriminant Power) ölçütü, duyarlık ve özgüllük değerleri kullanılarak hesaplanan bir ölçüttür. DP değerinin 1'den küçük olması zayıf, 1-2 arasında olması kabul edilebilir, 2-3 arasında olması adil, 3'ün üstünde olması sınıflandırıcının iyi olduğunu gösterir. DP'nin matematiksel gösterimi Eşitlik 13'de gösterilmiştir (Sokolova, Japkowicz, & Szpakowicz, 2006).

$$DP = \frac{\sqrt{3}}{\pi} \left( \log \left( \frac{\text{Duyarlık}}{1-\text{Duyarlık}} \right) + \log \left( \frac{\text{Özgüllük}}{1-\text{Özgüllük}} \right) \right) \quad (13)$$

MCC (Matthews Correlation Coefficient) dengesiz verilerden en az etkilenen, gerçek ve tahmin edilen sınıflandırmalar arasındaki bir korelasyon ölçütüdür. (-1,+1) arasında bir değer alırken, +1 değeri mükemmel bir tahmini, -1 olası en kötü tahmini temsil etmektedir. MCC'nin 0 değerine yakın olması ise modelin rastgele yapılacak bir tahminden daha iyi olmadığını gösterir (Akosa, 2017). MCC ölçütü, karşıtlık matrisinde yer alan ve sınıflandırma algoritmalarının performansının daha iyi bir özetini veren TP, TN, FP ve FN değerlerinin tamamını kullanır (Boughorbel, Jarray, & El-Anbari, 2017). Ding (2011), MCC'nin her iki sınıftaki hem doğruluğu hem de hata oranlarını dikkate alan güçlü bir ölçüm olması nedeniyle, MCC'nin şimdiye kadarki en iyi tekil değerlendirme ölçütü olarak kabul edilebileceğini belirtmiştir. Benzer şekilde Chicco ve Jurman (2020), MCC'nin ikili sınıflandırmaların değerlendirilmesinde, doğruluk ve F puanından daha bilgilendirici ve doğru bir ölçüt olduğunu, tüm bilim toplulukları tarafından ikili sınıflandırma görevlerinin değerlendirilmesinde doğruluk ve F puanı yerine MCC katsayısının tercih edilmesi gerektiğini belirtmiştir. Brown (2018) ise MCC'nin TP, TN, FP ve FN değerlerinin tamamını kullanmasına rağmen en iyi ölçüt olamayabileceğini, bazı tahmin problemlerinde metriğin kullanımının uygun olmadığını belirtmiştir. MCC'nin matematiksel gösterimi Eşitlik 14'de gösterilmiştir (Chicco ve Jurman, 2020).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (14)$$

Kesinlik ve hassasiyet arasındaki dengeyi sağlayan F ölçütü ise Eşitlik 15'de gösterilmiştir (Akosa, 2017). F değeri, kesinlik ve hassasiyet değerlerinin artışıyla orantılı olarak artar, F değerinin yüksek olması modelin pozitif sınıfta daha iyi performans sergilediğini gösterir (Bekkar vd., 2013).

$$F = \frac{2 \times \text{Duyarlık} \times \text{Kesinlik}}{\text{Duyarlık} + \text{Kesinlik}} \quad (15)$$

İnternette alışveriş yapan haneleri tespit etmek amacıyla kurulan üç farklı RO modelinin eğitim ve test veri setlerinden elde edilen karşıtlık matrisleri Tablo 3'de gösterilmiştir.

**Tablo 3.** Rastgele Orman Modellerine Ait Eğitim ve Test Verileri Karşıtlık Matrisi

Model No	Model İsmi	Eğitim Veri Seti		Test Veri Seti		
		1	0	1	0	
1	Rastgele Orman	1	359	622	218	405
		0	153	5778	114	3872
2	Rastgele Orman (SMOTE)	1	5588	343	348	275
		0	446	5485	278	3708
3	Rastgele Orman (Rastgele Yetersiz Örnekleme)	1	784	197	500	123
		0	152	829	750	3236

\* Satırlar hedef değişkenin gerçek değerlerini, sütunlar ise tahmin değerlerini temsil etmektedir.

