

**T.C.
PAMUKKALE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

**MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TÜRKİYE'DE
TURİZM GELİRİNE ETKİ EDEN PARAMETRELERİN
BELİRLENMESİ VE DEĞERLENDİRİLMESİ**

YÜKSEK LİSANS TEZİ

ÖZGÜR GÜLER

DENİZLİ, OCAK - 2021

**T.C.
PAMUKKALE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**



**MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TÜRKİYE'DE
TURİZM GELİRİNE ETKİ EDEN PARAMETRELERİN
BELİRLENMESİ VE DEĞERLENDİRİLMESİ**

YÜKSEK LİSANS TEZİ

ÖZGÜR GÜLER

DENİZLİ, OCAK - 2021

Bu tezin tasarımı, hazırlanması, yürütülmesi, arařtırmalarının yapılması ve bulgularının analizlerinde bilimsel etięe ve akademik kurallara özenle riayet edildiđini; bu alıřmanın dođrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etięe uygun olarak kaynak gösterildiđini ve alıntı yapılan alıřmalara atfedildiđine beyan ederim.

Özgür GÜLER

ÖZET

**MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TÜRKİYE’DE TURİZM
GELİRİNE ETKİ EDEN PARAMETRELERİN BELİRLENMESİ VE
DEĞERLENDİRİLMESİ
YÜKSEK LİSANS TEZİ
ÖZGÜR GÜLER
PAMUKKALE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

(TEZ DANIŞMANI:DOÇ. DR. OLCAY POLAT)

DENİZLİ, OCAK - 2021

Turizm sektörü ülkeler üzerinde geniş ekonomik etkileri olan en büyük sektörlerden biridir. Küresel etkiler ve gelişen teknolojik düzenle beraber daha da yaygınlaşan seyahat ve turizm sektörü büyüme ivmesini her geçen gün arttırmaktadır. Turizmde yabancı ziyaretçi sayısı bakımından üst sıralarda yer alan Türkiye turizm gelirinde ise dünya ortalamasının çok altında kalmaktadır. Bu durum göstermektedir ki çalışmalarda turizm etkinliğinin ölçülmesinde genellikle kullanılan ziyaretçi sayısının aksine turizm gelirleri üzerinde durulmalıdır. Bu çalışmada 2003-2019 dönemlerinde çeyrek yıllık verilerle ekonomi, turizm kapasitesi, özgürlük, güvenlik ve gelişmişlik düzeyi konularında belirlenen 20 adet faktörün turizm gelirin etkileri Türkiye için araştırılmıştır. Araştırmada birim kök ve eşbütünleşme analizleriyle desteklenmiş çoklu doğrusal regresyon modeli, esnek geri yayılım algoritmasına sahip çok katmanlı yapay sinir ağları modeli ve gradyan arttırılmış regresyon ağaçları modeli karşılaştırılmıştır. Çalışmada turizm gelirin etkileyen değişkenler küresel terör endeksi ve dış hatlar uçuş trafiği olarak belirlenmiştir. Çalışmanın sonucunda klasik istatistikî yöntemlerle belirlenen bağımlı değişkenlerle en iyi performansla sahip model yapay sinir ağları olarak bulunmuştur.

ANAHTAR KELİMELELER: Turizmi etkileyen faktörler, Doğrusal regresyon, Yapay sinir ağları, Gradyan arttırılmış ağaçlar

ABSTRACT

DETERMINATION AND EVALUATION OF PARAMETERS AFFECTING TOURISM REVENUES IN TURKEY BY MACHINE LEARNING METHODS

MSC THESIS

ÖZGÜR GÜLER

**PAMUKKALE UNIVERSITY INSTITUTE OF SCIENCE
INDUSTRIAL ENGINEERING**

(SUPERVISOR:ASSOC. PROF. DR. OLCAY POLAT)

DENİZLİ, JANUARY 2021

The tourism is one of the most important sector that affects economies of countries widely. Travel and tourism sectors, which has become more widespread with globalization and developing technology is growing day by day. Turkey is located high rankings in terms of tourist arrivals in the world, but it has lower tourism revenues compared to the world. This shows that tourism revenues should be focused on instead of the tourist arrivals, which is generally used in measuring tourism efficiency in studies. In this study, twenty factors related to economy, tourism capacity, development, freedom and security were investigated in terms of their impact on tourism revenues with quaterly data from 2003 to 2019 in Turkey. Multiple linear regression model supported by unit root tests and cointegration analysis, multilayer artificial neural network supported by resilient propagation algorithm and gradient boosted regression trees models were compared in the research. It was seen that the variables that significantly affect tourism revenues are the global terrorism index and international flight traffic. As a result of the study, neural networks with dependent variables determined by classical statistical methods gave the best performance.

KEYWORDS: Factors affecting tourism, Machine learning, Linear regression, Artificial neural networks, Gradient boosted trees

İÇİNDEKİLER

Sayfa

ÖZET.....	i
ABSTRACT	ii
İÇİNDEKİLER	iii
ŞEKİL LİSTESİ	v
TABLO LİSTESİ	vi
SEMBOL LİSTESİ	vii
ÖNSÖZ.....	viii
1. GİRİŞ.....	1
2. LİTERATÜR TARAMASI	4
2.1 Türkiye’de Yapılan Çalışmalar	4
2.2 Dünyada Yapılan Çalışmalar	6
2.3 Literatür Özeti	9
3. DÜNYADA VE TÜRKİYE’DE TURİZM SEKTÖRÜ.....	11
3.1 Küreselleşmenin Turizm Üzerindeki Etkisi	11
3.2 Turizmin Ekonomi Üzerindeki Etkisi	13
3.3 Türkiye Açısından Turizm Geliri	16
4. YÖNTEM.....	23
4.1 Çoklu Doğrusal Regresyon	23
4.1.1 Çoklu Doğrusal Regresyonun Gereklilikleri	24
4.1.1.1 Normal Dağılım Varsayımı.....	25
4.1.1.2 Korelasyon Varsayımı ve Çoklu Doğrusallık Problemi.....	25
4.1.1.3 Otokorelasyon Problemi.....	26
4.1.1.4 Sabit Varyans Varsayımı.....	28
4.1.1.5 Sahte Regresyon Problemi	28
4.1.2 Durağanlık ve Durağanlığın Ölçülmesi	29
4.1.2.1 Birim Kök Analizi.....	29
4.1.3 Eşbütünlüşme Analizi	32
4.2 Yapay Sinir Ağları.....	34
4.2.1 Yapay Sinir Hücreleri	34
4.2.2 Yapay Sinir Ağlarının İşleyişi	36
4.2.3 Yapay Sinir Ağlarının Sınıflandırılması	37
4.2.3.1 Akış Yönüne Göre Yapay Sinir Ağları	38
4.2.3.2 Katman Sayısına Göre Yapay Sinir Ağları	38
4.2.4 Geri Yayılım Algoritması	39
4.2.5 Esnek Geri Yayılım Algoritması	40
4.2.6 Yapay Sinir Ağlarının Avantajları ve Dezavantajları.....	42
4.2.7 Yapay Sinir Ağlarının Kullanım Alanları.....	43
4.3 Gradyan Arttırmalı Regresyon Ağaçları	44
4.3.1 Regresyon Ağaçları.....	44
4.3.2 Gradyan Arttırmalı Regresyon Ağaçlarının İşleyişi	45
4.3.3 Gradyan Arttırmalı Regresyon Ağaçları Modelinin Avantajları ve Dezavantajları	48
4.3.4 Gradyan Arttırmalı Regresyon Ağaçlarının Kullanım Alanları ..	49
5. UYGULAMA VE SONUÇLAR.....	51
5.1 Araştırma Tasarımı	51

5.2	Araştırmanın Sınırları ve Veri Derleme Süreci	54
5.3	Çoklu Doğrusal Regresyon Modeli Uygulaması.....	58
5.3.1	Çoklu Doğrusal Regresyon Gerekliliklerinin İncelenmesi	58
5.3.1.1	Durağanlık Analizi	60
5.3.1.2	Durağan Olmayan Değişkenlerin Eşbütünleşme Analizleri ...	61
5.3.1.3	Geriye Doğru Ayıklama ve Çoklu Doğrusallık Analizi.....	62
5.3.1.4	Normal Dağılım Testi	63
5.3.1.5	Varyansların Değişkenliği Testi.....	64
5.3.1.6	Otokorelasyon Analizi	65
5.3.2	Çoklu Doğrusal Regresyon Uygulamasının Çıktıları	65
5.4	Makine Öğrenmesi Modellerinin Uygulanması	67
5.4.1	Yapay Sinir Ağları Uygulaması.....	67
5.4.2	Gradyan Arttırmalı Regresyon Ağaçları Uygulaması	69
5.5	Modellerin Karşılaştırılması	70
5.5.1	Eğitim ve Test Veri Setlerinin Belirlenmesi	70
5.5.2	Performans Kriterlerinin Belirlenmesi ve Karşılaştırılması	72
6.	TARTIŞMA VE ÖNERİLER.....	78
7.	KAYNAKLAR.....	81
8.	ÖZGEÇMİŞ.....	89

ŞEKİL LİSTESİ

Sayfa

Şekil 3.1: Turizmin ekonomik etkileri.	14
Şekil 3.2: Türkiye kişi başı turizm geliri göstergesi.	20
Şekil 3.3: Türkiye toplam turizm geliri göstergesi.	21
Şekil 4.1: Otokorelasyon sorunu grafik gösterimi.	27
Şekil 4.2: Bazı aktivasyon fonksiyonları (Çayıroğlu 2015).	36
Şekil 4.3: Örnek yapay sinir ağları modeli gösterimi (Karakullukçu 2020).	37
Şekil 5.1: Uygulamanın akış şeması üzerinde gösterimi.	53
Şekil 5.2: Turizm geliri için mevsimsellikten arındırma grafiği.	57
Şekil 5.3: Turizm gelirinin mevsimsellik faktörü grafiği.	58
Şekil 5.4: Nihai regresyon denkleminin normallik dağılımı gösterimi.	64
Şekil 5.5: Turizm gelirinin çoklu doğrusal regresyon ile tahmin grafiği.	66
Şekil 5.6: Çoklu doğrusal regresyon Knime modeli akış gösterimi.	66
Şekil 5.7: Yapay sinir ağları Knime modeli akış gösterimi.	69
Şekil 5.8: Gradyan arttırımlı regresyon ağacı Knime modeli akış gösterimi. ...	70
Şekil 5.9: K-katlamalı çapraz doğrulama Knime modülleri.	71
Şekil 5.10: Modellerin tahmin değerlerinin karşılaştırılması grafiği.	74
Şekil 5.11: Modellerin mutlak yüzde hatalarının karşılaştırılması grafiği.	75
Şekil 5.12: Turizm gelirinin en iyi model ile tahmin grafiği.	75

TABLO LİSTESİ

Sayfa

Tablo 2.1: Literatür karşılaştırma tablosu.	9
Tablo 5.1: Bağımsız değişkenlerin genel özelliklerinin gösterimi.....	55
Tablo 5.2: Bağımsız değişkenlerin kaynak gösterimi.	56
Tablo 5.3: Değişkenlerin çarpıklık ve basıklık değerleri.	59
Tablo 5.4: ADF birim kök testi sonuçları.	60
Tablo 5.5: Zivot-Andrews birim kök testi sonuçları.	61
Tablo 5.6: Gregory-Hansen eşbütünleşme testi sonuçları.....	62
Tablo 5.7: Nihai regresyon denklemi ve t istatistik değerleri gösterimi.	63
Tablo 5.8: Nihai regresyon denkleminin VIF değerleri gösterimi.	63
Tablo 5.9: Nihai regresyon denkleminin varyans testinin gösterimi.	65
Tablo 5.10: Yapay sinir ağları parametre gösterimi.....	68
Tablo 5.11: Gradyan arttırılmalı regresyon ağaçları parametre gösterimi.	69
Tablo 5.12: Bölümleme parametreleri gösterimi.	71
Tablo 5.13: Çoklu doğrusal lineer regresyon test sonuçları.....	72
Tablo 5.14: Gradyan arttırılmalı regresyon ağaçları test sonuçları.	73
Tablo 5.15: Yapay sinir ağları test sonuçları.	74
Tablo 5.16: YSA ve GBRT modelleri test sonuçları.	76

SEMBOL LİSTESİ

EKK	:	En Küçük Kareler
GSYİH	:	Gayri Safi Yurtiçi Hasıla
GSMH	:	Gayri Safi Milli Hasıla
ADF	:	Genişletilmiş Dickey Fuller
YSA	:	Yapay Sinir Ağları
MLP	:	Çok Katmanlı Algılayıcı
GBRT	:	Gradyan Arttırılmış Regresyon Ağaçları
VIF	:	Varyans Şişirme Faktörü
MAPE	:	Ortalama Mutlak Yüzde Hata
APE	:	Mutlak Yüzde Hata
MSE	:	Ortalama Hata Kareleri
RPROP	:	Esnek Geri Yayılım

ÖNSÖZ

Gerek yüksek lisans sürecimde yapmış olduđu desteklerden dolayı gerekse iş hayatımda vermiş olduđu güzel tavsiyelerden dolayı değerli danışman hocam Doç. Dr. Olcay POLAT'a

Tez sürecindeki katkılarından ve değerlendirmelerinden dolayı değerli hocam Dr. Öğr. Üyesi Hande Mutlu ÖZTÜRK'e ve yine değerlendirmelerinden dolayı değerli hocam Doç. Dr. Utkucañ ŞAHİN'e

Tez sürecinde bilgilerinden faydalandığım ve yorulmadan bütün sorularıma cevap veren sevgili hocam Dr. Seyyide DOĞAN'a

Hayatımın her alanında beni geliştiren, her konuda beni destekleyen canım abim Dr. Ceyhun GÜLER'e

Son olarak da benden desteklerini hiç esirgemeyen ve bu süreçte sabırla yanımda olan aileme gönülden saygı ve sevgilerimi sunar teşekkür ederim.

1. GİRİŞ

Turizm, uzun yıllardır ülkeler arası döviz akışında önemli bir katkıya sahip olan, insanlara ziyaret ettikleri ülkeleri çoklu bakış açısıyla değerlendirme ve ülkelerin tarihlerine, kültürlerine ve doğal güzelliklerine daha yakından şahit olma gibi fırsatlar sunan çok yönlü bir sektördür. Turizm sayesinde insanlar teknolojinin zayıf olduğu tarihlerde bile farklı kültürleri ziyaret etme, onları tanıma gibi ihtiyaçlarını gidererek küreselleşmeye katkıda bulunmuşlardır. Küreselleşmenin oldukça yoğunlaştığı ve teknolojinin geliştiği bu dönemlerde insanların başka ülkelere ulaşımı kolaylaşmış, bir ülkenin turistik destinasyonlarıyla, kültürüyle ve insanlarıyla ilgili bilgi toplamak daha kolay hale gelmiştir. Dolayısıyla turizm günümüzde daha kolay, daha güvenilir ve daha yaygın bir hal almaktadır.

Turizm artışının en önemli etkilerinden bir tanesi kuşkusuz ekonomik getirilerdir. Turizmin istihdama ve dövize doğrudan katkısının yanı sıra ziyaretçilerin kişisel harcamaları, turizm sektörüne ürün ve hizmet sağlayan yardımcı sektörlerle etkisi gibi durumlardaki dolaylı katkıları da göz önüne alınmalıdır. Dünya Seyahat ve Turizm Konseyi'nin 2019 yılı için hazırladığı ekonomik etki raporuna göre; son beş yıldır yaratılan her dört yeni iş imkanından birini seyahat ve turizm sektörü oluşturmakta ve bu hükümetler için seyahat ve turizmi en iyi istihdam yaratan sektör haline getirmektedir. Bunun dışında raporda seyahat ve turizm sektörünün, 8,9 trilyon dolar ile dünyada Gayri Safi Yurtiçi Hasılanın (GSYİH) %10,3'ünü, küresel toplam ihracatın %6,8'ini ve küresel hizmet ihracatının %28,3'ünü oluşturduğu belirtilmektedir (WTTC 2019).

Türkiye kültürel mirası, tarihi dokusu, doğal güzellikleri, jeopolitik konumu nedeniyle turizm sektöründe en çok tercih edilen ülkeler arasında yer almaktadır. Turizm gelirinin ülkemizde hem ödemeler dengesine hem de istihdama katkısı göz önünde bulundurulduğunda sektörün en önemli ekonomik kaynaklardan biri olduğu gözler önüne serilmektedir. Buna rağmen ülkemiz turist sayısı olarak rekabet içinde olduğu ülkelere kıyasla yeterli turizm geliri seviyelerine ulaşamamaktadır. T.C. Kültür ve Turizm Bakanlığı (2019) genel raporunda 2018 yılı için Türkiye'nin ziyaretçi sayısı

olarak dünyada 6. sırada yer almasına rağmen turizm geliri olarak 15. sıraya kadar gerilediği görülmektedir. Başka bir şekilde ifade etmek gerekirse 2019 yılında dünyada kişi başı turizm geliri ortalaması 1020 \$ iken bu rakam Türkiye’de 580 \$, turist sayısı sıralaması olarak Türkiye’den sonrası gelen Almanya’da 1050 \$ ve Tayland’da 1520 \$ olarak hesaplanmıştır. Amerika Birleşik Devletleri için ise kişi başı ortalama turizm geliri 2700 \$ olarak gösterilmektedir (UNWTO 2020). Böylece açıkça görülmektedir ki yabancı ziyaretçi sayısı bakımından dünyada tatmin edici bir yere sahip olan Türkiye aynı başarıyı turizm geliri konusunda gösterememektedir.

Bu çalışmanın amacı, turizm gelirini etkileyen en önemli faktörlerin neler olduğunun anlamlı istatistiksel yöntemlerle belirlenmesi ve belirlenen bu değişkenlerin farklı modellerle ne kadar yüksek performansta turizm gelirini açıklayabileceğinin karşılaştırmalı olarak gösterilmesidir. Böylece turizmle ilgili faaliyet planlamaları yapılırken hangi faktörlere dikkat edilebileceği ve gelecekte yapılacak öngörülerde hangi modellerin avantaj sağlayabileceği konusunda ilgili kişilere yol gösterici sonuçların ortaya konulması amaçlanmıştır.

Turizm geliri üzerinde etkili olabileceği düşünülen 20 adet değişken daha önceki çalışmalardan esinlenilerek ve yabancı ziyaretçilerin hangi nedenlerle bir ülkeyi turistik ziyaret için tercih edebileceği veya o ülkeyi tercih etmekten vazgeçebileceği düşünülerek araştırma kapsamına alınmıştır. Bu bağlamda turizm gideri, Gayri Safi Yurtiçi Hasıla (GSYİH), ihracat, ithalat, enflasyon, işsizlik, uluslararası yatırım miktarı, euro ve dolar satış kurları, altın kurları, küresel terör endeksi, basın özgürlük endeksi, tutuklu ve hükümlü sayıları, insani gelişmişlik endeksi, beşeri sermaye endeksi, kentsel nüfus oranı, dış hatlar uçuş sayısı, tesis yatak kapasitesi, seyahat acenteleri sayısı, turizm teşvik miktarları araştırma kapsamına dahil edilmiştir.

Bu çalışmada turizm gelirini etkileyen değişkenler üzerinde araştırma yapıp en önemli değişkenler klasik istatistiksel yöntemlerle belirlendikten sonra kurulan modelin performansı makine öğrenmesi modelleriyle karşılaştırılmıştır. Önemli açıklayıcı değişkenler belirlenirken ekonometrik modellerle desteklenmiş çoklu doğrusal regresyon modelinden yararlanılmıştır. Kurulan çoklu doğrusal regresyon modelinin performansı, Yapay Sinir Ağları (YSA) ve Gradyan Arttırmalı Regresyon Ağaçları (GBRT) modelleriyle karşılaştırılmıştır.

Bu çalışma yapılan diğer çalışmalardan iki ana konuda farklılaşmaktadır. Literatürde genellikle makroekonomik değişkenlerin turizmle ilişkisi üzerine yapılmış çalışmalar mevcuttur (Uysal ve El Roubi 1999; Dritsakis 2004; Aydın ve diğ. 2015; Pekmezci ve Bozkurt 2016; Tengilimođlu ve Kuzucu 2019). Çalışmanın ilk aşaması olan turizm gelirini etkileyen parametrelerin belirlenmesinde açıklayıcı değişkenler olarak ekonomi, turizm kapasitesi, özgürlük, güvenlik ve ülke gelişmişliđi olmak üzere 4 grupta adlandırılabilcek 20 adet değişken ile çalışılmış ve önemli değişkenler belirlenirken ekonometrik yöntemlerden yararlanılmıştır. İkinci bir farklılaşma noktası ise bulunan bağımsız değişkenlerle bağımlı değişkenler arasındaki ilişkinin performansının ölçümü konusundadır. Diğer çalışmalarda turizmin gelecek tahminlemesi konusunda genel olarak Yapay Sinir Ağları (YSA) gibi makine öğrenmesi metotları uygulanırken açıklayıcı değişken olmadan sadece zaman serisi tahminleri kullanılmaktadır (Çuhadar ve diğ. 2009; Çuhadar 2013; Koutras ve diğ. 2016; Keskin 2019; Höpken ve diğ. 2020; Çuhadar 2020). Bu çalışmada turizm tahmini literatüründe çok sık göremediğimiz fakat klasik regresyon ağaçlarından daha iyi sonuçlar verebilen Gradyan Arttırmalı Regresyon Ağaçları (GBRT) modelleriyle çoklu doğrusal regresyon yöntemi ve YSA yöntemleri açıklayıcı değişkenler yardımıyla modellenerek karşılaştırılmıştır.