Dengeli veri setlerinde ikili sınıfların değerlendirilmesinde kullanılan ve karşıtlık matrisindeki verilerin Eşitlik 6-11'ye uygulanması sonucu elde edilen değerler ile dengesiz veri setlerinde kullanılan ve karşıtlık matrisindeki verilerin Eşitlik 12-15'a uygulanması sonucu elde edilen değerler Tablo 4'de gösterilmiştir. Tablo 4, pozitif sınıflar için (internetten alışveriş yapan hanelerin tespiti için) kurulan model sonuçlarını göstermektedir. Eğitim ve test veri setinde uygulanan model sonuçlarına göre tüm metriklerde en iyi değerler koyu renk ile gösterilmiştir.

**Tablo 4.** Eğitim ve Test Verileri Performans Metrikleri

Metrik	Eğitim Verileri			Test Verileri		
	Model No			Model No		
	1	2	3	1	2	3
Doğruluk	0.8879	<b>0.9335</b>	0.8221	<b>0.8874</b>	0.8800	0.8106
Duyarlık	0.3660	<b>0.9422</b>	0.7992	0.3499	0.5586	<b>0.8026</b>
Kesinlik	0.7012	<b>0.9261</b>	0.8376	<b>0.6566</b>	0.5559	0.4000
Özgüllük	<b>0.9742</b>	0.9248	0.8451	<b>0.9714</b>	0.9303	0.8118
NTO	0.9028	<b>0.9411</b>	0.8080	0.9053	0.9310	<b>0.9634</b>
PTO	0.7012	<b>0.9261</b>	0.8376	<b>0.6566</b>	0.5559	0.4000
F-Ölçütü	0.4809	<b>0.9341</b>	0.8179	0.4565	<b>0.5572</b>	0.5339
G-Means	0.1783	<b>0.4357</b>	0.3377	0.1699	0.2598	<b>0.3258</b>
DP	0.7379	<b>1.2692</b>	0.7370	<b>0.6958</b>	0.6768	0.6858
MCC	0.4533	<b>0.8671</b>	0.6449	0.4249	<b>0.4879</b>	0.4725

\*Tablo 4'deki Model Numaraları için Tablo 3'deki yazım dikkate alınmalıdır.

İkili sınıf problemlerinin değerlendirilmesinde modelin başarısı doğruluk değeri ile belirlenir. Fakat dengesiz veri setlerinde bu durum genellikle tercih edilmez (Akosa, 2017). İnternette alışveriş yapan hanelerin tespit edildiği bu problemde kesinlik ve duyarlık ölçütlerinin sonuçları doğruluktan daha önemlidir. Çünkü kesinlik ölçütü, toplam pozitif tahminler içerisinde doğru



pozitif tahminlerin oranını verirken, duyarlılık ölçütü ise toplam pozitif gözlemler içerisinde doğru tahmin edilen gözlemlerin oranını vermektedir. Eğitim veri setinde, duyarlılık ve kesinlik ölçütü sonuçlarına göre SMOTE yöntemi kullanılan modelin diğer modellere göre daha başarılı olduğu görülmektedir. Test veri setinde, duyarlılık ölçütü sonuçlarına göre Rastgele Yetersiz Örnekleme yöntemi uygulanan model, kesinlik ölçütü sonuçlarına göre ise hiçbir örnekleme yöntemi uygulanmayan modelin daha başarılı olduğu görülmüştür.

Kurulan modellerin başarısının değerlendirilmesinde Eşitlik 6-11’de belirtilen ölçütlerin dışında, özellikle veri setinin dengesiz olduğu durumlarda sınıflandırma yöntemlerinin performanslarının değerlendirilmesinde kullanılan DP, G-Means, F ve MCC ölçütleri kullanılmıştır. Eğitim veri setinde SMOTE yöntemi uygulanarak kurulan model diğer modellere kıyasla, tüm ölçütlerin (F, DP, G-Means ve MCC) sonuçlarına göre daha başarılı olduğu görülmüştür. Eğitim veri setinde maksimum  $F=0.9341$ ,  $DP=1.2692$ ,  $G\text{-Means}=0.4357$ ,  $MCC=0.8671$  olarak elde edilmiştir. Test veri setinde ise SMOTE yöntemi uygulanan model,  $F (=0.5572)$  ve  $MCC (=0.4879)$  ölçütlerinin sonucuna göre daha başarılı olduğu görülürken, Rastgele Yetersiz Örnekleme yöntemi ile kurulan modelin  $G\text{-Means} (=0.3258)$  ölçütü sonucuna göre, hiçbir yöntem kullanılmadan oluşturulan modelin ise  $DP (=0.6958)$  ölçütü sonucuna göre daha başarılı olduğu görülmüştür.

#### 4. SONUÇ

Gerçekleştirilen çalışmada, Türkiye’de internetten alışveriş yapan haneler makine öğrenmesi yöntemlerinden Rastgele Orman yöntemi ile tespit edilmiştir. TÜİK 2019 HBA veri setinde, internetten alışveriş yapan hanelerin dağılımı incelenmiş, elde edilen frekanslar sonucu dağılımın dengesiz olduğu görülmüştür. Dengesiz veri setlerinde doğru sınıflandırma tahmini için eğitim veri setinde Sentetik Aşırı Örnekleme Yöntemi-SMOTE, Rastgele Yetersiz Örnekleme Yöntemi kullanılarak veri seti dengeli hale getirilmiş, daha sonra eğitilen model RO yöntemi test edilmiştir. Ayrıca bu yöntemlerden hiçbirisi kullanılmadan RO modeli kurulmuş, kurulan üç model sonucu elde edilen metrikler ve modellerin performansları kıyaslanmıştır.

Birinci modelde herhangi bir örnekleme yöntemi kullanılmamıştır. Bu modelde, 11521 haneden rastgele seçilen 6912 hane ( $\approx\%60$ ) eğitim, 4609 hane ( $\approx\%40$ ) ise test için kullanılmıştır. Modelin eğitimi için ayrılan hanelerin 981 tanesi internetten alışveriş yaparken 5931 hane internet alışveriş yapmamaktadır. Eğitim veri setinde 10’lu çapraz doğrulama yöntemi kullanılarak hedef sınıflar tahmin edilmiştir. Daha sonra eğitim aşamasında elde edilen bilgiler ve deneyimlerle model test edilmiştir. Pozitif sınıflar (internetten alış

veriş yapan) için kurulan RO modeli sonucu oluşan karşıtlık matrisi Tablo 3’de gösterilmiştir.

İkinci model Sentetik Aşırı Örnekleme Yöntemi-SMOTE kullanılarak gerçekleştirilmiştir. Bu modelde, 11521 haneden rastgele seçilen 6912 hane ( $\approx\%60$ ) eğitim, 4609 hane ( $\approx\%40$ ) ise test için kullanılmıştır. Modelin eğitimi için ayrılan hanelerin 981 tanesi internetten alış veriş yaparken 5931 hane internet alış verişi yapmamaktadır. SMOTE algoritması ile alış veriş yapan hane sayısı 5931’e çıkartılarak eğitim aşamasında veri seti dengelenmiştir. Toplam 11862 haneli veri seti ile eğitim gerçekleştirilmiştir. Elde edilen yeni veri setinde 10’lu çapraz doğrulama yöntemi kullanılarak hedef sınıflar tahmin edilmiştir. Daha sonra eğitim aşamasında elde edilen bilgiler ve deneyimlerle model test edilmiştir. Pozitif sınıflar (internetten alış veriş yapan) için kurulan RO modeli sonucu oluşan karşıtlık matrisi Tablo 3’de gösterilmiştir.

Üçüncü model ise Yetersiz Örnekleme yöntemi kullanılarak gerçekleştirilmiştir. Bu modelde, 11521 haneden rastgele seçilen 6912 hane ( $\approx\%60$ ) eğitim, 4609 hane ( $\approx\%40$ ) ise test için kullanılmıştır. Modelin eğitimi için ayrılan hanelerin 981 tanesi internetten alış veriş yaparken 5931 hane internet alış verişi yapmamaktadır. Yetersiz Örnekleme Yöntemi ile alış veriş yapan hane sayısı 981’e düşürülerek eğitim aşamasında veri seti dengelenmiştir. Toplam 1962 haneli veri seti ile eğitim gerçekleştirilmiştir. Elde edilen yeni veri setinde 10’lu çapraz doğrulama yöntemi kullanılarak hedef sınıflar tahmin edilmiştir. Daha sonra eğitim aşamasında elde edilen bilgiler ve deneyimlerle model test edilmiştir. Pozitif sınıflar (internetten alış veriş yapan) için kurulan RO modeli sonucu oluşan karşıtlık matrisi Tablo 3’de gösterilmiştir.