Çalışmanın ilk bölümünde dünyada ve Türkiye’de turizmi etkileyen faktörlerin belirlenmesi ve turizm talebinin tahmin edilmesi üzerine yapılan ekonometrik modelleri, makine öğrenmesi modellerini ve klasik istatistik modellerini içeren çalışmalar literatür araştırması adına incelenmiştir. İkinci bölümde, dünyada ve Türkiye’de turizmin gelişimi, çok boyutlu etkileri ve ekonomik önemi üzerinde durulup turizmi etkileyen faktörler tartışılmaktadır. Üçüncü bölümde, uygulamada kullanılan yöntemler genel olarak ekonometrik analizler, çoklu doğrusal regresyon, YSA ve GBRT modelleri çerçevesinde açıklanmaktadır. Dördüncü bölümde, belirlenen veri setiyle bahsedilen yöntemler kullanılarak turizm gelirinin modellenmesi adına bir uygulama gerçekleştirilmekte ve model sonuçları karşılaştırmalı olarak açıklanmaktadır. Son bölümde ise sonuçlar tartışılıp çalışma önerileriyle sonlandırılmaktadır.

2. LİTERATÜR TARAMASI

Bu bölümde turizm çıktısı olarak turist sayısı, turizm geliri, konaklama doluluk oranı vb. ölçütler ile önceki bölümde belirtilen ekonomi, turizm kapasitesi, güvenlik vb. seçili etkenlerin arasındaki ilişkiyi inceleyen ekonometrik modeller, turizm ile ilgili tahminlemelerin ve modellemelerin araştırıldığı regresyon ve Yapay Sinir Ağları (YSA) yöntemleri ve çeşitli konularda performansı ölçülmüş Gradyan Arttırımlı Regresyon Ağaçları (GBRT) modelleri ile ilgili çalışmalar ulusal ve uluslararası ölçekte kronolojik sırayla verilmiştir. GBRT yönteminde özellikle ulusal çapta turizmle ilgili yapılan direkt bir çalışmaya rastlanamadığından benzer yöntemler ve farklı sektörlerdeki tahmin çalışmaları incelenmeye çalışılmıştır.

2.1 Türkiye’de Yapılan Çalışmalar

Turanlı ve Güneren (2003), turist sayısını tahmin edebilmek için 1983-2001 verilerini kullanarak çoklu logaritmik regresyon yöntemini uygulamıştır. Bağımsız değişken olarak terör ölüm sayısı, tesis yatak kapasitesi, reklam ve tanıtım bütçesi ve devalüasyon (kukla değişken) değişkenlerini kullanmıştır. Sonuç olarak kurulan modelin 0,97 R² değerine sahip olduğu gözlemlenmiştir.

Aktaş (2005), 1980-2000 yıllarında turizm gelirini çoklu doğrusal regresyon modeliyle açıklamaya çalışmıştır. Bu modelde bağımsız değişken olarak gelen turist sayısı, döviz kurları, dış yatırım, turizm tanıtım bütçeleri, acente sayısı ve tesis yatak sayıları ele alınmıştır. Çalışmanın sonucunda turist sayısı ve turizm seyahat acentelerinin sayısı turizm geliri için önemli açıklayıcı değişkenler olarak bulunmuştur.

Çuhadar ve diğ. (2009), 1992-2005 yılları arası Antalya’nın aylık turizm ziyaretçi sayısını üstel düzleştirme, Box-Jenkins ve YSA modellemesiyle karşılaştırmıştır. Karşılaştırmada en iyi değer %7,16 hata oranı ile 12 veri gecikme değerine ve sırasıyla katmanlarda [5:3:1] nöron sayısına sahip YSA modeli olduğu görülmüştür.

Çuhadar (2013), 1987-2012 tarihlerinde aylık olarak gelen yabancı ziyaretçi sayısı verilerini modellemek için sinir ağı modellerinden çok katmanlı yapay sinir ağı, radyal tabanlı model ve zaman gecikmeli yapay sinir ağı yöntemlerinden faydalanmıştır. Karşılaştırmada en iyi modelin 12 veri gecikme değerine ve sırasıyla katmanlarda [4:5:1] nöron sayısına sahip YSA modeli olduğu görülmüştür.

Aydın ve diğ. (2015), 1996-2013 dönemlerinde çeyrek yıllık verilerle Türkiye'ye en çok ziyaretçi gönderen ülkelerin ziyaretçi sayısını etkileyen ekonomik değişkenlerini belirlemek için bir panel veri analizi yöntemi uygulamıştır. Bu analizin sonucunda turizm giderlerinin artmasının ve ziyaretçi ülkelerdeki gelir seviyesinin yükselmesinin turizm talebinin düşmesine neden olduğu görülmüştür. Bu noktada çalışmadan Türkiye'deki turist profilinin az harcama yapan ve geliri arttıkça daha farklı ve yüksek maliyetli yerler seçme eğiliminde olan bir kitle olduğu çıkarımı yapılmıştır.

Pekmezci ve Bozkurt (2016) tarafından yapılan çalışmada 2005-2015 aylık dönemlerde turizm gelirleri ile euro ve dolar kuru arasındaki ilişki Genişletilmiş Dickey Fuller (ADF) birim kök testi, Johansen eşbütünleşme analizi ve Granger nedensellik analizi ile araştırılmıştır. Araştırma sonucu dolar kuru ile turizm geliri arasında herhangi bir ilişki bulunamamasına karşın, turizm gelirlerinin euro kurunu tek yönlü etkilediği bulunmuştur.

Çelik Uğuz (2017) tarafından yapılan çalışmada 2000-2015 yıllarında 15 şehir için yerli-yabancı turist sayıları bakımından suç verileriyle panel nedensellik analizi yapılmıştır. Pairwise Granger nedensellik analizi esas alınarak yapılan bu çalışmada suçun yabancı turizm talebini etkilemediği yönünde bir çıkarım yapılmıştır.

Bayram (2018), dönem olarak seçtiği 2006-2017 yılları arasında 24 ülke için Türkiye'nin turizm ziyaretçi sayısını fiyat endeksleri, ulaşım maliyeti, terörizm, alternatif destinasyonlar gibi değişkenlerle bir regresyon metodu olan LSDV modeli aracılığıyla tahmin etmiştir. Bu model farklı yaş, meslek, gelir düzeyi, geliş amacı gibi turist profilleriyle karşılaştırılmıştır. Terörün ise genel olarak turizmi %14 oranında etkilediği görülmüştür.

Tengilimođlu ve Kuzucu (2019), 2003-2018 yılları arasında eyrek dnemlik verilerle kiři baři turizm harcaması ve dolar kurunun arasındaki iliřiği Granger nedensellik testi ve regresyon modelleriyle aıklamaya alıřmıřtır. Analizler sonucunda dolar kurunun kiři baři turizm harcamasını negatif olarak %19 oranında etkilediđi bulunmuřtur.

Keskin (2019), 1984-2014 dnemlerinde grnrde iliřkisiz regresyon yntemiyle Trkiye'ye en ok turist gnderen lkelerin en az bir gece konaklayan ziyareti sayılarını incelemiřtir. Geleneksel iktisadi talep teorisine gre kurulan modelde farklı olarak ok denklemliler de kurulmuřtur. Uygulamada rakip turizm lkelerinin fiyat endeksi, ziyareti lkelerin kiři baři geliri, Trkiye'nin fiyat endeksi gibi aıklayıcı deđiřkenlerle alıřılmıřtır. Trkiye'ye turizm olarak rakip lkelerin fiyat endekslerinin artmasının Trkiye'nin genel olarak turizm talebini arttırıcı sonular dođurduđu, gelir seviyesi yksek olan lkelerin lkemize turizm talebinin rakip lkelere gre daha az olduđu sonuları bulunmuřtur.

Deniz (2019), 1969-2017 yılları iin aylık olarak turizm gelirlerini ok katmanlı bir yapay sinir ađı modeli erevesinde tahminlemeye alıřmıřtır. Tahminlemede zaman serisinde 12 veri gecikme deđeriyle yaklařık %20 hata oranında bir sonu elde edilmiřtir.

uhadar (2020), 2003-2020 arası aylık dıř turizm gelirlerini stel dzleřtirme, Box-Jenkins ve YSA modelleriyle karřılařtırmıřtır. Karřılařtırma sonucunda %6,51 Ortalama Mutlak Yzde Hata (MAPE) deđeriyle zaman serisinde 12 veri gecikme deđerine sahip ve katmanlardaki nron sayısı sırasıyla [5:7:1] olan YSA modelinin en iyi model olduđu sonucu ıkarılmıřtır.

2.2 Dnyada Yapılan alıřmalar

Uysal ve El Roubi (1999), 1980-1994 yılları arası eyrek dnemlerde Amerika'daki Kanadalı turist harcamalarını oklu regresyon ve Yapay Sinir Ađı (YSA) modelleriyle incelemiřtir. alıřma sırasında dviz kuru farkı, Kanada'nın reel kiři başına dřen geliri, lkeler arası tketiciler fiyat endeksi oranı kullanılmıřtır.

Çalışmada çok yakın hata kriterleri bulunmasına karşın YSA'nın %1'lik bir farkla daha iyi bir performans gösterdiği bulunmuştur.

Payne ve Mervar (2002), 1993-1999 yılları arası çeyreklik dönemlerde Hırvatistan'ın turizm gelirlerine yönelik bir regresyon modeli kurmuştur. Modelde Gayri Safi Yurtiçi Hasıla (GSYİH), reel efektif döviz kuru ve çeşitli kriz ve askeri hareket durumlarıyla ilgili kurulan kukla değişkenleri bulunmaktadır. Kurulan modelde R² değeri yaklaşık 0,95 olarak elde edilmiştir.

Dritsakis (2004), 1960-2000 yılları arasında İngiltere ve Almanya'dan Yunanistan'a yapılan turist gelişlerinin makroekonomik değişkenlerle ilişkisini araştırmıştır. Araştırmada Genişletilmiş Dickey Fuller (ADF) birim kök testi ve VAR eşbütünleşme analizini kullanan Dritsakis gelir düzeyleri, ulaşım maliyeti ve döviz kurunun turist gelişleriyle ilişkili olduğunu bulmuştur. Fakat bulunan bu ilişkinin yönüyle ilgili bir çalışma yapılmadığının belirtilmesi gerekmektedir.

Yin ve diğ. (2012), Washington D.C. bölgesinde 2011-2012 yılları arasında bisiklet paylaşım sistemindeki bisiklet kullanım sayılarını saatlik olarak tahmin etmek için Ridge doğrusal regresyon, destek vektör regresyonu, rassal orman ve gradyan arttırmalı ağaçlar modellerini kullanılmıştır. En düşük hata oranının %31 oranla grandyan arttırmalı ağaçlar ve rassal orman yöntemleri olduğu görülmüştür.

Raza ve Jawaid (2013), 1980-2010 yılları arası Pakistan için terör ve turizm talebi ilişkisini araştırmıştır. Yapılan araştırmada Johansen ve Jeuselius, ARDL gibi eşbütünleşme testleri ile Granger, Todo Yamamoto gibi nedensellik testleri yapılmıştır. Bu testlerin sonucunda terör olaylarının turizmi uzun dönemli olarak etkilediği ve bu ilişkinin tek yönlü bir ilişki olduğu, turizm gelirlerinin artması için terör olaylarının kontrol altına alınması gerektiği çıkarımları yapılmıştır.

Koutras ve diğ. (2016), 2005-2012 yılları arası Batı Yunanistan'ın aylık tesis konaklama doluluğu ve geceleme sayılarını modellemiştir. Modellemede doğrusal regresyon, YSA ve destek vektör makineleri kullanılarak karşılaştırmalar yapılmıştır. Sonuç olarak doluluk oranı için YSA'nın 15 aydan daha uzun süreli tahminlerde daha iyi olduğu görülmüştür.

Khan ve diğ. (2017), 1990-2014 yılları arasında uluslararası turizm gelişlerinin ve gidişlerinin, turizm gelirlerinin ve giderlerinin ulaşım gelir ve giderleriyle yolcu taşımacılığıyla ilişkisini incelemiştir. FMOLS regresyon yönteminin kullanıldığı çalışmada nedensellik çalışmaları üzerinde de durulmuş ve Granger nedensellik testleri yapılmıştır. İnceleme sonucunda hava yolu taşımacılığının turizm ve ticaret açıklığı üzerinde olumlu etkilerinin olduğu görülmüştür.

Fareed ve diğ. (2018), 1990-2017 yılları arası Tayland'da turist sayısı, global terör endeksi ve kişi başı GSYİH arasındaki eşbütünleşme ilişkisini ARDL yöntemiyle araştırmıştır. Yapılan eşbütünleşme analizi sonucunda terör, turist sayısı ve GSYİH değişkenlerinin ilişkili olduğu sonucu ortaya çıkarılmıştır. Nedensellik analizinin yapılmaması nedeniyle bu uzun dönemli ilişkilerin yönü belirtilmemektedir.

Kang ve Gu (2019) yaptıkları çalışmada Çin'in Guilin kenti için 2015-2018 yılları arası aylık turist akışını modellemiştir. Tahminlemede arttırma yöntemiyle çalışan bir regresyon ağacı algoritması olan MART modeli kullanılmıştır. Tahmin sonucunda kurulan modelin Ortalama Mutlak Yüzde Hata (MAPE) oranı %7,38 olarak bulunmuştur.

Abdualgalil ve Abraham (2020) yaptıkları çalışmada destek vektör makinesi, naive bayes, regresyon modelleri, rassal orman, karar ağaçları, en yakın komşuluk, gradyan arttırma gibi makine öğrenmesi modelleriyle yıllık ve ilk çeyreklik veriler için Hindistan'ın turist gelişlerini hesaplamıştır. Yapılan çalışmada bir gradyan arttırma modeli olan XGBoost modeliyle yaklaşık %88-90 doğruluk payında bir model kurulmuştur. En iyi model %90-96 doğruluk payıyla destek vektör makinesi olarak bulunurken regresyon modelleri ise yaklaşık %87-89 oranında bir başarı göstermiştir.

Höpken ve diğ. (2020), İsveç için 2008-2016 yılları arasında aylık turist gelişlerini ARIMA ve Yapay Sinir Ağları (YSA) modellerini karşılaştırarak tahmin etmeye çalışmıştır. Çalışmada otoregresif veri analizinin yanı sıra Google Trends büyük verileri de kullanılmıştır. Sonuç olarak geri yayımlı çok katmanlı algılayıcı ve tek gizli katman kullanan YSA Google Trends gibi büyük verilerde ARIMA modeline göre ortalamada %4,2 hata oranı farkla daha iyi sonuç vermiştir.

2.3 Literatür Özeti

Literatür taraması sonucunda yapılan çalışmalarda genellikle ziyaretçi sayıları üzerinde durulduğu, bağımsız değişkenler belirlenirken makroekonomik değişkenlere ağırlık verildiği ve Yapay Sınır Ağları (YSA) model karşılaştırmalarının çoğunlukla bağımsız değişken kullanılmadan zaman serisi yöntemleriyle karşılaştırıldığı gözlemlenmektedir. Literatür çalışmalarının daha kolay karşılaştırılabilmesi adına kritik başlıklarla çalışmaların özeti aşağıda Tablo 2.1 üzerinde gösterilmektedir.

Tablo 2.1: Literatür karşılaştırma tablosu.

Yazar	Dönem	Ülke/Bölge	Bağımlı Değişken	Bağımsız Değişken	Yöntem
Uysal ve El Roubi (1999)	1980-1994	Amerika	Turizm geliri	Kur farkı, gelir düzeyi, fiyat endeksi, gecikmeli seriler	Regresyon ve YSA
Payne ve Mervar (2002)	1993-1999	Hrvatistan	Turizm geliri	GSYİH, döviz kuru, kriz ve askeri hareket	Regresyon
Turanlı ve Güneren (2003)	1983-2001	Türkiye	Ziyaretçi sayısı	Terör, yatak kapasitesi, reklam bütçesi, devleşyon	Regresyon
Dritsakis (2004)	1960-2000	Yunanistan	Ziyaretçi sayısı	Gelir düzeyi, ulaşım maliyeti, döviz kuru	VAR eşbütünleşme
Aktaş (2005)	1980-2000	Türkiye	Turizm geliri	Ziyaretçi sayısı, döviz kuru, dış yatırım, reklam bütçesi, acente sayısı, yatak kapasitesi	Regresyon
Çuhadar ve diğ. (2009)	1992-2005	Antalya	Ziyaretçi sayısı	-	YSA, ARIMA ve üstel düzleştirme
Çuhadar (2013)	1987-2012	Türkiye	Ziyaretçi sayısı	-	Çok katmanlı, radyal tabanlı, zaman gecikmeli YSA
Raza ve Jawaid (2013)	1980-2010	Pakistan	Turizm geliri	Terör	Johansen, ARDL eşbütünleşme ve Granger, Todo Yamamoto nedensellik
Aydın ve diğ. (2015)	1996-2013	Türkiye	Ziyaretçi sayısı	Turizm gideri, gelir seviyesi	Panel nedensellik
Koutras ve diğ. (2016)	2005-2012	Batı Yunanistan	Tesis doluluğu ve geceleme sayısı	-	Regresyon, YSA ve destek vektör makineleri
Pekmezci ve Bozkurt (2016)	2005-2015	Türkiye	Turizm geliri	Döviz kuru	Johansen eşbütünleşme, Granger nedensellik
Çelik Uğuz (2017)	2000-2015	15 il (Türkiye)	Ziyaretçi sayısı	Suç oranları	Pairwise Granger nedensellik
Khan ve diğ. (2017)	1990-2014	Küresel	Ziyaretçi sayısı, turizm geliri	Ulaşım verileri	Regresyon, Granger nedensellik
Bayram (2018)	2006-2017	Türkiye	Ziyaretçi sayısı	Fiyat endeksi, ulaşım maliyeti, terör	Regresyon
Fareed ve diğ. (2018)	1990-2017	Tayland	Ziyaretçi sayısı	GSYİH, terör	ARDL eşbütünleşme
Deniz (2019)	1969-2017	Türkiye	Turizm geliri	-	YSA
Kang ve Gu (2019)	2015-2018	Guilin (Çin)	Ziyaretçi sayısı	-	Regresyon ağacı
Keskin (2019)	1984-2014	Türkiye	Ziyaretçi sayısı	Fiyat endeksi, kişi başı gelir	Regresyon
Tengilimoğlu ve Kuzucu (2019)	2003-2018	Türkiye	Kişi başı turizm geliri	Döviz kuru	Regresyon, Granger nedensellik
Çuhadar (2020)	2003-2020	Türkiye	Turizm geliri	-	YSA, ARIMA ve üstel düzleştirme
Höpken ve diğ. (2020)	2008-2016	İsveç	Ziyaretçi sayısı	-	YSA, ARIMA

Bu çalışma Tablo 2.1 üzerindeki çalışmalar da göz önünde bulundurulduğunda çeşitli noktalarda literatürden farklılaşmaktadır. Literatürde genellikle sınırlı sayıda değişkenle ziyaretçi sayısının açıklanması üzerinde durulmaktadır. Bu çalışmada ise makroekonomik değişkenlerin yanında ülke güvenliği-özgürlüğü, gelişmişliği ve turizm kapasitesi başlıklarında 20 adet değişken turizm gelirini açıklamak üzere karşılaştırılmıştır.

Bunun dışında çalışmanın literatürden en önemli farklarından bir tanesi de kullanılan yöntemlerdir. Literatürde görülen çalışmalarda ya belirli değişkenlerin regresyon veya ekonometrik yöntemlerle turizm verileri arasındaki ilişkileri incelenmiş ya da genellikle zaman serisi yöntemleriyle Yapay Sinir Ağları (YSA) metodu karşılaştırılmıştır. Bu çalışmada iki çalışma prensibi birleştirilerek hem önemli değişkenler tespit edilmiş hem de bu değişkenlerle kurulan modellerin performansı değerlendirilmiştir. Kullanılan yöntemlerde ise diğer çalışmalarda seyrek olarak görülen yapısal kırılmaya izin veren durağanlık ve eşbütünleşme testleriyle regresyon yöntemi desteklenmiştir. Ayrıca yine literatürden farklı olarak makine öğrenmesi yöntemlerinde genellikle kullanılan YSA ve destek vektör makinesi gibi yöntemlerin dışında Gradyan Arttırımlı Regresyon Ağaçları (GBRT) yöntemi de kullanılmıştır.

3. DÜNYADA VE TÜRKİYE'DE TURİZM SEKTÖRÜ

3.1 Küreselleşmenin Turizm Üzerindeki Etkisi

1900'lü yılların sonları itibariyle dünyada önemli bir dönüşüm süreci yaşanmıştır. Küreselleşme olarak tanımlanan bahsedilen dönüşüm süreci ekonomik, politik ve toplumsal alan üzerinde önemli etkiler oluşturmuştur (Çeken 2004). Küreselleşme, özellikle 1980'li yıllarla birlikte önemli bir tartışma alanı olarak görünür hale gelmeye başlamıştır. Bir tartışma alanı olmasının ötesinde gerek bireyler gerek toplumlar gerekse ülkeler üzerindeki somut etkileriyle birlikte temel bir gerçeklik olarak değerlendirilmeye başlanmıştır. Bu doğrultuda küreselleşmenin etkilerinin yaşam tarzları, üretim sistemleri, ekonomik faaliyetler ve politik ilişkiler üzerindeki etkisi karmaşık ve iç içe geçmiş bir görünümü ortaya çıkarmıştır. Ekonomik, politik ve sosyokültürel olguların ve faaliyetlerin liberalleşmesi ile mekân ve mesafe gibi olguların ortadan kalkması küreselleşmenin karakteristik özellikleri olarak dikkat çekmiştir. Dünyanın küresel bir köy haline gelmeye başladığına dair görüşler sıklıkla dile getirilmiştir (Bauman 2006). Küreselleşmenin belirli bir boyutu olarak küresel ekonomi, başta uluslararası ticaretin serbestleştirilmesi, rekabet, sermayenin ve iş gücünün serbest dolaşımı, yatırımlar vb. alanlarında çeşitli değişiklikleri tetiklemiştir (Cerovic ve diğ. 2015). Bu doğrultuda küreselleşme özellikle iktisadi faaliyetler alanında dünyada bütün sektörler üzerinde etkili hale gelirken, turizm sektörü küreselleşmenin etkilediği önemli alanlar içerisinde yer almıştır.