Dengeli veri setlerinde ikili sınıfların değerlendirilmesinde kullanılan ve karşıtlık matrisindeki verilerin Eşitlik 6-11’ye uygulanması sonucu elde edilen değerler ile dengesiz veri setlerinde kullanılan ve karşıtlık matrisindeki verilerin Eşitlik 12-15’a uygulanması sonucu elde edilen değerler Tablo 4’de gösterilmiştir. Tablo 4’den görüleceği üzere Eğitim veri setinde, SMOTE yöntemi uygulanarak kurulan model diğer modellere kıyasla, tüm ölçütlerin (F, DP, G-Means ve MCC ) sonuçlarına göre daha başarılı olduğu görülmüştür. Test veri setinde ise SMOTE yöntemi uygulanan model, F ve MCC ölçütlerinin sonucuna göre daha başarılı olduğu görülürken, Yetersiz Örnekleme yöntemi ile kurulan modelin G-Means ölçütü sonucuna göre, hiçbir yöntem kullanılmadan oluşturulan modelin ise DP ölçütü sonucuna göre başarılı olduğu görülmüştür.

Tablo 4’den görülebileceği üzere, eğitim ve test veri setlerinde yeniden örnekleme ve çapraz doğrulamanın doğru bir şekilde kullanımı sonucu kurulan modellerin sınıflandırma performanslarının daha başarılı olabileceği görülmüştür.

Dengesiz veri setlerinin tahmini gerçek dünya problemlerinde sıklıkla karşılaşılan durumlardır. Bu çalışmada, Sentetik Aşırı Örnekleme Yöntemi-SMOTE ve Rastgele Yetersiz Örnekleme Yöntemi kullanılarak veri seti dengeli hale getirilmiş ve Rastgele Orman modeline uygulanmıştır. Elde edilen deneyimler test veri setine uygulanmış ve modellerin sınıflandırma başarıları gözlenmiştir. Bu çalışmaya ek olarak SMOTE yöntemi ve türevleri, ROSE veya Aşırı Örnekleme gibi dengesiz veri setlerinin dengeli hale getirilmesinde kullanılan yöntemler kullanılabilir ve elde edilen sonuçlar karşılaştırılabilir. Ayrıca sınıflandırma problemlerinde sıklıkla tercih edilen Destek Vektör Makineleri, Yapay Sinir Ağları, Karar Ağaçları, XGBoost, GradientBoost gibi diğer makine öğrenmesi yöntemleri ile RO modelinin ürettiği sonuçlar kıyaslanabilir.

Çalışmadan elde edilen bilgilerin, internetten alışveriş yapan hanelere yönelik reklam, pazarlama çalışmalarında kullanılabilmesi, internetten alışveriş yapmayan haneleri ise internetten alışveriş yapmaya ikna edecek satış, pazarlama ve reklam çalışmaları ve yöntemlerinin geliştirilmesine yardımcı olabileceği düşünülmektedir.

#### **5. ÇIKAR ÇATIŞMASI BEYANI**

Yazarlar arasında çıkar çatışması bulunmamaktadır.

#### **6. MADDİ DESTEK**

Bu çalışmada herhangi bir fon veya destekten yararlanılmamıştır.

#### **7. YAZAR KATKILARI (İN, EK YERİNE YAZARLAR KENDİ İSİM VE SOYİSİM KISATLIMASINI KULLANACAKTIR)**

UE: Fikir;

UE: Tasarım;

UE: Denetleme;

UE: Kaynakların toplanması ve/veya işlemesi;

UE: Analiz ve/veya yorum;

UE: Literatür taraması;

UE: Yazıyı yazan;

UE: Eleştirel inceleme

#### **8. ETİK KURUL BEYANI VE FİKRİ MÜLKİYET TELİF HAKLARI**

Bu çalışmada yapılan analizler için etik kurul izni gerekmemektedir.