Küreselleşme etkisi bağlamında turizm faaliyetlerinin mümkün bir fırsat olarak bireyler tarafından gerçekleştirilme sıklığı belki de ilk ele alınabilecek konular arasındadır. Ulaşım ve seyahat imkânlarının artması, teknolojik gelişmeler, refah seviyesindeki artış beraberinde turizmi, belirli imkânlara sahip kişiler/gruplar tarafından gerçekleştirilen bir faaliyet olmasının ötesinde çok daha geniş kesimlerce gerçekleştirilebilen bir faaliyet alanı haline getirmiştir. Benzer şekilde küreselleşme tartışmaları beraberinde ülkeler arasında dolaşım imkânlarının artması da turizme yönelik hareketliliğin gelişmesine ve turizmin bir sektör olarak büyümesine yol açmıştır (Civelek 2020). Bu doğrultuda ülkeler arasındaki coğrafi sınırların

kalkmasının yanında, uluslararası yatırımların yaygınlaşması, kültürler arasındaki farklılıkların azalması, ortak iletişim kanallarının ve kullanılan ortak dillerin artması gibi unsurlar da turizm faaliyetlerinin gelişmesinde önemli bir etkiye sahip olmuştur. Bahsedilen gelişmeler uluslararası turizm gelirlerinin ve turist sayılarının artmasını sağlamıştır (Çeken ve diğ. 2009). Küreselleşme beraberinde şekillenen rekabet koşullarının turizm sektörü üzerindeki etkisi de önemli bir yere sahiptir. Ayrıca küreselleşme turizm pazarının gelişmesi ve genişlemesi adına yeni olanakların ortaya çıkmasını da sağlamıştır (Cerovic ve diğ. 2015).

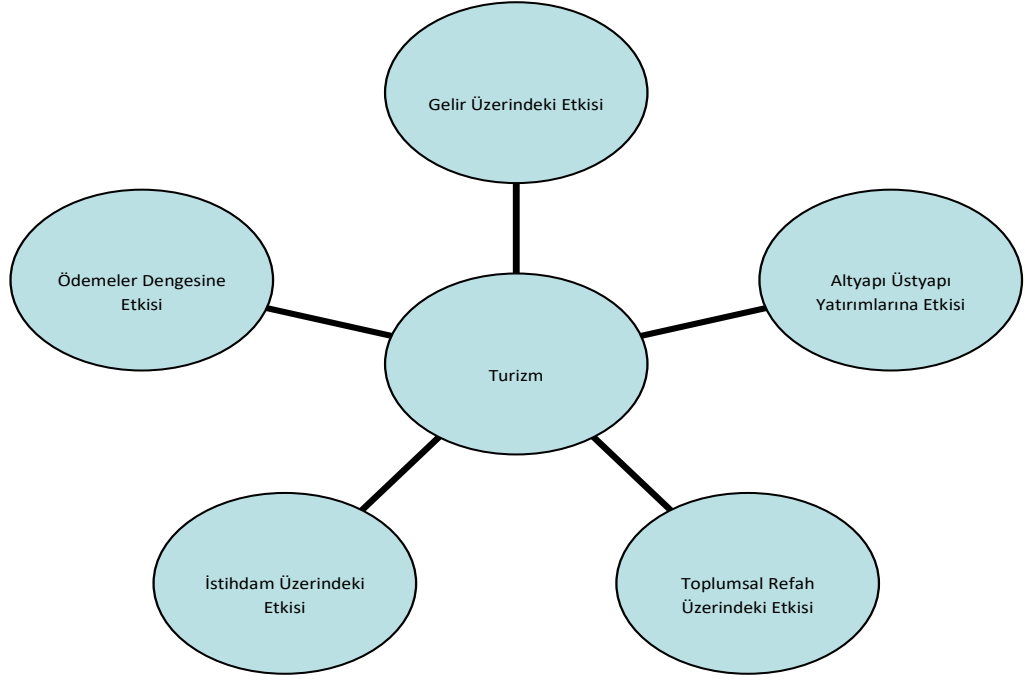
Küreselleşmenin yukarıda tartışılan boyutları paralelinde turizm 2000’li yıllar itibariyle küresel ekonomi içerisinde önemli bir paya sahip olmuştur. Bahsedilen dönemlerde turizm en hızlı büyüyen sektörler arasında yer almıştır. Turizm sektöründe yaşanan büyüme ilgili yıllarda mal ve hizmet üretiminde otomobil ihracatını da geçerek birinci sıraya yerleşmiştir. 2000’li yıllarda seyahat ve turizm dünya ekonomisine en büyük katkıyı sağlayan alanlar içinde yer almıştır. Daha önce de belirtildiği gibi bahsedilen dönemde dünyadaki turizm geliri ve turist sayısında gözle görülür bir artış yaşanmıştır (Çeken 2004). Bahsedilen gelişmeler küreselleşmenin turizm sektörü üzerindeki somut etkisi bağlamında değerlendirilmiştir.

Gerçekten de küresel turizm 2000’li yıllarda dile getirilen gelecek tahminleriyle uyumlu bir şekilde son yıllarda önemli ölçüde büyümüş, küreselleşmenin yükselişinden ve seyahatin kolaylaşmasını sağlayan teknolojik ilerlemelerden yararlanmışır. Ayrıca insanların kendi seyahatlerini planlamalarını ve rezervasyon yapmalarını, deneyimlerini gerçek zamanlı olarak arkadaşlarıyla paylaşmalarını kolaylaştıran uygulamalar da bu süreçte etkili olmuştur. Bahsedilen gelişmeler beraberinde küresel turizm istikrarlı bir şekilde büyümüştür. Bu büyüme neticesinde 2016 yılında küresel turizmin dünya ekonomisindeki payı tahmini 1,2 milyar dolar olmuştur. 2018 verileri dikkate alındığında ise OECD ülkelerinde turizmin, ortalama olarak GSYİH'nın %4,2'sini, istihdamın %6,9'unu ve hizmet ihracatının %21,7'sini oluşturduğu görülmektedir (OECD 2018). Küresel seyahat harcamaları ise 2000 ile 2016 arasında ikiye katlanarak 495 milyar dolardan 1,2 trilyon dolara yükselmiş ve mal ve hizmetler alanında küresel ihracatın %7'sini oluşturmuştur (UNWTO 2017). Belirtmek gerekir ki bahsedilen bu büyümenin devam edeceği öngörülürken, değişen çevresel koşullar, teknolojik yenilikler ve değişen demografik

özelliklerin, 2040 yılına kadar turizmin çehresini dramatik bir şekilde değiştireceği tahmin edilmektedir. İlerleyen dönemlerde gelişmekte olan ekonomi destinasyonlarına gelen uluslararası turistlerin, gelişmiş turizm ekonomilerindekinin iki katı oranında büyüyeceği öngörülürken, 2030 yılına kadar dünyadaki toplam turizm gelirlerinin 1,8 milyar dolara yükselmesi beklenmektedir (OECD 2018).

3.2 Turizmin Ekonomi Üzerindeki Etkisi

Ekonomi ve turizm arasındaki ilişki değerlendirilirken birçok farklı unsurun göz önünde bulundurulması gerekmektedir. Bu doğrultuda turistlerin turizm amaçlı seyahatleri sırasında transferleri, yolculukları ve bu süreç içerisinde gerçekleştirdikleri harcamalar nedeniyle; hava, kara ve deniz taşımacılığı yapan şirketler ile araç kiralama şirketleri, turizm acenteleri, vb. şirketler süreçte önemli oyuncular olarak karşımıza çıkabilmektedir. Bununla birlikte konaklama ve yeme içme alanında faaliyet gösteren işletmelerle, turistlerin ihtiyaçlarını karşılayan birçok sektör turizm faaliyetlerinden doğrudan veya dolaylı olarak etkilenebilmektedir (Bahar 2013). Kısacası turizm birçok farklı yönden ekonomik büyüme üzerinde önemli bir etkiye sahiptir. Turizm, Gayri Safi Yurtiçi Hasılayı (GSYİH) desteklemesi, istihdam yaratması ve döviz akışında önemli bir paya sahip olmasının yanında diğer ekonomik faaliyetleri tamamlayıcı bir özelliğe de sahiptir. Ayrıca belirtmek gerekir ki turizm sektörünün katkısı sadece ekonomik büyüme ile sınırlı da değildir. Turizm, ekonomik ve sosyokültürel ilerlemeyi etkilemekte ve toplumsal refahın iyileşmesine de katkı sunmaktadır (Cardenas-Garcia ve Pulido-Fernandes, 2019; Bento 2016). Kısacası ödemeler dengesi, gelir etkisi, ithalat etkisi, istihdam etkisi, ihracat ve gelir dağılımı etkisi, döviz geliri üzerindeki etkisi, küçük ölçekli işletmelerin geliştirilmesi, toplumsal refah genel itibariyle turizmin ekonomik etkileri arasında sayılabilecek alanlar arasındadır. Aşağıda Şekil 3.1 ile turizmin bahsedilen ekonomik etkileri genel olarak gösterilmektedir.



Şekil 3.1: Turizmin ekonomik etkileri.

Turizmin döviz geliri ile ödemeler dengesi üzerindeki etkisi arasındaki ilişki birbiri ile bağlantılı sonuçları içinde barındıran bir özelliğe sahiptir. Özellikle döviz arzı ve talebi üzerindeki etkisi bağlamında turizm, ödemeler dengesini düzenlemekte, ödemeler dengesinin açık verdiği durumlarda önemli bir döviz kaynağı olarak dikkat çekmektedir. Bununla birlikte turizm bir döviz geliri kaynağı olması nedeniyle ekonomik gelişme üzerinde de önemli bir etki oluşturmaktadır. Kısacası turizm sayesinde doğrudan veya dolaylı olarak ülkeler için önemli bir gelir kaynağı meydana gelmektedir (Arabacı 2018). Ortaya konulan tablo bir ülkenin turizm gelir ve giderlerinde farklı ülkelerle olan ilişkisini göstermesi münasebetiyle dış turizm bilançosunu ortaya koymaktadır. Bu noktada turizm döviz gelirleri ile ödemeler dengesi arasındaki sıkı bağlantının netleştirilmesi son derece önemlidir. Bir ülkenin turizminin ödemeler bilançosuna olumlu bir etkide bulunabilmesi için o ülkeye giren turizm döviz gelirinin ilgili ülkeden turizm adına çıkan döviz giderinden fazla olması gerekmektedir. Dolayısıyla bu kapsamda turizmin ödemeler dengesi üzerindeki olumlu etkisinden söz edebilmek ancak o ülkenin döviz gelirlerinin döviz kaybından büyük olduğu durumlarda söz konusudur (Kozak 2013).

İlgili başlık altında turizmin sadece döviz geliri ve ödemeler dengesi üzerindeki etkisine değil bununla birlikte ekonomi üzerindeki farklı etkileri üzerinde de

durulmalıdır. Bu kapsamda turizmin ihracat üzerindeki dolaylı veya doğrudan etkisi belki de ilk üzerinde durulması gereken hususlar arasındadır. Turistler tarafından ziyaret edilen ülkelerde o ülkeye özgü ürünlere olan talep beraberinde ilgili ürünleri kimi zaman mevcut ihracat fiyatlarının da üzerinde farklı ülkelere pazarlanmasını beraberinde getirmektedir. Bu durum özellikle farklı gelişmişlik seviyesine sahip bölgeler arasındaki turist akışı da göz önünde bulundurulduğunda, turizm gelir dağılımının düzenlenmesi anlamında da dengeleyici bir rol üstlenmektedir. Bahsedilen bu hareketlilik beraberinde turizmin en önemli olumlu etkilerinden birisi olan gelirin yaratılması durumunu da ortaya koymaktadır. Bu doğrultuda turizm harcamaları beraberinde belirli bir ülkede veya bölgede yaşayan kişilerin gelirlerinin oluşmasını sağlamaktadır. Dünya ekonomisi bağlamında bir değerlendirme yapıldığında ise turizmin ülke ekonomilerinin en temel gelir kaynakları arasında yer aldığı belirtilmektedir. Benzer bir şekilde turizm sektörü ülkeler açısından ihracat üzerindeki etkisi noktasında en büyük ilk beş sektör arasında telaffuz edilmektedir. Turizmin gelir üzerindeki etkisi sadece turizm sektörüyle bağlantılı olan alanlar üzerinden değil diğer bağlantılı sektörler noktasında da ele alınmalıdır. Turizm, ilişkili olduğu birçok sektör içerisinde de iktisadi bir hareketliliğe yol açmaktadır. Dolayısıyla elde edilen turizm geliri bahsedilen bu sektörlerde de pay sahipleri aracılığıyla değerlendirilmekte ve oluşan katma değer farklı şekillerde de el değiştirmektedir. Bu durum ise turizmin gelir ve harcama üzerindeki çarpan etkisini görünür hale getirmektedir (Kozak 2013).

Turizmin ekonomi üzerindeki doğrudan etkilerinin yanında istihdam ve refah seviyesinin yükselmesi üzerindeki etkisi de dikkate alınmalıdır. Bu kapsamda turist sayısındaki artış ve beraberinde turizm gelirindeki yükselme ilgili alandaki sektörel hacmi genişletmekte bu ise ilgili sektördeki istihdam hacminin artmasını da beraberinde getirmektedir. Turizmin emek yoğun bir sektör olduğu dikkate alındığında istihdam üzerindeki etkisinin dünya ekonomisi açısından önemi açıkça görülmektedir. Böylece hem doğrudan turizm sektöründe hem de bağlantılı sektörlerde işgücüne olan ihtiyaç artmaktadır. Kısacası turizm bir ülkede işsizliğin azalması üzerinde de önemli etkilere sahiptir (Arabacı 2018). İstihdam üzerinde oluşturduğu etki itibariyle belki de turizmin özellikle genç istihdamı ve kadın istihdamı üzerindeki etkisi de dikkate alınmalıdır. İlgili tartışmalardan da anlaşılacağı üzere turizm, konaklama, transfer, yeme-içme, ulaşım, seyahat acenteleri vb. kuruluşların faaliyetleri beraberinde istihdam alanında doğrudan bir etki oluşturmakla birlikte, turizm sektörünün etkilediği

farklı alanlar sayesinde dolaylı bir etki de oluşturmaktadır. Bunun yanında turizm gelirlerinin bölge ve diğer endüstriler üzerinde oluşturduğu olumlu etki beraberinde bu alanlarda yeni yatırımları iktisadi olarak teşvik etmekte bahsedilen durum ise uyarılmış istihdam üzerinde etkili hale gelebilmektedir (Paksoy ve diğ. 2018). Ayrıca turizmin geliştiği ve turizm yatırımlarının arttığı bölgelerde ulaşım ve altyapı yatırımlarının da hızla geliştiği ve arttığı görülmektedir. Kısacası turizm yoğun gelir sağladığı bölgelerde ekonomi ve istihdam üzerindeki etkileriyle bağlantılı olarak toplumsal refahın gelişmesi noktasında da önemli katkılar sunmaktadır (Arabacı 2018).

3.3 Türkiye Açısından Turizm Geliri

Turizm sektörünün Türkiye’deki gelişimi de küreselleşme süreciyle uyumlu bir şekilde seyretmiştir. Bu doğrultuda turizm sektörü Türkiye’de özellikle 1980’li yıllar sonrasında gelişmeye başlamıştır. 1980 yılı öncesinde turizm faaliyetlerinin Türkiye’de bugünkü durumundan çok uzak olduğunu belirtmek gerekir. 1980 öncesinde Türkiye’de turizmin toplam sabit yatırımlar içindeki payının %1’den daha az ve Gayri Safi Milli Hasıla (GSMH) içindeki payının ise %0,8 civarında olduğu görülmektedir. Bahsedilen dönemde turizm sektöründeki konaklama kapasitesinin ve Türkiye’yi ziyaret eden turist sayısının da oldukça düşük olduğu dikkat çekmektedir. 1980-2000 yılları arasında Türkiye’nin turizm yatırımlarını dikkat çekici ölçüde arttırması sonrasında turizm gelirlerinin GSMH içindeki payının da arttığı görülmüştür. Bahsedilen süreç sonrasında turizm gelirleri Türkiye ekonomisi açısından da önemli bir yere sahip olmaya başlamıştır. 1980’li yıllardan sonra turizmde yaşanan bu gelişmeler üzerinde turizm yatırımlarını teşvik eden yasal düzenlemelerin de önemli katkısı olmuştur (Bilgiçli ve Altınkaynak 2016).

Yatırımlar ve yasal düzenlemeler yanında Türkiye’nin doğal güzellikleri, coğrafi konumunun avantajlı yapısı, Türkiye’nin sosyokültürel özellikleri de turizm faaliyetlerinin gelişmesi ve Türkiye’ye gelen turist sayısının artması üzerinde son derece etkili olmuştur. Bunun yanında uluslararası alanda turizm faaliyetlerinin artması, yukarıda küreselleşme süreci bağlamında değinilen gelişmeler de Türkiye turizmini önemli ölçüde etkilemiştir. Bu doğrultuda dünya genelinde turizm

faaliyetlerinin gelişmesi ve turist sayılarının artması diğer gelişmekte olan ülkelere benzer bir şekilde Türkiye'nin de bu pazardan pay alma eğilimini yükseltmiştir. Türkiye özellikle döviz geliri beklentisi, sektörden alınacak katma değer in sürekli büyümesi, alana yönelik yatırım maliyetlerinin düşüklüğü ve sektörün belirli tarife ve kotalardan daha az etkileniyor olması gibi nedenlerle alana yönelik ilgisini canlı tutmuştur. Bütün bu nedenler yanında Türkiye ekonomisinin hizmet sektörü odaklı gelişim yapısı ve bu doğrultuda turizmin bahsedilen eğilimle uyumluluğu ile işsizlikle mücadele kapsamında turizmin emek yoğun bir alan oluşturması da Türkiye'nin turizm faaliyetlerini ve yatırımlarını ön planda değerlendirmesinde etkili olmuştur (Çımat ve Bahar 2003).

Bahsedilen hedefler ve bu hedeflerle uyumlu stratejiler beraberinde özellikle Türkiye'ye gelen turist sayısında önemli artışlar yaşandığına dikkat çekmiştir. 2000'li yıllarında başından 2010'lu yıllara kadar turist sayısının üç kat arttığı gözlenmiştir. 2000 yılında 10 milyon 400 bin civarında olan turist sayısının 2012 yılında 30 milyon 700 binin üzerine çıktığı görülmüştür. Bununla birlikte bahsedilen yıllarda 2006 yılı haricinde turist sayısındaki düzenli ve istikrarlı artış dikkat çekmiştir. 2006 yılında yaşanan sapmanın ise Almanya'da gerçekleşen Dünya Kupası nedeniyle olabileceği belirtilmiştir. Ayrıca 2008 yılında yaşanan küresel finansal krize rağmen Türkiye'ye gelen turist sayısında artış yaşanmaya devam etmesi de ayrıca dikkat çekilen unsurlar arasında yer almıştır (Şen ve Şit 2015).

Yıllar itibariyle değerlendirildiği turist sayısında olduğu gibi turizm gelirlerinde de artış yaşandığı belirtilmelidir. Ancak bu artıştaki istikrarın turist sayısındaki gibi net olmadığı ve kimi dönemlerde turizm gelirindeki belirli dalgalanmaların turist sayısı verilerine kıyasla daha hissedilir olduğuna da dikkat çekilmelidir. Örneğin 2008 finansal krizi nedeniyle her ne kadar turist sayısında bir azalma söz konusu olmamışsa da bu krizin turizm gelirlerini önemli ölçüde etkilediği görülmüştür. Türkiye'de ise ziyaretçi sayısında önde gelen ülkelere karşı kişi başı harcamada ve turizm gelirinde aynı sıraları koruyamamıştır. Bununla birlikte özellikle 2009 yılı ile 2016 yılları arasında yaşanan salgınlar, politika çekişmeleri vb. olayların da hem turist sayısı hem de turizm gelirleri üzerinde etkili olduğu dikkat çekmiştir. Özellikle 2016 yılında turizm gelirinde ve turist sayısında ciddi bir düşüş yaşandığı dikkat çekmiştir. Bahsedilen yıl içerisinde Türkiye ile Rusya arasında

yaşanan siyasi gerginliğin bu düşüş üzerinde oldukça etkili olduğu belirtilmektedir (Çullu Kaygısız 2019).

Yukarıda ele alınan bütün bu etkiler paralelinde ve genel olarak bir değerlendirme yapıldığında ise Türkiye'nin turist sayısında dünya genelinde ortaya koyduğu başarıyı aynı oranda turizm gelirlerinde ortaya koymadığı üzerinde durulmalıdır. Türkiye'nin 2008 yılında ziyaretçi sayısında 8. sırada, turizm gelirlerinde ise 9. sırada yer aldığı dikkat çekmektedir. 2009 yılında ise turizm gelirlerinde dünyadaki yerini koruyan Türkiye turist sayısı bakımından 7. sıraya yükselmiştir. 2011 yılından 2015 yılına kadar turist sayısı bakımından dünyada 6. sırada olan Türkiye turizm gelirleri bakımından ise 12. sırada yer almıştır. 2016 yılında ise turizm gelirleri bakımından beş sıra daha gerileyen Türkiye 17. sırada yer almıştır. Genel itibariyle bir değerlendirme yapıldığında Türkiye turist sayısı bakımında çoğu zaman ilk onda yer almasına rağmen aynı başarıyı turizm gelirlerinde gösterememiş ve özellikle 2010 yılı sonrasında dünya genelinde ilk onda yer alamamıştır. Böylece Türkiye'yi ziyaret eden yabancı turistlerin turizm harcamalarının düşük olduğu görülmüştür. Türkiye'de deniz turizminin ve her şey dahil sisteminin yaygınlığı, turistlerin ortalama harcamalarını düşüren önemli nedenler arasında sayılabilmektedir. Deniz turizminin yanında farklı turizm çeşitlerinin tercih edilirliliği artırılmalı ve turizm 12 aya çıkarılmalıdır. Bunun yanında Türkiye'de turizm gelirlerinin artırılması adına ziyaretçi turist profilinin de değişmesi gerekmektedir. Bu doğrultuda gelir düzeyi yüksek turist profilinin Türkiye'yi tercih etmesini sağlayacak stratejiler geliştirilmelidir (Şen ve Şit 2015; Çullu Kaygısız 2019). TÜİK (2020) verilerine göre inceleme aralığı olan 2003-2019 dönemleri arasında ülkemize gelen yabancı ziyaretçilerin ortalama %65'i orta düzey gelir seviyesine sahip olmakla beraber 2017, 2018 ve 2019 yıllarında orta düzey gelir seviyesine sahip yabancı ziyaretçilerin oranı yıllık ortalama %76'dır. Böylece açıkça görülmektedir ki turist profilinin değiştirilmesi adına yapılacak çalışmalar da turizm geliri açısından son derece kritik önem arz etmektedir.

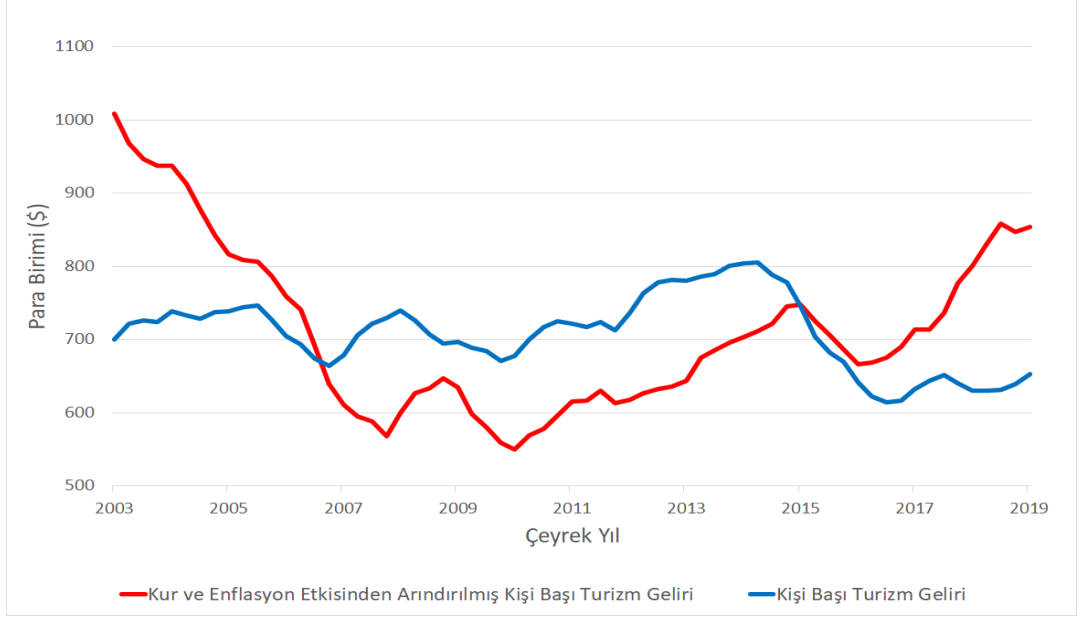
TÜİK (2020) verileri incelendiğinde, 2019 yılı için Türkiye'nin turizm gelirlerinin büyük bir bölümünün yeme içme ve ulaşım kalemlerine ait olduğu görülmektedir. Bu doğrultuda turizm gelirlerinin;

- %24'ü yeme-içme,

- %24'ü ülke içi ve uluslararası ulaşım,
- %14'ü giyim,
- %13'ü konaklama,
- %5'i hediyelik eşya,
- %4'ü sağlık,
- %16'sı diğer harcamalardan oluşmaktadır.

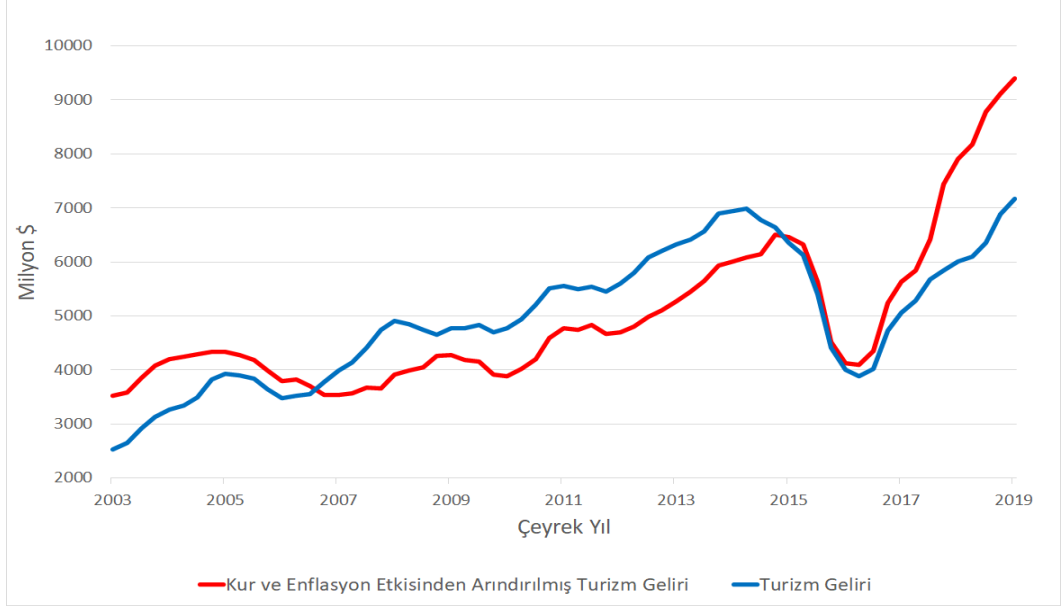
Yukarıda da görüldüğü gibi turizm gelirini oluşturan kalemlerin büyük çoğunluğu turistik faaliyetlerden ziyade, ziyaretler sırasında ulaşım, yeme-içme gibi gerekli ana ihtiyaçları karşılamak üzerine bir yapı sergilemektedir. Bunun dışında yine TÜİK (2020) verilerine göre 2019 yılı üçer aylık dönemlerinde en yüksek turizm geliri yıllık turizm gelirin %40'ına eşit olan 3. çeyrek (Temmuz-Ağustos-Eylül) dönemine aittir. En düşük turizm geliri ise %13'lük oran ile ilk çeyrek (Ocak-Şubat-Mart) verilerine aittir. Bu da gösteriyor ki Türkiye'de sadece yaz aylarında otel turizmine gösterilen ilgi turizm gelirin büyük bir kısmını oluşturmaktadır. Dolayısıyla sağlık turizmi, spor turizmi, kültür turizmi, eğitim turizmi, din turizmi, konferanslar vb. alanlarda ilgi görecektir. Çalışmaların kalitesinin, tanıtımının ve yaygınlığının artırılmasıyla ilgili gerekli aksiyonlar alınıp gelir kalemlerinin turizm gelirene katkısının yükseltilmesi turizm getirisi açısından son derece önemlidir.

Yurtdışından ziyaretçi alan ülkelerin makroekonomik göstergelerinin ise ziyaretçi sayılarını etkilemekten ziyade turizm gelirlerini etkileyebileceği düşünülmektedir. Bu göstergelerden Türkiye için yabancı ziyaretçinin alım gücünü ve turizm gelirini etkileyecek iki gösterge; enflasyon ve döviz kuru değerleridir. Turizm gelirleri açısından yaklaşık 17 yıllık bir aralığın gözlemlendiği bu çalışmada, bu tarihler arasında döviz kuruyla alışveriş yapan yabancı ziyaretçilerin alım gücünden bahsetmek gerekmektedir. Aşağıda Şekil 3.2'de enflasyon ve döviz kuru etkisinden arındırılıp düzenlenmiş kişi başı yabancı ziyaretçi harcamalarıyla dolar cinsinden kişi başı yabancı ziyaretçi harcamaları hareketli ortalama yöntemiyle mevsimsellikten arındırılıp karşılaştırılmaktadır.



Şekil 3.2: Türkiye kişi başı turizm geliri göstergesi.

Şekil 3.2’de de görülebileceği gibi yabancı kişi başı turizm gelirlerinde enflasyon ve döviz kuru değerleri hem olumlu hem de olumsuz etkiler yaratabilmektedir. Türkiye’deki turist profilinin tercih sebeplerinden birisinin de yüksek döviz kurları olduğu gösterilebilir. Türkiye’deki turist profilinin fiyat endekslerinin düşük olduğu ülkeleri tercih etme ve gelir seviyesinin yükseldikçe başka ülkelere yönelme gibi eğilimlerinin olduğu çeşitli çalışmalarca tespit edilmiştir (Aydın ve diğ. 2015; Keskin 2019). Öte yandan Şekil 3.3’te de görülebileceği gibi turizm gelirleri genel olarak döviz kuru ve enflasyon etkilerinden arındırılıp düzenlendiklerinde eğilimlerde kişi başı turizm gelirinde görüldüğü gibi keskin değişikliklerle karşılaşılmamaktadır. Bu ise bahsedilen turist profilinin döviz kuru artışına karşı gösterdiği tepkiden kaynaklanmaktadır.



Şekil 3.3: Türkiye toplam turizm geliri göstergesi.

Kısacası Türkiye’de fiyat endekslerinin azalması kişi başı turist harcamasını azaltabilirken toplam gelen turist sayısını arttırdığından döviz kuru ve enflasyon değerlerinin artmasının turizm geliri üzerindeki etkisinin tam anlamıyla negatif ya da pozitif olduğu kanısına varmak kesinlik içermemektedir. Bu nedenle hem olumlu ve olumsuz etkilerin faktör bazında daha iyi anlaşılabilmesi adına hem de küresel bir karşılaştırma yapabilmek ve yapısal kırılmaların nedenini daha iyi açıklayabilmek adına turizm gelirinde düzeltme yapmadan dolar kurunu kullanmanın daha doğru olacağı düşünülmüştür.

Turizm gelirlerinin ülke ekonomisi üzerindeki yadsınamaz etkisi ve önemi kabul edilmelidir. Ülkeye gelen turist sayısının yanında, bu ziyaretçilerin ülke ekonomisi için önemli bir ekonomik değer oluşturmalarının yolları aranmalıdır. Bu kapsamda turizm gelirini etkileyen faktörler belirlenirken sadece ekonomik ölçütler değil aynı zamanda yabancı ziyaretçilerin turizm için Türkiye’yi tercih sebepleri ve turizm harcamalarını etkileyen çok boyutlu unsurlar da göz önüne alınmalıdır.

Yukarıda değinilen nedenler ve sonuçlar beraberinde turizm gelirleri üzerinde belirleyici olan birçok farklı unsur bulunabileceği unutulmamalıdır. Bu bağlamda Türkiye’de hangi etkenlerin turizm gelirine etki edebileceği detaylı bir şekilde tartışılmak istenmiştir. Ülkenin ekonomik durumunun yanı sıra turizm kapasitesinin, ülkedeki özgürlük ve güvenlik ortamının ve ülkenin gelişmişlik seviyesinin de turizm

gelirleri üzerindeki etkisinin detaylı bir şekilde sorgulanması gerektiği düşünölmüştür. Her bir grup içerisindeki farklı unsurların turizm geliri üzerinde ne ölçüde anlamlı bir etki oluşturduğu ele alınmalıdır.

Çalışmanın bundan sonraki bölümünde bahsedilen niyet temelinde ekonomi, turizm kapasitesi, özgürlük, güvenlik ve gelişmişlik seviyesi bağlamında belirlenecek kriterler/değişkenler eşliğinde bir sorgulama gerçekleştirilecek, hangi değişkenlerin turizm gelirleri üzerinde anlamlı ve yüksek performansta bir etkiye sahip olduğu anlaşılmaya çalışılacaktır. Bu tespitler yapılırken turizm gelirleri üzerindeki etkisi belirlenen bu faktörlerin gelecekte daha etkin bir tahmin için kullanılabilmesi amacıyla çeşitli yöntemler altında test edilip etkili faktörlerden en çok faydalanılabilecek modeller belirlenmeye çalışılacaktır. Bir sonraki bölümde yukarıda bahsedilen turizm gelirlerini etkileyen faktörleri belirlemek ve bunların turizm geliri üzerindeki etkisini en iyi şekilde açıklayabilmek için kullanılacak yöntemler üzerinde durulmuştur.

4. YÖNTEM

4.1 Çoklu Doğrusal Regresyon

Regresyon, bağımsız değişkenlerle bağımlı değişken arasındaki ilişkiyi bir matematiksel model ile ifade eden klasik istatistiksel bir yöntemdir. Değişik amaçlarla kullanılabileceği bilinen regresyon modelleri genel olarak tahmin alanında kullanılmaktadır. Regresyon bağımlı ve bağımsız değişkenler arasındaki ilişkiyi doğrusal veya doğrusal olmayan eğrilerle ifade edilebilmektedir. Bu ilişkileri açıklamak için kullanılan denklemlerden bazıları;

- Doğrusal Denklem,
- Logaritmik Denklem,
- Üstel Denklem,
- Polinomal Denklem,
- S Denklem şeklinde olabilmektedir.

Genel olarak doğrusal ve eğrisel denklemleri birbirinden ayırmak gerekirse, iki değişken içerisinde bir değişken bir birim değiştirildiğinde diğer değişkende sabit bir değişim gözleniyorsa bu ilişki doğrusal bir ilişki olarak adlandırılabilir (Zaman 2017).

Doğrusal regresyonda ise tek bir bağımsız değişken ile bağımlı değişken açıklanmaya çalışılıyorsa basit doğrusal regresyon, model birden fazla bağımsız değişken barındırıyor ise bu çoklu doğrusal regresyon olarak adlandırılmaktadır. Bu çalışmada çoklu doğrusal regresyon modeli kurulduğu için bundan sonraki aşamalarda çoklu doğrusal regresyon modeli üzerinden gidilecektir. Bu bağlamda bir çoklu doğrusal regresyon denklemi aşağıdaki gibidir.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n + c \quad (4.1)$$

(4.1) denklemine göre n tane bağımsız x değişkeni olan y bağımlı değişkeni çoklu doğrusal regresyonla açıklanmaktadır ve β_0 sabit katsayıyı $\beta_{1+2+\dots+n}$ bağımsız değişkenlerin katsayısını ve c hata miktarını göstermektedir. Denklemden de

anlaşılacağı gibi her y gözlem değeri çizilen regresyon doğrusu üzerinde yer almayabilir. Bu denklem sadece x değişkenleriyle y değişkenini bir kural çerçevesinde optimum hatayla anlatabilmenin matematiksel bir yolunu bulmayı amaçlamaktadır. Bu çalışmada regresyon modelinde tahmin için En Küçük Kareler (EKK) yöntemi kullanılmıştır. Bu yöntemde hata (c) karelerinin toplamının en küçük olması istenilmektedir. Hata değeri denklemde eşitliğin sağ tarafına alınıp karesi alınır, toplanır ve elde edilen formül için kısmi türev uygulanır. Böylece bulunan parametreler denklemde yerine yazılır ve regresyon modeli kurulur.

Regresyon modelleri bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ifade eder fakat bilmek gerekir ki regresyon modeli bu değişkenler arasındaki nedenselliği açıklamamaktadır. Regresyon genel olarak değişkenler arasındaki ilişkiyi bir matematiksel denklemle ortaya koymaktadır. Değişkenlerin aralarındaki ilişkinin gücü yani bağımsız değişkenlerin bağımlı değişkeni ne ölçüde karşılayabildiği determinasyon katsayısı R^2 değeri ile ölçülür. Aslında bu R^2 değeri değişkenler tarafından paylaşılan varyans oranını göstermektedir. Dolayısıyla modelde bir nedensellikten bahsedebilmek için etkili olabileceği düşünülen kanıtı dayalı bilgiler denkleme dahil edilmelidir. Çoklu regresyon modelinde düzeltilmiş R^2 değerleri oransal olarak denklemin açıklayıcılık yüzdesini temsil edebilmektedir (Cevahir 2020).

4.1.1 Çoklu Doğrusal Regresyonun Gereklilikleri

Çoklu doğrusal regresyon modeliyle doğru ve gerçeğe yakın bir çıkarsama yapabilmek için bazı ön gerekliliklerin sağlanması gerekmektedir. Özellikle zaman serilerinde bu durum oldukça önemlidir. Çoklu doğrusal regresyon analizinin gerekliliklerinden aşağıda bahsedilmektedir.

- Değişkenler normal dağılıma sahip olmalıdır.
- Bağımlı ve bağımsız değişkenler arasındaki ilişki doğrusal olmalıdır.
- Bağımsız değişkenlerin birbirleri arasındaki ilişki çok yüksek olmamalıdır (çoklu doğrusallık sorunu).
- Model artıkları otokorelasyona neden olmamalıdır.

- Modelin artıklarının yani tahmin değeri ile istenilen değer arasındaki farkların normal bir dağılıma sahip olması gerekmektedir.
- Modelin artıklarının değişken bir varyansa sahip olmaması gerekmektedir. (Aktaş 2005; Cevahir 2020).

4.1.1.1 Normal Dağılım Varsayımı

Değişkenleri normal dağılıma sahip olup olmadıklarını test etmenin veya gözlemleyebilmenin birçok yolu mevcuttur. Bunlardan bazıları;

- Histogram grafiğinin çan eğrisi şeklinde olması,
- Q-Q plot grafiğindeki sapmaların kontrol edilmesi,
- Ortalama, medyan ve mod değerlerinin birbirlerine yakınlığının kontrol edilmesi,
- Kolmogorov Smirnov, Shapiro Wilk vb. testlerin uygulanması,
- Çarpıklık ve basıklık değerlerinin kontrol edilmesi gibi yöntemlerdir.

Tabachnick ve Fidell (2013) verilerin normal dağılıma sahip olabilmesi için çarpıklık ve basıklık değerlerinin +1,5 ve -1,5 aralığında olmasının yeterli olacağını açıklamıştır. Bunun dışında verilerin hata değerlerinin (artıklarının) de normal dağılım gösterip göstermediğinin takibi için birçok normallik testinden yararlanılmaktadır. Normal dağılımın saptanmasında bu kadar farklı yolun geliştirilmiş olmasının nedeni ise daha sağlam parametrik testlerden faydalanabilmek adına daha çok doğru verinin sisteme dahil edilmesi isteği olduğu düşünülmektedir.

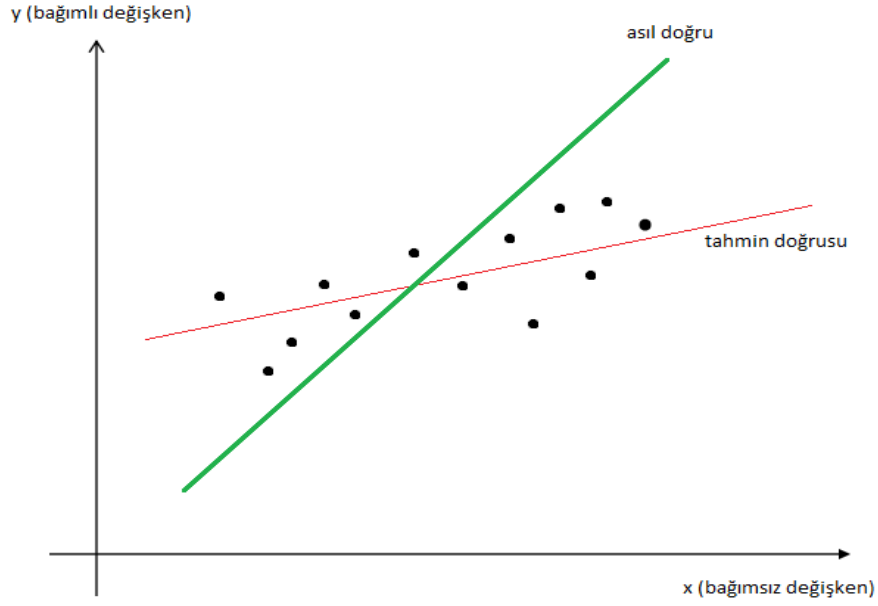
4.1.1.2 Korelasyon Varsayımı ve Çoklu Doğrusallık Problemi

İkişerli olarak değişkenler arasındaki doğrusal ilişkinin gücü ve yönü hakkında bilgi alınabilmesi için Pearson korelasyon analizinden faydalanılmaktadır. Korelasyon analizinde iki değişken arasındaki ilişkinin anlamlılığı, ne oranda olduğu ve yönü belirlenmektedir. Fakat bu analizde bir neden-sonuç ilişkisi çıkarımı kesinlikle yapılamamaktadır. Korelasyon katsayısı (r) ilişkinin gücü ve yönüne göre -1 ile +1 arasında bir değer almaktadır. İşaretler ilişkinin yönünü gösterirken değer

1'e yaklaşması iki değişken arasında çok yüksek bir ilişkinin olduğunu ortaya koymaktadır. Eğer p olasılık değeri anlamlılık değerinin (bu çalışma için 0,05) altındaysa bu iki değişken arasında anlamlı bir ilişkinin olduğu anlamına gelmektedir. Daha sonra iki değişken için anlamlılığın gücüne ve yönüne bakılmaktadır. Negatif işaret değişkenler arasında ters yönlü bir ilişkinin olduğunu gösterirken pozitif işaret ise değişkenlerin aynı yönlü bir ilişkisinin bulunduğunu göstermektedir. Korelasyon analizi regresyon modeli üzerinde değişkenlerin hem bağımlı değişkenle anlamlı doğrusal ilişkiye sahip olmayan bağımsız değişkenleri belirlemede hem de çoklu doğrusallık sorununu tespit etmede kullanılmaktadır. Bağımsız değişkenler arası çoklu doğrusallığın varlığı 0,8 r korelasyon katsayısında olması ihtimaline karşın 0,9 r korelasyon katsayısında büyük sorunlara neden olabilmektedir (Cevahir 2020). Bunların dışında özellikle bu çoklu doğrusallık problemini analiz etmek için kullanılan bazı yöntemler de mevcuttur. Bunlardan en önemlisi Varyans Şişirme Faktörü (VIF) hesaplamasıdır. Üst limit kabul olarak VIF değeri 10'dan büyük olan değişkenlerin çoklu doğrusallık sorunu yaratabileceğinden bahsedilmektedir (Büyükuysal ve Öz 2016).