#### **9. KAYNAKÇA**

Akhter, S. H. (2003). Digital divide and purchase intention: why demographic psychology matters. *Journal of Economic Psychology*, 24, 321-327.

Akhter, S. H. (2012). Who spends more online? The influence of time, usage variety, and privacy concern on online spending. *Journal of Retailing and Consumer Services*, 19, 109-115.

- Akın, P., & Terzi, Y. (2020). Dengesiz veri setli sağkalım verilerinde cox regresyon ve rastgele orman yöntemlerin karşılaştırılması. *Veri Bilimi Dergisi*, 3(1), 21-25.
- Akosa J. S. (2017). Predictive Accuracy: A misleading performance measure for highly imbalanced data. *Proceedings of The SAS Global Forum 2017 Conference*, 942. Cary, North Carolina: SAS Institute Inc.
- Al-Maghrabi, T., Dennis, C., Halliday, S. V., & BinAli, A. (2011). Determinants of Customer Continuance Intention of Online Shopping. *International Journal of Business Science & Applied Management*, 6(1), 41-65.
- Alam, M. S., & Vuong, S. T. (2013). Random forest classification for detecting android malware. 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing in (pp. 663-669). Beijing, China.
- Arafat, M. Y., Hoque, S., & Farid, D. M. (2017). Cluster-Based under-sampling with random forest for multi-class imbalanced classification. *2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)* in (pp. 1-6), Colombo, Sri Lanka.
- Armağan, E. A., & Turan, A. H. (2014). İnternet üzerinden alışveriş: demografik faktörlerin, bireysel ihtiyaçların etkisi üzerine ampirik bir değerlendirme. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 28(3), 1-22.
- Ayas, S. (2014). *Mikroskopik imgelerde tüberküloz bakterisinin rastgele ormanlar yöntemiyle sınıflandırılması*. Yayınlanmamış yüksek lisans tezi, Karadeniz Teknik Üniversitesi, Trabzon.
- Aydilek, İ. B. (2018). Yazılım hata tahmininde kullanılan metriklerin karar ağaçlarındaki bilgi kazançlarının incelenmesi ve iyileştirilmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(5), 906-914.
- Barua, L., Zou, B., & Liu, Y. (2021). Modeling Household Online Shopping Demand in the US: A Machine Learning Approach and Comparative Investigation between 2009 and 2017. arXiv preprint arXiv:2101.03690.
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, 3(10), 27-39.
- Berry, M. J. A. and Linoff, G. S. (2004). *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management (Second Edition)*. Indianapolis, Indiana: Wiley Publishing Inc.
- Bhatti, A., & Akram, H. (2020). The moderating role of subjective norms between online shopping behaviour and its determinants. *International Journal of Social Sciences and Economic Review*, 2(1), 1-9.
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25, 197-227.
- Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PloS One*, 12(6), 1-17.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Brown, J. B. (2018). Classifiers and their metrics quantified. *Molecular Informatics*, 37, 1-11.

- Buyrukoğlu, S. (2021). Early detection of alzheimer's disease using data mining: comparison of ensemble feature selection approaches. *Konya Mühendislik Bilimleri Dergisi*, 9(1), 50-61.
- Cengiz E., & Şekerkeya A. (2010), İnternet kullanıcılarının internette alış-verişe yönelik satın alma karar süreçlerinin incelenmesi ve kullanım yoğunlukları açısından sınıflandırılması üzerine bir araştırma, *Öneri Dergisi*, 9(33), 33-49.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- Chawla, N. V., Lazarevic, A., Hall, L. O., & Bowyer, K. W. (2003). SMOTEBoost: Improving prediction of the minority class in boosting. *Knowledge Discovery in Databases: PKDD 2003, 7th European Conference on Principles and Practice of Knowledge Discovery in Databases* in (pp. 107-119). Dubrovnik, Croatia.
- Chen, S., He, H., & Garcia, E. A. (2010). RAMOBoost: ranked minority oversampling in boosting. *IEEE Transactions on Neural Networks*, 21(10), 1624-1642.
- Chicco, D., & Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(6), 1-13.
- Çiçek, R., & Mürütsoy, M. (2014). İnternet tüketicisinin satın alma davranışlarının incelenmesi üzerine bir araştırma: Niğde Üniversitesi Akademik ve İdari Personeli Örneği. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 15(2), 291-305.
- Daş, B., Türkoğlu, İ. (2014). DNA dizilimlerinin sınıflandırılmasında karar ağacı algoritmalarının karşılaştırılması, *Elektrik – Elektronik - Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu (ELECO 2014)* içinde (ss.381-383), Bursa.
- Ding, Z. (2011). *Diversified ensemble classifiers for highly imbalanced data learning and their application in bioinformatics*, Unpublished dissertation, Georgia State University, The USA.
- Dogru, N., & Subasi, A. (2018). Traffic accident detection using random forest classifier. *2018 15th Learning and Technology Conference (L&T)* in (pp. 40-45), Jeddah, Saudi Arabia.
- Doğrul, Ü. (2012). Elektronik alışveriş davranışında faydacı ve hedonik güdülerin etkisi. *Sosyal ve Beşeri Bilimler Dergisi*, 4(1), 321-331.
- Ecer, O., Yetgin, Z., & Celik, T. (2018). Air write letter recognition using random forest classification on arduino dataset. *International Journal of Scientific and Technological Research*, 4(7), 1-9.
- Ekelik, H., & Altaş, D. (2019). Dijital reklam verilerinden yararlanarak potansiyel konut alıcılarının rastgele orman yöntemiyle sınıflandırılması. *İktisat Araştırmaları Dergisi*, 3(1), 28-45.
- Ertürk, R., & Aktepe, C. (2020). Sosyal medyanın tüketicilerin satınalma davranışları üzerinde etkisi: Türkiye ve İran'daki üniversite öğrencileri üzerine bir uygulama. *İşletme Araştırmaları Dergisi*, 12(4), 4289-4304.

- EUROSTAT (2021). Internet purchases by individuals (until 2019), 04 Temmuz 2021 tarihinde [https://ec.europa.eu/eurostat/databrowser/view/isoc\\_ec\\_ibuy/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/isoc_ec_ibuy/default/table?lang=en) adresinden erişildi.
- Farnaaz, N., & Jabbar, M. A. (2016). Random forest modeling for network intrusion detection system. *Procedia Computer Science*, 89, 213-217.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2004). Random forest classification of multisource remote sensing and geographic data. *IGARSS 2004. IEEE International Geoscience and Remote Sensing Symposium Vol. 2*, in (pp. 1049-1052).
- Goldsmith, R. E., & Flynn, L. R. (2005). Bricks, clicks, and pix: apparel buyers' use of stores, internet, and catalogs compared. *International Journal of Retail & Distribution Management*, 33(4), 271-283.
- Göktaş, M. E., & Yağanoğlu, M. (2020). Veri bilimi uygulamalarının hastalık teşhisinde kullanılması: kalp krizi örneği. *Bilişim Sistemleri ve Yönetim Araştırmaları Dergisi*, 2(2), 26-32.
- Gültaş, P., & Yıldırım, Y. (2016). İnternette alışverişte tüketici davranışını etkileyen demografik faktörler. *Dicle Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 6(10), 32-50.
- Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning Advances in Intelligent Computing, International Conference on Intelligent Computing, ICIC 2005, in (pp. 878-887). Hefei, China.
- Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77, 5198-5219.
- Jothi, C. A., & Gaffoor, A. M. (2017). Impact of social media in online shopping. *Journal on Management Studies*, 3(3), 576-586.
- Kahramanli, H., & Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications*, 35, 82-89.
- Kantardzic, M. (2019). *Data Mining Concepts, Models, Methods, and Algorithms Third Edition*. New Jersey: John Wiley & Sons, Inc.
- Kartal E. ve Özen Z., (2017) Dengesiz veri setlerinde sınıflandırma, Orhan Torkul, Sevinç Gülseçen, Yılmaz Uyaroğlu, Gültekin Çağıl, Muhammed Kürşad Uçar (Ed.) *Mühendislikte Yapay Zekâ ve Uygulamaları* içinde (ss. 109-131). Sakarya: Sakarya Üniversitesi Yayınları, No: 184.
- Koehn, D., Lessmann, S., & Schaal, M. (2020). Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*, 150, 1-16.
- Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: a review. *GESTS International Transactions on Computer Science and Engineering*, 30, 25-36.