4.1.1.3 Otokorelasyon Problemi

En Küçük Kareler (EKK) tahminleyicisinin kullanıldığı bir regresyon modelinde artıkların birbirleri arasındaki ilişkiye otokorelasyon denilmektedir. Özellikle zaman serisi içeren regresyon modellerinde t anındaki hata değeri ile gecikmeli hata değerleri arasında pozitif bir ilişkinin bulunması varyansların olduğundan küçük bulunmasına sebep olabilmektedir. Bu durum ise t ve F istatistik değerlerinin ve R^2 değerinin olduğu değerden büyük çıkmasına sebep olabilmektedir. Böylece aslında herhangi bir açıklayıcılığı bulunmayan bir katsayının anlamlı ve yüksek düzeyde bir ilişki değerine sahip çıkması sorunu ile karşı karşıya kalınabilir. Otokorelasyon sorunu veri ölçüm hatası, gerekli bağımsız değişkenlerin sisteme dahil edilmemesi, yanlış regresyon modelinin kullanılması vb. sorunlardan dolayı kaynaklanabilmektedir (Yavuz 2009). Şekil 4.1 üzerinde otokorelasyon sorununun daha iyi anlaşılabilmesi adına regresyon doğrularının bulunduğu bir grafik verilmiştir.



Şekil 4.1: Otokorelasyon sorunu grafik gösterimi.

Otokorelasyon probleminin saptanması için grafik ile saptama, sıra testi, LM testi, Durbin-Watson testi, Wallis testi, Von-Neumann testi gibi birçok yöntem mevcuttur. Modellerin veri sayısı, sabit katsayının olup olmaması, modelin doğrusal olup olmaması gibi özelliklerine göre uygulanabilecek otokorelasyon saptama testi de değişmektedir. Buna karşın parametrik olarak sabit katsayıya sahip modellerde genellikle Durbin-Watson yöntemi kullanılmaktadır.

$$d = \frac{\sum_{t=2}^n (c_t - c_{t-1})^2}{\sum_{t=1}^n (c_t)^2} \quad (4.2)$$

(4.2) denkleminde t zamanındaki artık değerler c ile ifade edilmekte ve hesaplanan d değeri Durbin-Watson tablosundaki kritik değerlerle karşılaştırılmaktadır. Buna göre;

- $0 < d < d_L$ durumunda pozitif otokorelasyondan,
- $d_L \leq d \leq d_U$ durumunda kararsızlıktan,
- $d_U < d < 4 - d_U$ durumunda otokorelasyonun yokluğundan,
- $4 - d_U \leq d \leq 4 - d_L$ durumunda kararsızlıktan,
- $4 - d_L < d < 4$ durumunda ise negatif otokorelasyondan bahsedilmektedir (Yavuz 2009).

Wallis (1972) yaptığı çalışmada Durbin-Watson testini verilerin üçer aylık olduğu durumlar için uyarlamıştır. Wallis testinin uygulanması için gerekli tek şart bağımsız değişkenlerin rassal olmaması durumudur. Wallis testi Durbin-Watson testinin aynı formülüyle (4.2) dördüncü dereceden otokorelasyon durumu için düzenlenmiştir ve kritik değerler belirlenmiştir.

4.1.1.4 Sabit Varyans Varsayımı

Artıkların sabit bir varyansa sahip olması gerekliliği de regresyon için gerekli şartlar arasında sayılmaktadır. Burada hataların dağılımının herhangi bir değişkenden etkilenmemesi ve sapmaların sabit olması beklenmektedir. Bu durumun saptanabilmesi için de saçılım grafikleri, değişen varyans testleri gibi birçok yöntem mevcuttur. Bu testler içerisinde en çok kullanılan yöntemler White ve Breusch-Pagan-Godfrey testleridir (Uçan ve Saraç 2020).

Regresyon modelinin ana gereklilikleri yukarıda anlatılan şartlar olmakla beraber sağlam ve gerçeği yansıtan bir regresyon modeli oluşturabilmek için dikkat edilmesi gereken başka gerekliliklerin de olabileceği göz önünde bulundurulmalıdır. Bu gereklilikler verinin yapısı, kullanım şekli gibi etkenlere göre irdelenmesi gereken konulardır.

4.1.1.5 Sahte Regresyon Problemi

Regresyon tanımına ilave olarak üzerinde durulması gereken kavramlardan bir tanesi de zaman serisinin uzun dönemde gösterdiği ana eğilimi açıklamak için kullanılan uzun devre eğilimidir (trend). Belirtmek gerekir ki birden fazla değişken arasında anlamlı bir ilişki elde edebilmek için kullanılan serilerin negatif ya da pozitif yönlü güçlü bir eğilim göstermemesi gerekmektedir. Aksi halde değişkenler kapsamındaki zaman serileri içinde bir trendin söz konusu olması halinde eğilimin yönüne bağlı olarak gerçek dışı bir regresyon oluşmaktadır. Ekonometrik analizlerde zaman serisi içeren iki değişken arasında güçlü bir trendin olması, bağımsız değişkenin bağımlı değişkeni gerçek hayatta karşılayamamasına rağmen açıklayıcılık değeri olan R^2 değeri yüksek çıkabilmektedir. Granger ve Newbold (1974) çalışmalarında bu

durumu sahte regresyon olarak adlandırmışlardır. Bahsedilen R^2 değerinin bu denli yüksek çıkması iki değişken arasındaki ilişkiden ziyade güçlü eğilimin getirdiği sonuçtan kaynaklanmaktadır. Bu nedenle bir eğilime sahip zaman serilerinde regresyon modeli oluştururken sadece yüksek bir R^2 değerinin ve anlamlılığın o değişkenler arasındaki ilişkinin kuvvetli olduğunu açıklamaya yetmeyeceği bilinmektedir. Granger ve Newbold (1974) tarafından açıklanan modele göre bir regresyon analizinde Durbin-Watson istatistik değeri R^2 değerine göre çok düşük ise bu regresyonun sahte bir ilişki sergileyebileceği savunulmuştur. Dolayısıyla zaman serilerinin durağanlığı bir regresyon denkleminin gerçeği yansıtabilme gücüyle doğrudan orantılıdır (Uzgören ve Uzgören 2005; Granger 2003).

4.1.2 Durağanlık ve Durağanlığın Ölçülmesi

Yukarıda dile getirilen sahte regresyon tanımı paralelinde üzerinde durulması gereken bir diğer kavram ise durağanlıktır. Durağanlığı kısaca tanımlamak gerekirse güçlü bir eğilim yani trend özelliği göstermeyen, sabit bir varyans ve ortalama özelliği gösteren ve kovaryansı dönem arasındaki fark üzerinden hesaplanan zaman serileri durağan seriler olarak ifade edilir. Durağanlığın test edilmesi için çok sayıda yöntem olmakla birlikte korelogramdan ve örnek otokorelasyon katsayısından faydalanılabilmektedir (Uğurlu 2009). Ancak geliştirilen birim kök testleri aracılığıyla verilerin seviyede ve n. dereceden farklarında durağan olup olmadıkları anlamlı bir şekilde tespit edilmektedir.

4.1.2.1 Birim Kök Analizi

Dickey ve Fuller (1979) parametrelerin En Küçük Kareler (EKK) tahmincisinin dağılımına dayanan bir test geliştirmişlerdir. Başlangıç için regresyon denklemi aşağıda verilmektedir.

$$y_t = py_{t-1} + c_t \quad (4.3)$$

(4.3) formülünde y_t durağanlığın araştırılacağı seriyi y_{t-1} onun bir dönem gecikmesini, c_t ise regresyon hata değerini göstermektedir. Eğer p parametresi 1'e eşit

olursa y_t birim köke sahip ve durağan olmayan bir seridir. Değişimi göstermek adına denklemi,

$$\Delta y_t = \delta y_{t-1} + c_t \quad (4.4)$$

Şeklinde gösterirsek δ parametresi (p-1)'e eşit olacaktır. Birim kök testi sırasında δ parametresinin 0'a eşitliği araştırılır.

Dickey ve Fuller (1979) üç farklı modelden hareketle testi oluşturmuşlardır. Bunlar:

$$\Delta y_t = \delta y_{t-1} + c_t \quad (4.5)$$

$$\Delta y_t = \beta_1 + \delta y_{t-1} + c_t \quad (4.6)$$

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + c_t \quad (4.7)$$

(4.5) formülünde sabitsiz ve trendsiz bir model, (4.6) formülünde sadece sabitli bir model, (4.7) formülünde ise hem sabitli hem de trendli bir model ifade edilmiştir. Birim kök testi y_t 'nin gecikmeli değerinin t istatistiği ile test edilmesidir. Bu yapılırken kritik değerlerden faydalanılmaktadır.

$H_0 = \delta \geq 0$ ise birim kök vardır ve seri durağan değildir.

$H_1 = \delta \leq 0$ ise birim kök yoktur ve seri durağandır.

Analiz sonucu hesaplanan istatistiğin değeri belirlenen kritik değerden küçük ise H_0 hipotezi reddedilecektir ve bu durumda serinin durağan olduğu söylenebilir.

Daha sonraki dönemlerde Dickey ve Fuller (1981) yeni bir çalışmayla bu testi geliştirip gecikmeli değerleri de denkleme dahil etmişlerdir ve Genişletilmiş Dickey Fuller (ADF) testi oluşturulmuştur. Bu testin bir önceki modellerden tek farkı daha uzun dönemlerde gecikmeli değerleri modele dahil edebilmesidir. Böylece otokorelasyonlu hata terimleri denetlenebilecektir. Modellerde gecikme uzunluğu belirlenirken Akaike ve Schwartz bilgi kriterlerinden yararlanılmaktadır. Veri uzunluğu n olan bir seri için maksimum gecikme uzunluğu belirlenirken ise Schwert (1989) tarafından aşağıdaki denklem önerilmiştir.

$$l_{max} = \left[12 \left(\frac{n}{100} \right)^{1/4} \right] \quad (4.8)$$

Zaman serilerinde özellikle geniş aralıklar kullanıldığında belirli dönemlerde konjonktürel değişimler saptanmaktadır. Bu durum ekonometrik verilerde ülkeler için gerek siyasi gerekse ekonomik şok unsurlarının (terör, ekonomik kriz vb.) bir göstergesidir. Klasik birim kök analizlerinde bu yapısal kırılma durumu göz ardı edilmektedir. Perron (1989) geliştirdiği modelde dışsal olarak belirlenebilecek bir yapısal kırılmanın modele dahil edilebileceğini öne sürmüştür.

Zivot ve Andrews (1992) ise böyle bir yapısal kırılmanın dışsal olarak belirlenmesinin birim kökün reddine sebep olabileceğini ileri sürmüştür. Önerilen analizde trendde durağanlık içeren hipotez için en yüksek ağırlık adına kırılma yansımaları tahmin etmek istemiştir. Bu kırılma yansıması en küçük t istatistik değerini veren değer seçilir. Bu test de tıpkı ADF testi gibi çalışmaktadır. İçsel olarak belirlenen yapısal kırılma için kukla değişkeni atanmaktadır. Böylece yapısal kırılma etkisi yok edilerek standart birim kök testi uygulanmaktadır (Yonar 2012).

Model içerisinde yapısal kırılmaları açıklamaya yarayan düzeydeki kukla değişkenleri DU ve eğimdeki kukla değişkenleri ise DT ile ifade edilmiştir.

$$DU(\lambda) = \begin{cases} 1 & t > T_B \\ 0 & t \leq T_B \end{cases} \quad (4.9)$$

$$DT(\lambda) = \begin{cases} t - T\lambda & t > T\lambda \\ 0 & t \leq T\lambda \end{cases} \quad (4.10)$$

Yukarıda denklemlerde T_B kırılmanın gerçekleştiği tarihi, $\lambda = T_B/T$ kırılma noktasını göstermektedir. Bu doğrultuda içsel olarak yani dışarıdan bir müdahale olmadan direkt olarak yöntemin hesapladığı tekli yapısal kırılma modelleri aşağıdaki gibi hesaplanmaktadır.

$$y_t = \mu + \beta t + \delta y_{t-1} + \theta_1 DU(\lambda) + \sum_{i=1}^n \delta_i \Delta y_{t-i} + c_t \quad (4.11)$$

$$y_t = \mu + \beta t + \delta y_{t-1} + \theta_2 DT(\lambda) + \sum_{i=1}^n \delta_i \Delta y_{t-i} + c_t \quad (4.12)$$

$$y_t = \mu + \beta t + \delta y_{t-1} + \theta_1 DU(\lambda) + \theta_2 DT(\lambda) + \sum_{i=1}^n \delta_i \Delta y_{t-i} + c_t \quad (4.13)$$

(4.11) denkleminde seviyede, (4.12) denkleminde trendde, (4.13) denkleminde ise hem seviyede hem de trendde tek yapısal kırılmaya izin veren Zivot-Andrews birim kök testi modelleri gösterilmektedir. Denkleminde belirlenen δ 'nin test edilmesi için t istatistik değeri hesaplanmaktadır. Yapısal kırılmanın tarihi hesaplanan en küçük t istatistik değerine sahip tarih olarak belirlenmektedir. Kırılmanın gerçekleştiği tarih T_B belirlenip denkleminde yerine koyulduktan sonra hesaplanan t istatistik değerinde ise kritik değerlere göre ADF testi ile aynı şekilde birim kökün olup olmadığına karar verilmektedir (Yıldırım Tıraşoğlu 2014; Zivot ve Andrews 1992).

4.1.3 Eşbütünlüme Analizi

Zaman serileri analizinde bir regresyon denkleminin kurulabilmesi için değişkenlerin seviye değerlerinin durağan olması veya aynı entegrasyon derecelerinde eşbütünlüklük olmaları gerekmektedir (Arı ve Yıldız 2017). Eşbütünlüklük iki veya daha fazla zaman serisi arasındaki uzun dönemli ilişkiyi saptamak için kullanılmaktadır. Özellikle seviyede durağan olmayan verilerin entegrasyon derecelerinde durağanlaştığı zaman serilerinde aralarındaki ilişkinin güçlü eğilimlere yani trende bağlı değil de gerçek bir ilişkiye bağlı olduğunu saptamak için kullanılmaktadır. Eğer analizlerde En Küçük Kareler (EKK) yöntemi kullanılacaksa verilerin durağan olmaması analizin yanlış yorumlanmasına neden olacaktır. Fakat durağan olmayan bir verinin analize alınmaması o veriden elde edilebilecek önemli uzun dönemli birlikteliklerin de göz ardı edilmesine neden olacaktır. Eşbütünlüklük analizleri durağan olmayan zaman serileri arasındaki hataların birlikteliğine odaklanarak bahsedilen bu ilişkiyi bulmakta kullanılmaktadır (Işık ve diğ. 2004).

Engle ve Granger (1987) eşbütünlüklüğü açıklamak için aşağıdaki formülü kullanmıştır.

$$y_t = \beta_0 + \beta_1 x_t + c_t \quad (4.14)$$

(4.14) denkleminde x ve y birinci düzeyde durağanlaşan değişkenlerdir. Bu iki değişken arasında uzun dönemli bir ilişkinin olabilmesi için c_t hata değerlerinin seviyede durağan olması gerekmektedir. Kısacası iki değişkenin hatalarının seviye değerleri birim kök içermiyorsa iki değer eşbütünleşiktir. Burada dikkat edilmesi gereken nokta tıpkı korelasyon analizi gibi iki değer arasında da eşbütünleşmenin olması bu ilişkinin yönünü belirtmemektedir. Bunun için nedensellik ilişkilerine başvurulabilir.

Yukarıdaki modellerde yapılan varsayım eşbütünleşik vektörün zaman içerisinde sabit kaldığı yönündedir. Gregory ve Hansen (1996), yapısal kırılmaya izin veren sabit ve trende sahip setler için üç farklı eşbütünleşme modeli önermiştir. Bu model de tıpkı Zivot-Andrews birim kök testinde olduğu gibi içsel olarak yapısal kırılmayı tespit etmektedir.

Yapısal kırılmanın model içerisinde kullanılabilmesi adına aşağıdaki kukla değişkeni kullanılmıştır.

$$\varphi_{1t} = \begin{cases} 1 & t > nT \\ 0 & t \leq nT \end{cases} \quad (4.15)$$

Denklem (4.15) üzerinde T yapısal kırılma noktasını, nT ise yapısal kırılma noktasındaki tam sayıyı ifade etmektedir. Sabitte kırılma ile eşbütünleşmeyi gösteren model (C) denklemi aşağıdaki gibidir.

$$y_{1t} = \mu_1 + \mu_2 \varphi_{1t} + \delta^T y_{2t} + c_t \quad (4.16)$$

$$y_{1t} = \mu_1 + \mu_2 \varphi_{1t} + \beta t \delta^T y_{2t} + c_t \quad (4.17)$$

$$y_{1t} = \mu_1 + \mu_2 \varphi_{1t} + \delta_1^T y_{2t} + \delta_2^T y_{2t} \varphi_{1t} + c_t \quad (4.18)$$

Burada denklem (4.16) ile sabitte kırılma modeli gösterilmektedir. Yapısal kırılmadan önceki sabit değeri μ_1 , yapısal kırılma için sabitte yapılan değişikliği ise μ_2 değeri göstermektedir. Denklem (4.17) ise hem sabitte hem de trenddeki kırılmaları işleme dahil eden (C/T) modeldir. Rejim değişikliği (C/S) modelinde kurulan (4.18) denklemde ise sabitte kırılma (C) modelinin yanı sıra δ_1 ile yapısal kırılma öncesi

trendin katsayısı, δ_2 ile ise yapısal kırılmada trendin katsayısında meydana gelen değişiklik üzerine bir model kullanılmıştır. Bu modeller sonucunda Genişletilmiş Dickey Fuller (ADF) test istatistiğinin en küçük olduğu değer yapısal kırılma tarihi olarak kullanılmaktadır. Daha sonra yapısal kırılmanın modele yerleştirildiği denklemden çıkan sonucun test istatistiği Gregory-Hansen kritik değerleri ile karşılaştırılıp değerlerin eşbütünleşik olup olmadığına karar verilmektedir. Bu yöntemde ikiden fazla değişken sayısına izin verilip her bir değişken sayısı, model türü ve güven aralığı için kritik değerler değişim göstermektedir (Ünlü ve Bozdağ 2016; Gregory ve Hansen 1996).

4.2 Yapay Sinir Ağları

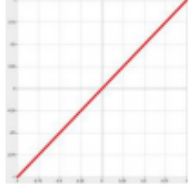
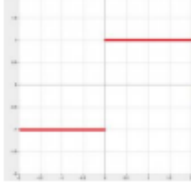
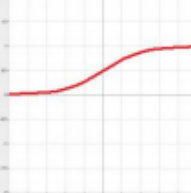
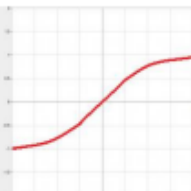
Makine öğrenmesi, geçmiş verilerden faydalanarak bilgisayar üzerinde kendi tecrübesini yaratan ve bu tecrübe doğrultusunda bir modelleme yapabilen ve edindiği tecrübelerle yani veri sayısına ve kalitesine göre modeli iyileştirebilen bir yöntemdir. Genel tanımıyla yapay zeka biliminin bir alt dalı olan Yapay Sinir Ağları (YSA), insan beyninin öğrenme sistemini bilgisayar aracılığıyla taklit ederek modellemede bulunan bir makine öğrenmesi metodudur. McCulloch ve Pitts tarafından 1943 yılında ortaya atılan makalede, sinir hücrelerinin girdi ağırlıklarıyla çıktısı 1 ve 0 değeri alabilen ilk YSA modelinin matematiksel olarak açıklandığı belirtilmiştir. İlk denemeler 1958 yılında Rosenblatt tarafından kavramların ağlara algılatılması yoluyla gerçekleşmiştir. 1980'li yıllardan sonra tekrar yaygınlaşmaya başlayan YSA modelleri 1986 yılında Rummelhart ve McClelland tarafından detaylı olarak açıklanabilen geri yayılım yöntemiyle diğer yöntemlere göre çok büyük bir avantaj sağlamıştır ve problemlerin çözümünde YSA uygulamasını çok etkili hale getirmiştir (Demir 2007).

4.2.1 Yapay Sinir Hücreleri

Biyolojik olarak birçok sinir hücresinin (nöron) bir araya gelmesi sonucunda sinir ağları oluşmaktadır. Gerçek sinir hücreleri aralarındaki iletişim sinapslar aracılığıyla sağlanmaktadır. Sinir hücreleri işlenmiş bilgileri hücrelere aktarmaktadır. Yapay Sinir Ağları (YSA) da bu yapıya benzer şekilde yapay sinir hücrelerinin (nöron)

aralarında bağ kurmasıyla bir yapı oluşturmaktadır. Yapay nöronlar giriş verilerini dış ortamdan sağlayıp, bir toplama fonksiyonu ile verileri toplayıp, aktivasyon fonksiyonuyla işler ve aktarımını sağlarlar. Aktarımın sonunda bir çıktı ile elde edilen bilgiyi dış ortamla paylaşırlar. YSA'nın yine tıpkı gerçek hayattaki gibi birbirleri arasında bağlantıları ve bu bağlantıların birer ağırlık değeri vardır. Genel olarak özetlemek gerekirse yapay sinir hücreleri girdi, ağırlık, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktı olmak üzere 5 aşamaya sahiptir (Öztemel 2003).