- Kumar, A., Kabra, G., Mussada, E. K., Dash, M. K., & Rana, P. S. (2019). Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention. *Neural Computing and Applications*, 31(2), 877-890.
- Lessmann, S. (2004). Solving imbalanced classification problems with support vector machines. International Conference on Artificial Intelligence (ICAI), in (pp. 214-220), The USA: Las Vegas, Nevada.
- Magidi, J., Nhamo, L., Mpandeli, S., & Mabhaudhi, T. (2021). Application of the random forest classifier to map irrigated areas using google earth engine. *Remote Sensing*, 13, 876, 1-16.
- Maheswari, K., & Priya, P. P. A. (2017). Predicting customer behavior in online shopping using SVM classifier. In 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1-5).
- Marins, M. A., Barros, B. D., Santos, I. H., Barrionuevo, D. C., Vargas, R. E. V., Prego, T. D. M., Lima, A. A. D., Campos, M. L. R. D., Silva, E. A. B. D., & Netto, S. L. (2021). Fault detection and classification in oil wells and production/service lines using random forest. *Journal of Petroleum Science and Engineering*, 197, 1-10.
- Mukherjee, M., & Khushi, M. (2021). SMOTE-ENC: A Novel SMOTE-based method to generate synthetic data for nominal and continuous features. *Applied System Innovation*, 4, 18, 1-12.
- Oke, T., & Ramachandran, T. (2021). Determinants of Decision to Use and Continued Use of Online Shopping Medium: A Bivariate Probit Approach, 19 February 2021, PREPRINT (Version 1), 1-13. available at Research Square [https://doi.org/10.21203/rs.3.rs-217453/v1]
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? *International Workshop on Machine Learning and Data Mining in Pattern Recognition* in (pp. 154-168). Springer-Verlag, Berlin, Heidelberg.
- Özdarıcı Ok, A., Akar, Ö., & Güngör, O. (2011). Rastgele orman sınıflandırma yöntemi yardımıyla tarım alanlarındaki ürün çeşitliliğinin sınıflandırılması. *TUFUAB 2011 VI. Teknik Sempozyumu* içinde (ss.1-7). Antalya, Türkiye.
- Özgüven, N. (2011). Tüketicilerin online alışverişe karşı tutumları ile demografik özellikleri arasındaki ilişkinin analizi. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 13(21), 47-54.
- Özhan, Ş., & Altuğ, N. (2015). Tüketicilerin demografik özelliklerinin online alışveriş davranışları üzerine etkileri. *Ege Akademik Bakış*, 15(4), 481-493.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217-222.
- Ramyachitra, D., & Manikandan, P. (2014). Imbalanced dataset classification and solutions: a review. *International Journal of Computing and Business Research (IJCBR)*, 5(4), 1-29.