Girdi başka bir nöron veya dış ortamdan ilk kez sağlanan verinin nörona girişi aşamasıdır. Sinir hücreleri arasındaki bağlantıların önem düzeyine göre bir ağırlık değeri vardır. Bu ağırlık değerleri o bilginin çıktığı ne kadar etkileyeceğini belirler. Toplama fonksiyonları nöronların net girdisini hesaplamak için kullanılan fonksiyonlardır. Bu fonksiyonlarda net girdi belirlenirken genellikle ağırlık ve girdi değerlerinin birbirleriyle çarpımlarının toplanması, çarpılması, en büyüğünün alınması vb. yöntemler kullanılır. Aktivasyon fonksiyonu ise işlenen girdi sonucunda üretilen çıktıyı hesaplayan bir fonksiyondur. Bu değer ise çıktı ile diğer hücrelere veya dışarıya iletilir (Öztemel 2003). Aktivasyon fonksiyonunun belirlenmesinde de çeşitli yöntemler kullanılmakla beraber doğrusal problemlerin çözümünde doğrusal aktivasyon fonksiyonu, doğrusal olmayan problemlerin çözümünde ise diğer yöntemler uygulanmaktadır. Bu yöntemler içerisinde en çok tercih edileni Sigmoid yani lojistik formüle sahip yöntem olmakla beraber bazı yöntemlerin genel gösterimi Şekil 4.2'de yer almaktadır.

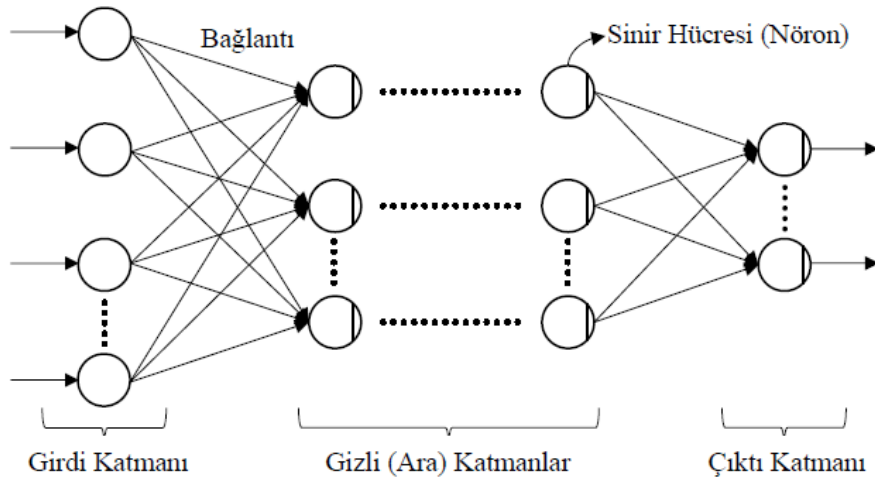
Doğrusal (Lineer) Aktivasyon Fonksiyonu		$F(\text{Net})=A \cdot \text{NET}$ (A sabit bir sayı)	Doğrusal problemler çözmek amacıyla aktivasyon fonksiyonu doğrusal bir fonksiyon olarak seçilebilir. Toplama fonksiyonundan çıkan sonuç, belli bir katsayı ile çarpılarak hücrenin çıktısı olarak hesaplanır.
Adım (Step) Aktivasyon Fonksiyonu		$F(\text{Net})= \begin{cases} 1 & \text{if Net} > \text{Eşik Değer} \\ 0 & \text{if Net} \leq \text{Eşik Değer} \end{cases}$	Gelen Net girdinin belirlenen bir eşik değerinin altında veya üstünde olmasına göre hücrenin çıktısı 1 veya 0 değerini alır.
Sigmoid Aktivasyon Fonksiyonu		$F(\text{Net})= \frac{1}{1+e^{-\text{Net}}}$	Sigmoid aktivasyon fonksiyonu sürekli ve türevi alınabilir bir fonksiyondur. Doğrusal olmayışı dolayısıyla yapay sinir ağı uygulamalarında en sık kullanılan fonksiyondur. Bu fonksiyon girdi değerlerinin her biri için 0 ile 1 arasında bir değer üretir.
Tanjant Hiperbolik Aktivasyon Fonksiyonu		$F(\text{Net})= \frac{e^{\text{Net}} + e^{-\text{Net}}}{e^{\text{Net}} - e^{-\text{Net}}}$	Tanjant hiperbolik fonksiyonu, sigmoid fonksiyonuna benzer bir fonksiyondur. Sigmoid fonksiyonunda çıkış değerleri 0 ile 1 arasında değişirken hiperbolik tanjant fonksiyonunun çıkış değerleri -1 ile 1 arasında değişmektedir.
Eşik Değer Fonksiyonu		$F(\text{Net})= \begin{cases} 0 & \text{if Net} \leq 0 \\ \text{Net} & \text{if } 0 < \text{Net} < 1 \\ 1 & \text{if Net} \geq 1 \end{cases}$	Gelen bilgilerin 0 dan küçük-eşit olduğunda 0 çıktısı, 1 den büyük-eşit olduğunda 1 çıktısı, 0 ile 1 arasında olduğunda ise yine kendisini veren çıktılar üretilebilir.
Sinüs Aktivasyon Fonksiyonu		$F(\text{Net}) = \text{Sin}(\text{Net})$	Öğrenilmesi düşünülen olayların sinüs fonksiyonuna uygun dağılım gösterdiği durumlarda kullanılır.

Şekil 4.2: Bazı aktivasyon fonksiyonları (Çayiroğlu 2015).

4.2.2 Yapay Sinir Ağlarının İşleyişi

Yapay Sinir Ağları (YSA) katmanlardan oluşan ve her katmanda birbirlerine bağlantıları bulunan nöronlara sahip bir yapıdır. Bir YSA işleyişi üç katman üzerinden ilerlemektedir. Girdi katmanında dışarıdan bilgiler elde edilir ve genellikle bir işleme uğramadan bir sonraki katmana iletilir, gizli katman bazı yapılarda bulunmayabileceği gibi içerisinde barındırdığı nöron sayısı diğer katmanlardan bağımsız olarak değişebilir ve bilginin işlendiği yerdir, çıkış katmanı ise gizli katmandan veya girdi katmanından aldığı bilgileri işleyip bulduğu çıktıyı dışarıya aktaran katmandır. Her bir katmanda kaç adet nöron olacağı, gizli katman sayısı gibi değerler probleme özgü

olarak değişmekte ve kesin bir yöntem barındırmamaktadır. Yapıdaki nöron sayısının veya katman sayısının artmasının hesaplamayı ve işlem sürecini zorlaştıracağı bilinmektedir. Dolayısıyla katmanlardaki nöron sayısı veya gizli katman sayıları belirlenirken problemin karmaşıklık derecesi göz önünde bulundurulur. Fakat genellikle problem doğrusallıktan uzak ve çok kompleks olmadığı sürece tek gizli katman yeterli görülmektedir (Öztemel 2003). Şekil 4.3 ile genel olarak Çok Katmanlı Algılayıcı (MLP) sisteminin akış görünümü gösterilmektedir.



Şekil 4.3: Örnek yapay sinir ağları modeli gösterimi (Karakullukçu 2020).

Modellerde Şekil 4.3'te gösterilen nöronların sayısı, gizli katmanların varlığı ve sayısı problemin genel durumuna göre değişiklik gösterebilmektedir. Ayrıca akış yönü de yine problemin karmaşıklığı ve verilerin türüne göre değişmektedir.

4.2.3 Yapay Sinir Ağlarının Sınıflandırılması

Uygulamalarda iki farklı öğrenme yönteminden bahsedilmektedir. Elde edilmesi istenen çıktılarla ilgili herhangi bir bilgi mevcut değilse sistem geçmiş verilerden yararlanarak ve bu veriler ile elde edilen tecrübeler doğrultusunda bir genelleme yapar ve çıkarımda bulunur. Bu yöntemde denetimsiz öğrenme denilmektedir ve genellikle sınıflandırma, kümeleme gibi problemler için kullanılmaktadır. Diğer öğrenme yöntemi ise denetimli öğrenmedir. Bu yöntemde ise çıkarım yapılacak veriler mevcuttur. Böylece sistemin tahmin verileri gerçek verilerle karşılaştırılarak bir sonraki iterasyonlarda gerçek veriye daha yakın bir model üretebilmek için algoritma güncellenir. Yapay Sinir Ağları (YSA) yönteminde bu

durum her seferinde daha az hata oranı için nöronlar arasındaki ağların ağırlık değerlerinin değiştirilmesiyle gerçekleşmektedir. YSA modellerinin çeşitlendirilmesinde de öğrenme yöntemleri, problemin karmaşıklığı ve doğrusallığı önemli kriterler arasındadır.

4.2.3.1 Akış Yönüne Göre Yapay Sinir Ağları

Yapay sinir ağlarının sınıflandırılması için kullanılan ilk yöntem bağlantıların akış yönü olarak nitelendirilmektedir. Eğer bahsi geçen akış giriş katmanından çıkış katmanına doğru tek yönlü ve bir katman çıkışından bir sonraki katman girişine ağırlıklar dikkate alınarak yapılıyorsa ve veriler gizli katmanlar ve çıkış katmanlarında işleniyorsa bu ileri beslemeli ağı temsil etmektedir (Ataseven 2013; Rojas 1996).

Geri beslemeli ağlarda ise akış tek yönlü değildir. Katmanlardan elde edilen çıktı bir önceki katmana, kendi katmanına ve bir sonraki katmana bilgi oluşturmak için dağıtılabilir (Öztürk ve Şahin 2018). Bu özelliğiyle geri beslemeli ağlar doğrusal olmayan ve denetimsiz öğrenme yöntemine sahip problemler için kullanılmaya elverişlidir.

4.2.3.2 Katman Sayısına Göre Yapay Sinir Ağları

Bahsedilen özelliklerin dışında katman sayısı da Yapay Sinir Ağlarının (YSA) sınıflandırılması için kullanılan bir özelliktir. Eğer sistemde sadece giriş ve çıkış katmanından oluşan bir yapı varsa bu tek katmanlı YSA modeli olarak adlandırılmaktadır. Fakat giriş ve çıkış katmanlarının dışında gizli katmanlar da bulunuyor ise bu çok katmanlı YSA yani Çok Katmanlı Algılayıcı (MLP) olarak adlandırılır. Tek katmanlı YSA daha basit ve doğrusal bir problem üzerinde kullanılabilirken, MLP doğrusal olmayan, kompleks problemler için kullanıma elverişli hale getirilmiştir.

4.2.4 Geri Yayılım Algoritması

Doğrusal olmayan ve denetimli bir öğrenme metodunda yapay sinir ağlarının oluşturdukları çıktılar için bir uygunluk değeri hesaplanmaktadır. Bu uygunluk fonksiyonu aşağıdaki gibidir.

$$U = \frac{1}{2} \sum_i c_i^2 n \quad (4.19)$$

(4.19) denkleminde i verisi için c gerçek değerden sapma miktarını, n iterasyon sayısını vermektedir. Bu uygunluk değeri sistemin çıktısının gerçek sonuçlarla karşılaştırılmasından elde edilir ve minimum olması istenilir. Her seferinde uygunluk değerinin azalması için ağların ağırlık değerleri güncellenir ve böylece geriye doğru tekrar tekrar bir düzeltme meydana gelir. Bu ağırlık değişme miktarı gradyan azalma yöntemi kullanılarak bulunmaktadır (Ataseven 2013).

$$\Delta w_{ij} = -m \frac{\delta U(w)}{\delta w_{ij}} \quad (4.20)$$

(4.20) denkleminde gradyan azaltma denkleminde m öğrenme katsayısı ve w_{ij} ağırlık değerini göstermektedir. Geri yayılım metodunda her iterasyonda hem ileri hem de geri yayılım aşamaları takip edilir. İleri yayılım YSA için anlık durumun giriş sinyallerine karşı çıkışta değerler oluşturur. Geri yayılım aşamasına ise çıkış kısmında oluşan hatalar ağırlığın yeniden düzenlemesini sağlar. Her ağırlık değişimi aşağıdaki gibi yapılır.

$$\Delta w_{ij} = -m \theta_j y_i \quad (4.21)$$

(4.21) denklemindeki θ_j çıkış ve gizli katmanlarda farklı şekillerde hesaplanmaktadır. Çıkış katmanı için aşağıdaki gibidir.

$$\theta_j = c_j(n) f_j \quad (4.22)$$

Gizli katman sınırları için ise (4.23) denkleminde hesaplanmaktadır.

$$\theta_j = f_j \sum_p \theta_p + w_{pj} \quad (4.23)$$

Burada f_j j siniri için kullanılan aktivasyon fonksiyonudur. Böylece hataların çıkıştan geriye yani girişe doğru akışı ilk aşamadaki ileri akış yöntemine benzetilmektedir. Bu şekilde bir işleyişe sahip olup her iterasyonda bir önceki iterasyonun uygunluk değerine göre davranış sergileyen bu algoritma geri yayılım algoritması olarak adlandırılmaktadır (Ataseven 2013).

Geri yayılım algoritmasında ağın asıl çıktısı ileri besleme ile belirlenirken oluşan hataların gradyan azaltma ile uygunluk değeri çıkartılır ve bu geri besleme olarak sistem ağırlıklarının güncellenmesinde kullanılır. Bu tür bir algoritma uygulanırken giriş katmanındaki nöron sayıları girdi verisi kadar olmalıdır. Böylece ileri beslemeli bu sistemde girdi verileri giriş katmandaki her bir nörona dağıtılır.

4.2.5 Esnek Geri Yayılım Algoritması

Denetimli öğrenme sürecinde ağırlıkların güncellenmesiyle ilgili çok fazla algoritma ileri sürülmüştür. Bu algoritmalar genellikle yerel ve küresel olmak üzere iki mantıkla çalışmaktadır. Küresel stratejilerde önceki ağırlık adımının yönü gibi tüm ağın durumunu ilgilendiren bilgiler göz önünde bulundurulurken, yerel stratejilerde kısmi türev gibi yalnızca ağırlığa bağlı bilgilerle ilgilenilmektedir. Bir algoritmada küresel olarak sisteme yoğunlaşmak sistemin yavaşlamasına, yerel çözümlere odaklanmak ise iyi olmayan çözümler üzerinde duraksamalara neden olabilmektedir.

Esnek Geri Yayılım (RPROP) algoritması diğer algoritmalara göre bulanık olmayan, gradyan bilgisine dayalı doğrudan adaptasyon sağlayan bir öğrenme algoritmasıdır. Daha hızlı bir öğrenme için RPROP algoritmasıyla her ağırlık güncellemesi için yalnızca ağırlık güncellemesinin büyüklüğünü belirleyen bir güncelleme değeri (Δ_{ij}) atanmaktadır. Bu uyarlanabilir güncelleme değeri E hata fonksiyonunun yerel bilgilerine göre bir öğrenme kuralı geliştirmektedir (Riedmiller ve Braun 1993).

$$\Delta_{ij}^{(t)} = \begin{cases} n^+ * \Delta_{ij}^{(t-1)}, & \text{eğer } \frac{\delta E^{(t-1)}}{\delta w_{ij}} * \frac{\delta E^{(t)}}{\delta w_{ij}} > 0 \\ n^- * \Delta_{ij}^{(t-1)}, & \text{eğer } \frac{\delta E^{(t-1)}}{\delta w_{ij}} * \frac{\delta E^{(t)}}{\delta w_{ij}} < 0 \\ \Delta_{ij}^{(t-1)}, & \text{değilse} \end{cases} \quad (4.24)$$

$$0 < n^- < 1 < n^+$$

Riedmiller ve Braun (1993) yaptıkları çalışmada kısmi türevin işaret değiştirdiği durumlarda ağırlık değerinin (w_{ij}) çok büyük bir değer almasıyla algoritmanın yerel bir minimum değeri pas geçtiğini ve güncelleme değerinin (Δ_{ij}) n^- kadar azaltılması gerektiğini göstermektedir. Türevin işaretinin korunduğu durumlarda ise o bölgedeki taramayı yoğunlaştırmak için güncelleme değeri (Δ_{ij}) arttırılır. Her ağırlık için bu durum tekrarlandıktan sonra türev pozitif çıkarsa yani hata artarsa ağırlık değeri güncelleme değeri kadar azaltılır. Türev negatif çıktığı takdirde ise güncelleme değeri eklenir.

$$\Delta_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, & \text{eğer } \frac{\delta E^{(t)}}{\delta w_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, & \text{eğer } \frac{\delta E^{(t)}}{\delta w_{ij}} < 0 \\ 0, & \text{değilse} \end{cases} \quad (4.25)$$

Burada bir sonraki ağırlık değeri güncellemesi ise aşağıdaki gibidir.

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} \Delta w_{ij}^{(t)} \quad (4.26)$$

Ancak kısmi türev değişikliğinde önceki adım çok büyükse ve minimum değer kaçırılmışsa oluşabilecek hatayı önlemek için ağırlık güncellemesi bir adım geri alınır.

$$\Delta w_{ij}^{(t)} = -\Delta w_{ij}^{(t-1)}, \text{ eğer } \frac{\delta E^{(t-1)}}{\delta w_{ij}} * \frac{\delta E^{(t)}}{\delta w_{ij}} < 0 \quad (4.27)$$

(4.27) denklemindeki geri adım işlemi nedeniyle bir sonraki aşamada türevin işareti tekrar değişebilir. Güncelleme değerinin iki kez aynı işleme maruz kalmasını engellemek için sonraki adımda güncelleme değerine uyarılama yapılmamaktadır. Bunu sağlayacak bir eşitlik kullanılmaktadır. Sonuç olarak hem öğrenmeyi hem de

yakınsamayı sağlamak için kısmi türevin işaretinden faydalanılmış olmaktadır. Sigmoid aktivasyon fonksiyonunda üssel azalmadan dolayı çıktı katmanına daha uzak ağırlıklar daha yavaş değişir bu da sistemi yavaşlatır. RPROP algoritmasında ise türevin büyüklüğü değil yönü yani işareti kullanıldığı için tüm ağ öğrenmek için eşit şansa sahiptir. Bu diğer algoritmalara göre hem daha hızlı hem de güçlü bir sonuç doğurmaktadır (Riedmiller ve Braun 1993).

4.2.6 Yapay Sinir Ağlarının Avantajları ve Dezavantajları

Yapay Sinir Ağları (YSA) diğer geleneksel modellere göre oldukça fazla avantaja sahiptir. Bunlardan bazıları;

- Doğrusal olmayan karmaşık problemlerin çözümünde kullanılmaktadır.
- Seri halinde olan sistemlerin bir bölümündeki yavaşlığın diğer bölümleri de etkilemesi durumu YSA'daki paralel yapı sayesinde yerini daha hızlı bir yapıya bırakmaktadır.
- YSA sürekli öğrenme ve uyum sağlayabilme özelliği sayesinde tekrar tekrar veri setlerine göre değiştirilmeden başka veri setleri için de adapte edilebilmektedir.
- Bilgi klasik yöntemlerdeki gibi veri tabanlarında veya kodlarda değil nöronlardaki bağlantılarda ağa dağıtılmış şekilde saklanmaktadır.
- Parametre ve değişken sayısı kısıtlaması olmadan çalışabilmektedir.
- Hata toleransı ile çalışma yeteneği klasik yöntemlere göre oldukça yüksektir (Doğan 2020).

Bunların dışında YSA uygulamalarının bazı dezavantajları da bulunmaktadır. Bunlardan bahsetmek gerekirse;

- Parametrelerin belirlenmesi konusunda belirlenmiş kesin kuralların olmaması ve bu işlemin kullanıcı deneyimine bırakılması en iyi çözümlerin bulunmasını zorlaştırabilmektedir.

- YSA numerik verilerle çalışabilmektedir. Bu yüzden verilerin numerik dönüşümü sırasında yapılan hatalar hatalı bir modelin oluşumuna sebep olabilmektedir (Eşidir 2020).
- Öğrenmeye dayalı bir sistem olduğu için bazı verilerin aşırı öğrenilmesi sonucu ezberleme sorunu ile karşılaşılabilir.
- Farklı birimlere sahip veriler kullanılamaz. Bu yüzden verilerin normalizasyon vb. ön işleme durumlarından geçmeleri gerekmektedir.
- Sistem içerisinde eksik veri barındırılmasına izin vermemektedir.

4.2.7 Yapay Sinir Ağlarının Kullanım Alanları

Yapay Sinir Ağları (YSA) kompleks problemlerin çözümünde iyi bir alternatiftir. Öğrenme özelliği sayesinde geçmiş verilerden yararlanılarak gelecek tahmini yapılabilmektedir. Yalnızca sayısal verilerle çalışabilme özelliğine sahip, bilgiyi hücreler içerisinde saklayabilen YSA gelecek verilerinin tahmini, verilerin sınıflandırılması, görüntü işleme gibi konularda finanstan, mühendisliğe, tıptan üretime birçok alanda kullanılabilir. YSA sayesinde sistemlerdeki arıza tespitleri, ekonomik verilerin tahmini, kredi değerlendirmeleri, risk analizi, kanser hastalığının analizi, nesnelerin görüntü işlemeyle ayırt edilmesi, ürünlerin analizi, sınıflandırılması, veri madenciliği vb. birçok karmaşık konu çözüme kavuşturulmaktadır. Temelde en çok kullanıldığı alanlar ise sayısal tahminleme, verilerin sınıflandırılması, ilişkilendirilmesi, filtrelenmesi ve yorumlanması olarak belirtilmiştir (Ağyar 2015).

Genel olarak bahsetmek gerekirse YSA kişilerin yazım tarzının ayrıştırılmasından, saflık testlerine kadar çok geniş alanlarda kullanılabilir. Üretimde, kalite kontrolü, ürün dizaynı, planlama, makine arıza onarım takibinde, sağlıkta protez dizaynı, nakil gibi konularda, güvenlikte görsel veri işleme gibi konularda, finasta para, borsa, yatırım analizleri gibi çok çeşitli alanlarda uygulamalarına rastlanmaktadır (Demir 2007).

4.3 Gradyan Arttırmalı Regresyon Ağaçları

Topluluk algoritmaları öğrencilerin birbirinden bağımsız olarak eğitildiği ve tahmin sonuçlarının tek bir model altında toplandığı algoritmalarıdır. Topluluk algoritmalarında iki ana modelden bahsedilmektedir. İlki rastgele orman yöntemini de içerisinde barındıran torbalama modelidir. Bu modelde öğrenciler basit ortalama teknikleri kullanılarak birleştirilir ve rastgele alt örneklerle çalışır. İkinci model olan arttırma modelinde ise tahminler rastgele değil sıralı olarak önceki tahminlerden sonraki tahminlerin çıkarıldığı bir sistemle çalışmaktadır. Gradyan arttırmalı algoritmalar bu modellere örnek olarak verilmektedir (Grover 2017). Yaygın olarak kullanılan gradyan arttırma yöntemlerinin bazıları;

- Gradyan Arttırma Makineleri
- AdaBoost
- Gradyan Arttırmalı Regresyon Ağaçları (GBRT)
- XGBoost şeklinde sıralanabilmektedir.

Friedman (1999) tarafından ortaya konulan ve bir makine öğrenmesi algoritması olan gradyan arttırmalı ağaçlar, başta zayıf olarak üretilen bir modelin daha sonra hata fonksiyonlarına göre davranarak modeli iyileştirmesi prensibine dayanır. Başka bir şekilde açıklamak gerekirse güçlü öğrencilerin oluşturulabilmesi için zayıf öğrencilerin ürettikleri tahminlerden yararlanan bir yöntemdir.

Bu model regresyon üzerinde de değerlendirilebilir. Burada kayıp fonksiyonunda bulunan negatif yönlü gradyan ile en yüksek ilişkiye sahip güncel bir öğretici oluşturulmak istenmektedir. Hata oranını belirlemede ortalama hata kareleri yöntemi kullanılmaktadır (Gökler 2020; Zhou ve diğ 2017).

4.3.1 Regresyon Ağaçları

Aslında sınıflandırma ve regresyon için kullanılan karar ağaçları kompleks olan verileri çok aşamalı bir hale getirerek basit bir karar süreci oluşturma sistemine dayanmaktadır. Eğer sistemde kategorik değişken varsa sınıflandırma, sürekli değişken var ise regresyon ağacı adı verilmektedir. Regresyon ağacı içerisinde basit

olarak düğüm, dal ve yaprak olmak üzere üç adet unsur bulunmaktadır. Kök düğümünden başlanmak üzere karar sistemi üzerine yapraklara kadar dallanma devam etmektedir. Bu sistemdeki en önemli faktör ağacın hangi kritere göre dallanacağıdır. Regresyon ağaçları için yaprak kısmındaki tahmin edilecek bağımlı değişken sürekli bir değer almaktadır. Twoing algoritması ile ortalama hata karelerini en küçüklemek için bölünmeler hesaplanarak dallanma gerçekleştirilir. Her yaprak için tahmin değerleri düğümdeki ağırlıklı ortalama değerlere göre hesaplanmaktadır. Twoing algoritmasında eğitim verileri iki alt dala ayrılır ve regresyon verilerinin iki daldan hangisine ait olabileceği olasılığı hesaplanır. Daha sonra t düğümü içerisindeki aday bölünmeler için uygunluk değeri hesaplanır. Uygunluk değeri içerisinde en büyük değer seçilip o değer temsil ettiği aday bölünme satırı dallanmayı oluşturur. Böylece ilk dallanma oluşmuş olur ve daha sonra işlemler tekrarlanır (Kavzoğlu ve diğ. 2012).

Regresyon ağaç modelinde amaç bağımlı değişken için bağımsız değişkenlerin bir matrisini oluşturup alt gruplar elde etmektir. Dallanmalardan oluşan bu sistemde kritik değerler belirlenir ve yapraklarda bağımlı değişkenin değerleri gösterilir. İlk düğümünden (kök) son düğüme kadar (yaprak) oluşturulan hatlarda maksimum ayırım minimum varyasyon kuralı işlemektedir. Regresyon ağaçlarında iki ayrımlı düğümler için tahminlerin toplam varyansının minimum olması istenir. Bu yüzden her düğümden artıkların karelerini azaltma yöntemi uygulanır (Özkan 2012).

4.3.2 Gradyan Arttırmalı Regresyon Ağaçlarının İşleyişi

Gradyan arttırma yöntemi genel olarak üç unsur barındırmaktadır. Bunlar;

- Kayıp fonksiyonu (optimizasyon için)
- Zayıf öğrenciler (tahmin için)
- Yeni zayıf öğrenciler (hata oranını azaltmak için) unsurlarından oluşmaktadır.

Bu yöntemde karar ağaçları yerine zayıf öğrenciler kullanılmaktadır. İterasyonlar boyunca yeni zayıf öğrenciler eklenerek hatanın azalması amaçlanır (Soylu 2018).

Gradyan Arttırımlı Regresyon Ağaçları (GBRT) yönteminde tahminler bağımsız değil sırayla yapılır ve topluluk algoritmaları sınıfına dahil edilmektedir. İlk iterasyonda tahmin için bir fonksiyon oluşturulur. Tahmin ile istenilen değer arasındaki sapmalardan Ortalama Hata Karelerini (MSE) hesaplayan fonksiyon tahmin fonksiyonuyla birleştirilip tekrar sapmalar hesaplanır ve hata oranı sıfıra yaklaştırılmaya çalışılır.

$$MSE = \sum (y_i - \hat{y}_i)^2 \quad (4.28)$$

(4.28) kayıp fonksiyonunda y_i istenilen değer, \hat{y}_i ise tahmin değeridir. Bu denkleme öğrenme oranı (α) dahil edilerek yeni formül oluşturulur.

$$\hat{y}_i = \hat{y}_i - \alpha * 2 * \sum (y_i - \hat{y}_i) \quad (4.29)$$

Böylece (4.29) denkleminde öğrenme oranı iterasyonlar boyunca güncellenerek kayıp fonksiyon değerinin minimuma yaklaşması sağlanmaktadır (Keleş ve diğ. 2020).

GBRT algoritmasında ana amaç her adımda istenilen değer ile zayıf öğrencilerin yaptıkları tahmin değerleri toplamındaki farka bir regresyon ağacı eklemektir. Yani burada zayıf öğrenciler olarak sabit boyutlarda karar ağaçları algoritması seçilmektedir. GBRT algoritması ağırlıklandırılmış zayıf öğrencilerin toplamı ile fonksiyonu bulmaya çalışır.

$$F(x) = \sum_{n=1}^N \beta_n h(x; a_n) \quad (4.30)$$

(4.30) denkleminde N zayıf öğrencilerin sayısı, β her bir öğrenci için ağırlık değeri, $h(x; a_n)$ parametre kümesi a_n tarafından tanımlanan küçük bir regresyon ağacıdır. Diğer arttırma algoritmalarına benzer şekilde, GBRT denklem (4.31)'deki gibi bir ek model oluşturur.

$$F_n(x) = F_{n-1}(x) + \beta_n h(x; a_n) \quad (4.31)$$

Böylece her aşamada daha iyi bir model oluşturmak için önceki fonksiyonun üzerine $h(x; a_n)$ tahmincisi eklenerek ilerlenir. Karar ağacı ve ağacın ağırlık değeri β_n kayıp fonksiyonunun değerini en aza indirmek için eğitilmektedir.

$$[h(x; a_n), \beta_n] = \operatorname{argmin} \sum_{i=1}^M L(y_i, F_{n-1}(x_i) + \beta_n h(x_i; a_n)) \quad (4.32)$$

$L(y, F(x))$ burada kayıp fonksiyondur ve (4.32) zor bir optimizasyon problemi olduğu için gradyan iniş algoritması ile kolay bir hale getirilmelidir. Gradyan iniş algoritmasını uygulamak için $h(x; a_n)$ eğitim seti $\{x_i, g_n(x_i)\}$ kullanılarak eğitilir. Burada $g_n(x_i)$ (4.33) denkleminde hesaplanan kayıp fonksiyonun negatif gradyanıdır.

$$g_n(x_i) = \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{n-1}(x)} = [y_i - F_{n-1}(x_i)] \quad (4.33)$$

(4.33) denklemin sağ tarafı MSE ile değiştirilerek hesaplanır (Dabiri ve Abbas 2018).

Belirli koşullar altında uygulanabilen farklı kayıp fonksiyonlar vardır. Bu seçim yapılırken ilk olarak aykırı değerlere karşı tutum göz önüne alınmaktadır. Kayıp fonksiyonu çıktı değişkeninin özelliklerine göre değişebilmektedir. Zaman serisi analizleri için de farklı algoritmalar üretilmiştir. Sürekli değişkenlerde en çok kullanılan kayıp fonksiyonları aşağıdaki gibidir:

- Gauss Kayıp Fonksiyonu
- Laplace Kayıp Fonksiyonu
- Huber Kayıp Fonksiyonu
- Nicelik Kayıp Fonksiyonu

Makine öğrenmesi modellerinin ön önemli konularından birisi ise modelin genelleme yeteneğidir. Eğer algoritma iyi bir genelleme modeli kuramazsa veriler aşırı öğrenme problemiyle karşı karşıya kalacaklardır. Gradyan arttırma ağaçları temelli modeller için geliştirilen en basit düzenleme yöntemi alt örneklemedir. Bu yöntemin temelinde modellemeye rastgelelik katmak yatmaktadır. Her iterasyonda belirlenen eğitim verilerinin rasgele olarak sadece bir kısmı kullanılır. Diğer bir yöntem olan

büzülme yönteminde ise hatalı verilerin büyüklüğü düşürülerek hataya göre olan modellemede büyük hataların etkisi azaltılmaktadır. Böylece daha çok ve daha küçük adımlarla daha doğru sonuçlar hedeflenmektedir. Son ve en çok tercih edilen yöntem ise erken durdurma ya da diğer adıyla önceden budamadır. Bu yöntemle eğitim ve test setleri arasındaki hata farkları önceden belirlenen durdurma değerine geldiğinde sistem durdurulur (Natekin ve Knoll 2013).

GBRT algoritmaları için en önemli parametreler aşağıdaki gibidir;

- Öğrenme oranı
- Ağaç sayısı
- Ağaç derinliği
- Alt örnekleme

Gradyan inişi yöntemiyle algoritmanın yakınsama uzunluğunu öğrenme oranı aracılığıyla kontrol edilir. Büyük değerler verilmesi sistemi yerel bir çözüme itebilmektedir. Ağaç sayısı kullanılan zayıf öğrencilerin kayıp fonksiyonu en küçüklemesi için kullanılmaktadır. Ağaç sayısının artması sistemi karmaşık bir hale getirip yavaşlatabilir. Ağaç derinlikleri ise her bir ağacın ne kadar dallanabileceğiyle ilgili bir parametredir. Bu konularla ilgili her ne kadar önleyici yöntemler bulunsa da Hastie ve diğ. (2009) tarafından yapılan çalışmada ağaç derinliğiyle ilgili boyutu 2 olan ağaçların performansının genel olarak düşük olduğu ve 10'dan büyük bir derinliğe ise ekstrem koşullar dışında ihtiyaç olmayacağı belirtilmektedir. Alt örnekleme ise eğitim için belirlenen veri setinin içerisinden her seferinde belirli bir miktarının işlenmek ve sınanmak için alınması işlemidir. Böylece tüm verilerin etkisi altında kalmayan model ezberleme sorunundan uzaklaşmaktadır.

4.3.3 Gradyan Arttırmalı Regresyon Ağaçları Modelinin Avantajları ve Dezavantajları

Sıradan ağaç modellerinde yüksek varyansa sahip ağacın ortalaması alınmaktadır. Gradyan arttırmalı ağaç modelinde ise az sayıda ağaç sırayla sisteme dahil edilmektedir. Böylece karar ağaçlarının neden olabileceği aşırı öğrenme (ezberleme) sorununun önüne geçilmeye çalışılmaktadır (Mohan ve diğ. 2011).

Genel olarak GBRT avantajlarını özetlemek gerekirse;

- Parametrik olmadıkları için varsayımlarla uğraşılmasına gerek yoktur.
- Kategorik ve sürekli verilerle karışık olarak çalışabilir.
- Çoklu doğrusallık sorunu modelin performansını etkilemez.
- Uyumunun yetersiz olduğu zayıf öğreniciler tahminlerini daha güçlü bir öğrenici ile birleştirdikleri için modelin varyansı azaltılır.
- Dolayısıyla ezberleme problemine karşı bir önlem alınmış olunur.
- Girişlerin ön işlenmesi ve ölçeklenmesine gerek yoktur.
- Aykırı ve gürültülü değerler iyi işleneceği için genel performansı az etkileyecektir.
- Eksik verilerle çalışabilmektedir.

Bunun dışında diğer makine öğrenmesi yöntemleri gibi bazı dezavantajları da bulunmaktadır.

- Problem birçok ağaç yapısıyla çalıştığından bellek kullanımı açısından sıkıntılar doğurup sistemi yavaşlatabilir.
- Parametrelerin belirlenmesinde ağaç sayısı, ağaç derinliği gibi karmaşık yapılar olduğundan ve herhangi bir kesin kural olmadığından erken uydurma, yakınsama süresini uzatma gibi sorunlarla karşılaşılabilir.
- Diğer yöntemlere göre daha karmaşık parametrelerle çalışılmaktadır.

4.3.4 Gradyan Arttırmalı Regresyon Ağaçlarının Kullanım Alanları

Bu yöntem ile çok farklı alanlarda normal karar ağaçları modellerine nazaran daha etkili sonuçlar elde edilmektedir. Özellikle gradyan arttırma ağaçları olarak bakıldığında gereksiz maillerin saptanması, eşleştirme, satış tahmini, kullanıcı davranışı tahmini, metin sınıflandırılması, ürünleri kategorilere ayırma, risk tahmini, güvenli yazılım tahmini vb. pek çok konuda Kaggle tarzı yarışmalarda diğer yöntemlere göre üstünlük sağladığı görülmektedir (Abar 2020; Grover 2017).

Özellikle veri bilimi, mühendislik gibi alanlarda parametre yönteminin iyi yapılması koşuluyla çok verimli sonuçlar alınabileceği düşünülen bir yöntemdir. GBRT algoritması genel olarak sıralama, sınıflandırma ve regresyon tabanlı problemlerin çözümünde kullanılmaktadır. Sağlık, tarım, veri bilimi, üretim, yazılım sistemleri vb. birçok alanda kullanılabilir.

5. UYGULAMA VE SONUÇLAR

5.1 Araştırma Tasarımı

Uygulamanın ilk aşamasında toplanan değişkenlerle bir çoklu doğrusal regresyon modeli kurulmuştur. Modelin kurulması aşamasında bir önceki bölümde bahsedilen ana gerekliliklerden normallik varsayımı, değişkenler arası doğrusal ilişki analizleri değerlendirilmiştir. Normallik analizi için Tabachnick ve Fidell (2013) tarafından belirlenen çarpıklık ve basıklık değerlerinden yararlanılmıştır. Doğrusal bağlantı için ise Pearson Korelasyon yöntemi kullanılmıştır.

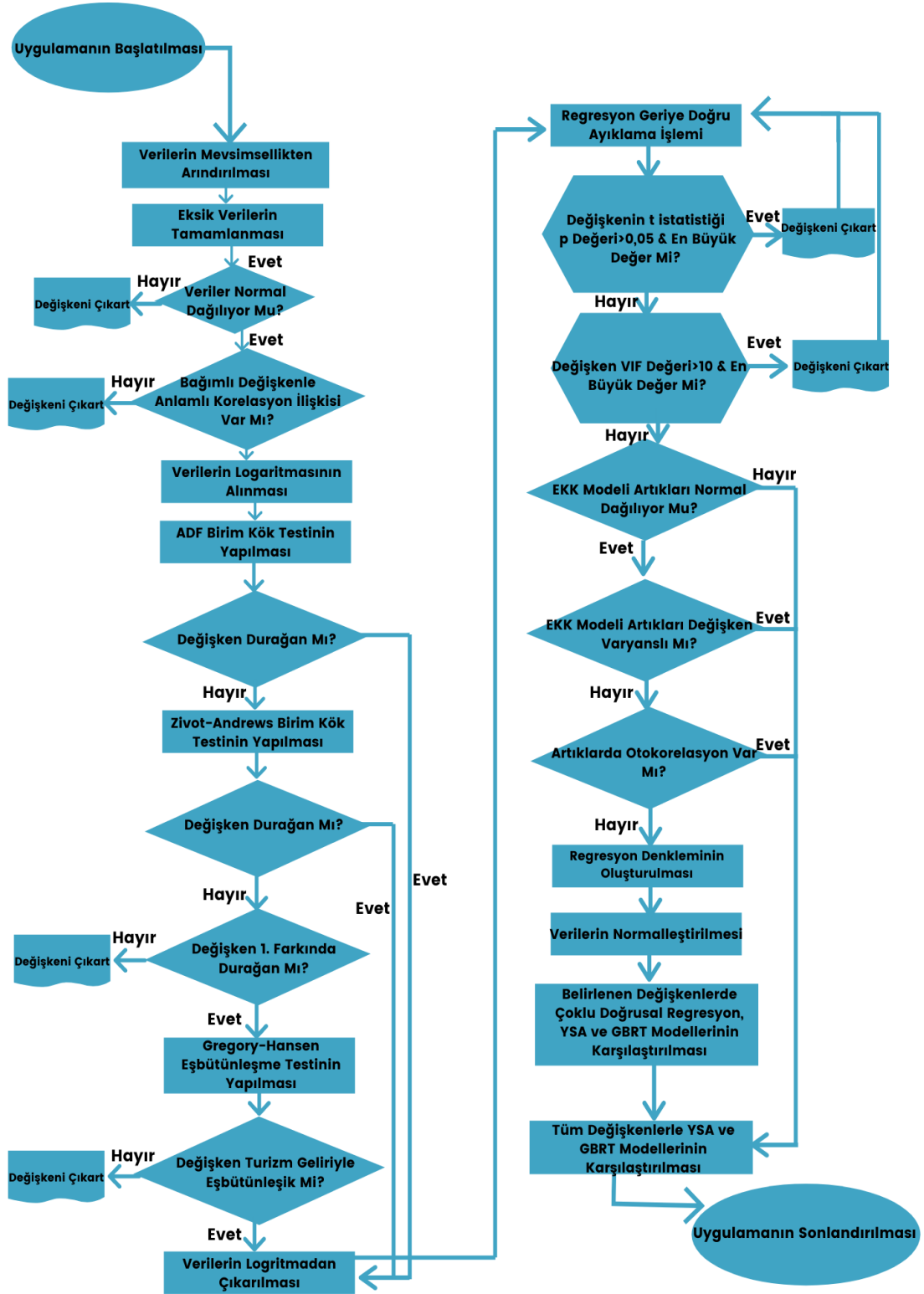
Bir sonraki aşamada Granger ve Newbold (1974) tarafından ortaya atılan sahte regresyon probleminin nedeni olan verilerin durağanlık durumu kontrol edilmiştir. Durağanlık Dickey ve Fuller (1981) tarafından geliştirilen Genişletilmiş Dickey Fuller (ADF) birim kök testi ile kontrol edildikten sonra durağan olmayan yani birim kök barındıran değişkenler için verilerin içerisinde bulunan ekonomik krizler vb. olayların göz ardı edilmemesi adına yapısal kırılmaya izin veren Zivot ve Andrews (1992) birim kök testi uygulanmıştır. Bu iki birim kök testi sonucu durağan olmayan verilerin hepsi birinci farklarında durağanlaştığı için bir sonraki aşama olan eşbütünleşme testlerine geçilmiştir.

Eşbütünleşme testlerinde yine tek yapısal kırılmaya izin veren bir yöntem olan Gregory ve Hansen (1996) eşbütünleşme testi uygulanmıştır. Test sonucu turizm geliriyle aralarında uzun dönemli ilişki bulunan değişkenler belirlenerek daha sonra regresyon modelinde kullanılmak üzere değerlendirilmiştir.

Regresyona sokulan değişkenler geriye doğru ayıklama yapılarak adım adım en düşük t istatistiği anlamlılık değerine sahip veriler elenerek ve kalan değişkenlerin çoklu doğrusallıktan kurtulması için Varyans Şişirme Faktörü (VIF) değeri 10'dan büyük en yüksek değer çıkartılarak aynı işlemler anlamlılık olasılık değerleri 0.05'in ve VIF değeri 10'un altına düşürülünceye kadar tekrarlanmıştır. Daha sonra Jarque-Bera testi ile kalıntıların normallik değerleri, Breusch-Pagan-Godfrey testi ile değişen varyans değerleri ve Wallis (1972) testi ile otokorelasyon değerleri kontrol edilerek kurulan modelin sağlaması yapılmıştır.