- Rausch, T. M., Derra, N. D., & Wolf, L. (2020). Predicting online shopping cart abandonment with machine learning approaches. *International Journal of Market Research*, 1-24.
- Richards, J., & Shen, D. (2006). E-commerce adoption among chinese consumers: an exploratory Study. *Journal of International Consumer Marketing*, 18(3), 33-55.
- Rokach, L., & Maimon, O. (2014). *Data Mining With Decision Trees Theory and Applications (2nd Edition)*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Saki, F., Sehgal, A., Panahi, I., & Kehtarnavaz, N. (2016). Smartphone-based real-time classification of noise signals using subband features and random forest classifier. *41st IEEE International Conference On Acoustics, Speech and Signal Processing (ICASSP)* in (pp. 2204-2208). Shanghai, China.
- Sokolova M., Japkowicz N., Szapkowicz S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: Sattar A., Kang B. (eds) *AI 2006: Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science*, vol 4304. Springer, Berlin, Heidelberg.
- Soler, J. R. L., Christidis, P., & Vassallo, J. M. (2021). Teleworking and online shopping: socio-economic factors affecting their impact on transport demand. *Sustainability*, 13, 1-24.
- Spelmen, V. S., & Porkodi, R. (2018). A review on handling imbalanced data. *2018 International Conference on Current Trends toward Converging Technologies (ICCTCT)* in (pp. 1-11), Coimbatore, India.
- Sreerekha, T., Saranya, R., & Prabhu, V. S. (2019). Consumer behaviour in online shopping. *International Journal of Trend in Scientific Research and Development*, 3(5), 460-464.
- Su, C. T., & Hsiao, Y. H. (2007). An evaluation of the robustness of MTS for imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 19(10), 1321-1332.
- Subudhi, A., Dash, M., & Sabut, S. (2020). Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier. *Biocybernetics and Biomedical Engineering*, 40, 277-289.
- Sujithra, M., & Padmavathi, G. (2016). Enhanced permission based malware detection in mobile devices using optimized random forest classifier with PSO-GA. *Research Journal of Applied Sciences, Engineering and Technology*, 12(7), 732-741.
- Sullivan, W. (2017). *Machine Learning For Beginners Guide Algorithms: Supervised & Unsupervised Learning Decision Tree & Random Forest Introduction*. Healthy Pragmatic Solutions Inc.
- Torres, F. R., Carrasco-Ochoa, J. A., Martínez-Trinidad, J. F. (2016) SMOTE-D a Deterministic Version of SMOTE. In: Martínez-Trinidad J., Carrasco-Ochoa J., Ayala Ramirez V., Olvera-López J., Jiang X. (eds) *Pattern Recognition. MCP 2016. Lecture Notes in Computer Science*, vol 9703. Springer, Cham.
- Torun, E. (2017). Tüketici satın alma davranışları üzerinde internet ve sosyal medyanın yeri ve önemi. *Elektronik Sosyal Bilimler Dergisi*, 16(62), 955-970.



- TÜİK (2019). Türkiye İstatistik Kurumu, Hanehalkı Bütçe Anketi 2019, Ankara.
- TÜİK (2021a). Hanehalkı Bilişim Teknolojileri Kullanım Araştırması 2011. 08 Kasım 2021 tarihinde [https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-\(BT\)-Kullanim-Arastirmasi-2011-8572](https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2011-8572) adresinden erişildi.
- TÜİK (2021b). Hanehalkı Bilişim Teknolojileri Kullanım Araştırması 2015. 08 Kasım 2021 tarihinde [https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-\(BT\)-Kullanim-Arastirmasi-2015-18660](https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2015-18660) adresinden erişildi.
- TÜİK (2021c). Hanehalkı Bilişim Teknolojileri Kullanım Araştırması 2020. 08 Kasım 2021 tarihinde [https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-\(BT\)-Kullanim-Arastirmasi-2020-33679](https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2020-33679) adresinden erişildi.
- Vasić, N., Kilibarda, M., & Kaurin, T. (2019). the influence of online shopping determinants on customer satisfaction in the serbian market. *Journal of Theoretical and Applied Electronic Commerce Research*, 14(2), 70-89.
- Wang, C., Shu, Q., Wang, X., Guo, B., Liu, P., & Li, Q. (2019). A random forest classifier based on pixel comparison features for urban LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148, 75-86.
- Wei, Y., Yang, Y., Xu, M., & Huang, W. (2021). Intelligent fault diagnosis of planetary gearbox based on refined composite hierarchical fuzzy entropy and random forest. *ISA Transactions*, 109, 340-351.
- Yang, P., Wang, D., Zhao, W. B., Fu, L. H., Du, J. L., & Su, H. (2021). Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification. *Biomedical Signal Processing and Control*, 63, 102138, 1-13.
- Yavaş, M., Güran, A. & Uysal, M. (2020). Covid-19 Veri Kümesinin SMOTE Tabanlı Örnekleme Yöntemi Uygulanarak Sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*, Ejosat Özel Sayı 2020 (HORA), 258-264.
- Zhang, L., Liu, Z., Ren, T., Liu, D., Ma, Z., Tong, L., Zhang, C., Zhou, T., Zhang, X., & Li, S. (2020). Identification of seed maize fields with high spatial resolution and multiple spectral remote sensing using random forest classifier. *Remote Sensing*, 12, 362, 1-19.
- Zhang, Y., Trusov, M., Stephen, A. T., & Jamal, Z. (2017). Online shopping and social media: friends or foes?. *Journal of Marketing*, 81, 24-41.