Elde kalan deęişkenlerle regresyon denklemi kurulup çeşitli parametre kombinasyonlarında Esnek Geri Yayılım (RPROP) algoritmasına sahip Yapay Sınır Aęı (YSA) ve Gradyan Arttırılmış Regresyon Aęaçları (GBRT) modelleri ile karşılaştırılmıştır. Karşılaştırma yapılırken hata deęeri olarak birim farklılıklarından ve yapısal kırılmalardan etkilenilmemesi adına Ortalama Mutlak Yüzde Hata (MAPE) yöntemi kullanılmıştır. Bunun yanı sıra R^2 açıklayıcılık deęerine de dikkat edilmiştir.

Son olarak makine öğrenmesi yöntemlerinin herhangi bir ön koşul gerektirmemesi nedeniyle tüm veriler RPROP algoritmalı YSA ve GBRT modelleriyle karşılaştırılarak deęişimler gözlemlenmiştir. Yukarıda anlatılan tüm analizlerin genel akış gösterimi aşağıda Şekil 5.1 ile görselleştirilmiştir.



Şekil 5.1: Uygulamanın akış şeması üzerinde gösterimi.

5.2 Araştırmanın Sınırları ve Veri Derleme Süreci

Çuhadar (2020)'in da çalışmasında belirttiği gibi önceki veri sisteminde banka kayıt yöntemiyle toplanan turizm verileri yabancı pasaportlu kişilerin bozdurdıkları döviz üzerinden hesaplanmaktadır. 2003 yılından itibaren “Çıkış Yapan Ziyaretçi Anketi” adı altında yurtdışında ikamet eden 15 yaş ve üstü yabancı ve yurtdışında yaşayan vatandaşlara çeyrek yıllık olarak anketler uygulanmaktadır (Çuhadar 2020). Bu veriler içerisinde geliş nedenlerinden, mesleki durumlarına kadar pek çok detay bulunmaktadır. Dolayısıyla detaylı ve doğru verilere ulaşabilmek adına veri seti 2003-2019 yılları arası çeyrek yıllık verilerden oluşturulmuştur. Bağımlı değişken olarak yabancı (yurtdışında yaşayan vatandaşlar dahil olmadan) turizm geliri verileri TÜİK (2020) üzerinden alınmıştır. Bağımsız yani açıklayıcı değişkenler için ise ekonomi (GSYİH, ihracat, ithalat, dolar, euro, altın, dış yatırım, enflasyon, turizm gideri, işsizlik) turizm kapasitesi (dış hatlar uçuş trafiği, yatak kapasitesi, acente sayısı, teşvik tutarları), özgürlük ve ülke güvenliği (terör ve basın özgürlük endeksleri, tutuklu sayıları) ülkenin gelişmişlik düzeyi (kentsel nüfus oranı, insani gelişmişlik ve beşeri sermaye endeksleri) olmak üzere 4 ana başlık üzerinden 20 adet değişken belirlenmiştir.

Bu değişkenlerin verilerinin toplanması aşamasında çeşitli kaynaklardan yararlanılmıştır. Tablo 5.1'de de görüldüğü gibi bazı veriler yıllık olarak elde edilmiştir. Bunun nedeni ise insani gelişmişlik endeksi, basın özgürlük endeksi vb. verilerin yılda sadece bir kere yayınlanmasıdır. Bu tür veriler çeyreklik yıllar için aynı oranda kullanılacaktır. Endeks değerinin 10 üzerinden hesaplandığı bilinen küresel terörizm endeksi Türkiye'nin yıllar içerisinde geniş dönemlerde terör saldırılarına maruz kalması ve yapısal kırılmalarının turizm geliriyle daha iyi açıklanabilmesi nedeniyle ters çevrilerek terörün azlığı durumlarını yansıtmaya sağlanmıştır. Açıklayıcı değişkenlerin genel özellikleri Tablo 5.1'de gösterilmektedir. Bunun dışında yine açıklayıcı değişkenlerin alıntılı olduğu linkler ve asıl kaynakları Tablo 5.2'de gösterilmektedir.

Tablo 5.1: Bağımsız değişkenlerin genel özelliklerinin gösterimi.

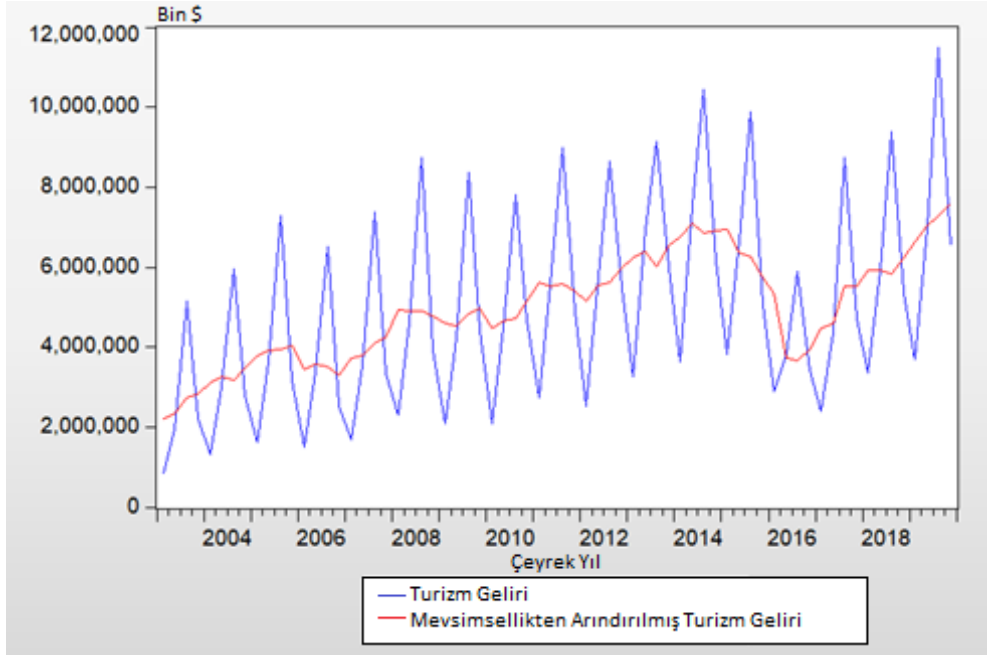
EKONOMİ			
Değişken	Kısaltma	Birim	Aralık
Mevsimsellikten Arındırılmış Gayri Safi Yurtiçi Hasıla (Sabit Fiyatlarla Harcamaya Göre)	GSYİH	2000 yılı zincirlenmiş ulusal para birimi	2003Q1-2019Q4
Mevsimsellikten Arındırılmış İhracat (Sabit Fiyatlarla Harcamaya Göre)	İHRACAT	2000 yılı zincirlenmiş ulusal para birimi	2003Q1-2019Q4
Mevsimsellikten Arındırılmış İthalat (Sabit Fiyatlarla Harcamaya Göre)	İTHALAT	2000 yılı zincirlenmiş ulusal para birimi	2003Q1-2019Q4
Tüketici Fiyat Endeksi (Tüm Kalemler)	TÜFE	Endeks 2015=100	2003Q1-2019Q4
Külçe Altın Gram Satış Fiyatı (Düzy Ortalama)	ALTIN	Türk Lirası	2003Q1-2019Q4
USD Satış Fiyatı (Düzy Kur)	USD	Türk Lirası	2003Q1-2019Q4
EURO Satış Fiyatı (Düzy Kur)	EURO	Türk Lirası	2003Q1-2019Q4
Uluslararası Doğrudan Yatırım	YATIRIM	Milyon USD	2003Q1-2019Q4
İşsizlik Oranı	İŞSİZLİK	Yüzdesel Değer	2003Q1-2019Q4
Turizm Gideri	TGDR	Bin USD	2003Q1-2019Q4
TURİZM KAPASİTESİ			
Değişken	Kısaltma	Birim	Aralık
Türkiye Seyahat Acenta Sayıları	ACENTE	Adet	2003-2019
Turizm Tesis Yatak Sayısı	TYATAK	Adet	2003-2019
Turizm Yatırım Teşvik Belgeleri Tutarı	TEŞVİK	Milyon TL	2003-2018
Türkiye Dış Hatlar Uçuş Trafığı	DIŞHAT	Adet	2003Q1-2019Q4
ÜLKE GELİŞİMİŞLİĞİ			
Değişken	Kısaltma	Birim	Aralık
Kişi Başı Beşeri Sermaye Endeksi	HCI	Endeks	2003-2017
İnsani Gelişmişlik Endeksi	HDI	Endeks	2003-2018
Kentsel Nüfus Oranı	KENTN	Yüzdesel Değer	2003-2019
ÜLKE ÖZGÜRLÜĞÜ VE GÜVENLİĞİ			
Değişken	Kısaltma	Birim	Aralık
Ters Yönlü Küresel Terörizm Endeksi	TERORİT	Endeks	2003-2019
Tutuklu ve Hükümlü Sayısı	TUTUKLU	Kişi	2003-2019
Basın Özgürlük Endeksi	BASIN	Endeks	2003-2019

Tablo 5.2: Bağımsız değişkenlerin kaynak gösterimi.

Değişken	Asıl Kaynak	Kaynak Alınan Yer
Mevsimsellikten Arındırılmış Gayri Safi Yurtiçi Hasıla (Sabit Fiyatlarla Harcamaya Göre)	OECD	https://fred.stlouisfed.org
Mevsimsellikten Arındırılmış İhracat (Sabit Fiyatlarla Harcamaya Göre)	OECD	https://fred.stlouisfed.org
Mevsimsellikten Arındırılmış İthalat (Sabit Fiyatlarla Harcamaya Göre)	OECD	https://fred.stlouisfed.org
Tüketici Fiyat Endeksi (Tüm Kalemler)	OECD	https://fred.stlouisfed.org
Külçe Altın Gram Satış Fiyatı (Düzyer Ortalama)	TCMB	https://evds2.tcmb.gov.tr
USD Satış Fiyatı (Düzyer Kur)	TCMB	https://evds2.tcmb.gov.tr
EURO Satış Fiyatı (Düzyer Kur)	TCMB	https://evds2.tcmb.gov.tr
Uluslararası Doğrudan Yatırım	TCMB	https://evds2.tcmb.gov.tr
İşsizlik Oranı	TÜİK	https://data.tuik.gov.tr/
Turizm Gideri	TÜİK	https://data.tuik.gov.tr/
Türkiye Seyahat Acenta Sayıları	T.C. KÜLTÜR VE TURİZM BAKANLIĞI	https://yigm.ktb.gov.tr/
Turizm Tesis Yatak Sayısı	T.C. KÜLTÜR VE TURİZM BAKANLIĞI	https://yigm.ktb.gov.tr/
Turizm Yatırım Teşvik Belgeleri Tutarı	T.C. SANAYİ VE TEKNOLOJİ BAKANLIĞI	http://ttyd.org.tr/
Türkiye Dış Hatlar Uçuş Trafiği	DHMİ	https://dhmi.gov.tr/
Kişi Başı Beşeri Sermaye Endeksi	University of Groningen, University of California, Davis	https://fred.stlouisfed.org
İnsani Gelişmişlik Endeksi	BİRLEŞMİŞ MİLLETLER GELİŞME PROGRAMI	https://countryeconomy.com/
Basın Özgürlük Endeksi	REPORTERS WITHOUT BORDERS	https://knoema.com/
Küresel Terörizm Endeksi	INSTITUTE FOR ECONOMİCS & PEACE	https://visionofhumanity.org/
Tutuklu ve Hükümlü Sayısı	TÜİK	https://data.tuik.gov.tr/
Kentsel Nüfus Oranı	WDB	https://data.worldbank.org/

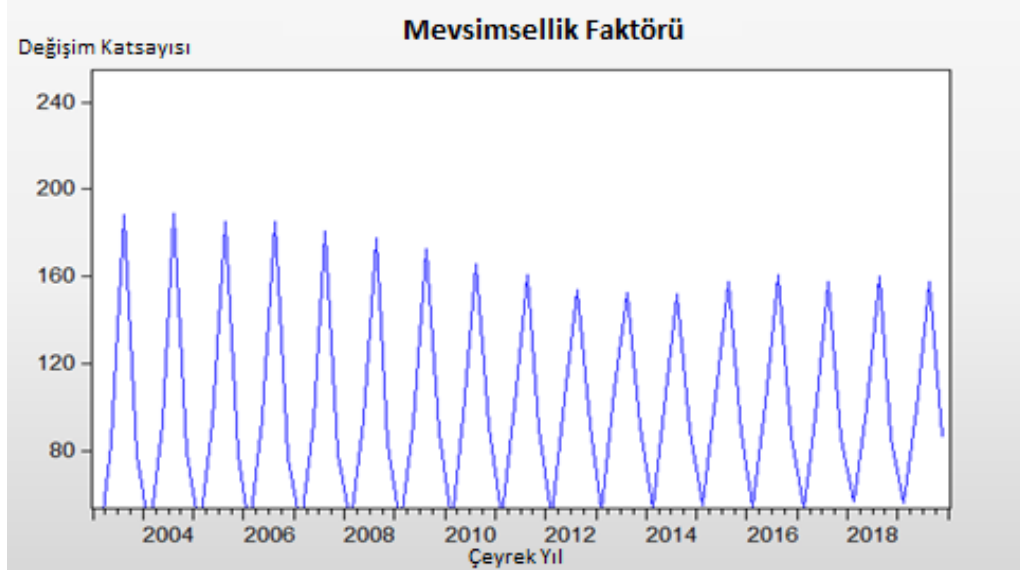
Veriler Eviews 10.0 eklentisi Tramo/Seats modülü ile mevsimsellikten arındırılmış ve eksik veriler tamamlanmıştır. Tramo/Seats modülünde otoregresyon, fark değerleri ve hareketli ortalama değerleri içsel olarak belirlenip hem verilerin mevsimsellik katsayıları belirlenir hem de zaman serisi tahminleri oluşturulur. Daha

sonra uygulamanın ilk adımı olan ekonometrik analizlerin daha esnek sonuçlar verebilmesi ve birim farklılıklarından etkilenmemesi adına tüm verilerin logaritmaları alınmıştır. Regresyon analizleri sırasında logaritmadan çıkarılan veriler makine öğrenmesi denemeleri yapılırken min-max normalizasyon yöntemiyle 0-1 arası değerlerle ifade edilmiştir. Böylece işlemler sırasındaki birim farklılıklarının doğuracağı olumsuz sonuçlar engellenmiştir. Aşağıda Şekil 5.2’de ham turizm geliri verileri ve verilerin mevsimsellikten arındırılmış hali gösterilmektedir.



Şekil 5.2: Turizm geliri için mevsimsellikten arındırma grafiği.

Şekil 5.2’de de görüldüğü gibi veriler bu yöntemle mevsimsel dalgalanmalardan kurtarılmıştır. Bu analiz sürecinde daha doğru sonuçlara ulaşmamızı sağlayacaktır. Aşağıda ise turizm gelirinin ayıklanan mevsimsellik faktörü Şekil 5.3 ile görülmektedir.



Şekil 5.3: Turizm gelirin mevsimsellik faktörü grafiği.

Şekil 5.3'te açıkça görüldüğü üzere 2003-2012 yılları içerisinde turizm gelirin mevsimsel farklarının düşüşte olduğu ve bu düşüşün 2012 yılı sonrası durağanlaştığı görülmektedir. Buradan 2003 yılında turizmin 2012 yılına oranla daha tekil ve büyük oranda yaz turizmi kaynaklarından beslendiği ve 2012 yılı sonrasında ise bahsedilen bu tekil turizme dair bir değişim olmadığı görülmektedir.

5.3 Çoklu Doğrusal Regresyon Modeli Uygulaması

Bu başlık altında regresyon modelinin tüm gereksinimleri test edilerek şartları sağlayan bağımsız değişkenler ile bağımlı değişken arasındaki ilişki bir matematiksel model vasıtasıyla açıklanmaktadır.

5.3.1 Çoklu Doğrusal Regresyon Gerekliliklerinin İncelenmesi

İlk olarak SPSS 25.0 programı ile normallik analizi ve doğrusal bağlantının olmadığı değişkenleri tespit etmek için Pearson korelasyon analizi yapılmıştır. Veriler mevsimsellikten arındırılmış ve eksik verileri tamamlanmış şekilde kullanılmıştır.

Normal bir dağılım eğrisinde sapmalar arttıkça verinin normallikten uzaklaştığı görülmektedir. Çarpıklık durumu normal dağılım eğrisindeki simetrik bozukluğu yani

grafığın sağa veya sola yönelmesini, basıklık ise grafiğin normale göre daha sivri veya daha yatay olma durumlarını göstermektedir. Hesaplama yapılırken ortalama, medyan ve standart sapma değerlerinden faydalanılır. Çarpıklık ve basıklık değerlerinin +1,5 ve -1,5 değerleri arasında olması verilerin normal dağıldığının bir göstergesi sayılmaktadır (Tabachnick ve Fidell 2013). Tüm veriler için Tablo 5.3 üzerinde SPSS 25.0 programı ile hesaplanmış çarpıklık ve basıklık değerleri gösterilmektedir.

Tablo 5.3: Değişkenlerin çarpıklık ve basıklık değerleri.

Değişkenler	Çarpıklık Değeri	Basıklık Değeri
TGLR	-0,039	-0,837
EURO	1,651	1,838
TGDR	-0,245	-0,829
TYATAK	0,205	-1,282
HDI	-0,041	-1,436
HCI	0,253	-1,189
TUTUKLU	0,715	-0,407
DISHAT	0,039	-1,339
TERORİT	-0,867	-0,165
ACENTE	0,588	-0,896
GSYİH	0,260	-1,250
İHRACAT	0,408	-0,947
İTHALAT	-0,140	-0,816
İŞSİZLİK	1,360	1,167
KENTN	-0,095	-1,205
TÜFE	0,825	-0,091
USD	1,627	1,686
YATIRIM	0,037	-1,413
TEŞVİK	-0,154	-1,551
ALTIN	1,238	1,125
BASIN	0,067	-0,313

Yukarıdaki Tablo 5.3 değerlerine bakıldığında EURO, USD ve TEŞVİK değerleri kritik değerlerin dışında kalmaktadır. Dolayısıyla regresyon modelinin bir şartı olan normallik varsayımını sağlamadıkları için sistemden çıkarılmışlardır. Pearson korelasyon analizi sonucunda turizm geliri bağımlı değişkeniyle t istatistik olasılık değeri 0,05'in üzerinde çıkan tek değer 0,227 oranla İŞSİZLİK değişkeni olmuştur. Bu da göstermektedir ki turizm geliriyle işsizlik verileri arasında anlamlı bir ilişki bulunmamaktadır.

5.3.1.1 Durağanlık Analizi

Bir zaman serisinde bağımsız değişkenler ile bağımlı değişkenler arasında gerçek bir regresyon ilişkisinin bulunabilmesi için ya verilerin seviyede durağan olmaları ya da aralarında eşbütünleşik bir ilişkinin olması gerekmektedir. Bu koşullar sağlanmadığında regresyon modelinin sahte olması beklenebilir (Granger ve Newbold 1974).

Bu başlık altındaki bütün işlemler Eviews 10.0 programı üzerinden gerçekleştirilmiştir. Durağanlık analizi, eşbütünleşme analizi sırasında tespitlerin daha esnek yapılabilmesi ve birim farklılıklarından etkilenmemesi adına verilerin logaritması alınmıştır. Modelde ilk olarak Genişletilmiş Dickey Fuller (ADF) ve yapısal kırılmaya izin veren Zivot-Andrews birim kök testlerine göre durağanlıkları sınanmıştır. Yöntem başlığı altına birim kök testlerinin nasıl yapıldığı ve formüller anlatılmıştır. Boş hipotez birim kökün varlığı yani serinin durağan olmadığı üzerine kurulmaktadır. Tablo 5.4 üzerinde ilk olarak yapısal kırılmaları göz ardı eden ADF birim kök testi sonucu verilerin istatistik değerleri gösterilmektedir. Değerlerden herhangi biri belirlenen kritik değerlerden küçük ise birim kök içermez yani durağandır (Dickey ve Fuller 1981).

Tablo 5.4: ADF birim kök testi sonuçları.

Modeller	Düzye			Birinci Dereceden Farkta		
	Sabitli	Sabitli ve Trendli	Sabitsiz ve Trendsiz	Sabitli	Sabitli ve Trendli	Sabitsiz ve Trendsiz
LOGTGLR	-2,28	-2,55	1,89	-6,68	-6,67	-6,45
LOGACENTE	2,95	-1,79	13,60	-8,00	-8,49	-1,38
LOGALTIN	0,48	-2,12	4,37	-7,59	-7,52	-5,95
LOGBASIN	-1,54	-2,39	0,29	-8,02	-7,96	-8,06
LOGDIŞHAT	-1,99	-1,98	5,88	-6,66	-6,80	-4,92
LOGGSYİH	-1,16	-2,02	5,35	-7,17	-7,21	-5,37
LOGHCI	3,70	-1,60	3,22	-2,31	-5,74	0,19
LOGKENTN	-2,88	2,10	-0,53	-0,06	-7,77	-4,68
LOGTGDR	-2,76	-2,85	0,53	-10,52	-10,69	-10,51
LOGTERORİT	-1,49	-2,00	-0,29	-8,02	-7,96	-8,06
LOGTÜFE	1,76	0,02	13,11	-8,02	-8,45	-0,73
LOGTYATAK	-2,09	-0,90	14,5	-8,00	-8,49	-1,38
LOGYATIRIM	-1,76	-0,73	4,64	-8,18	-8,49	-6,28
LOGİHRACAT*	-0,47	-4,07	3,76	-9,28	-9,20	-9,34
LOGİHALAT*	-2,24	-4,40	-1,33	-5,13	-5,10	-4,91
LOGTUTUKLU*	-0,65	-4,27	2,29	-2,75	-2,60	-1,47
LOGHDI*	-0,40	-1,63	-7,13	-8,15	-8,10	-1,39
0.05 p ile kritik değerler	-2,90	-3,48	-1,94	-2,90	-3,48	-1,94

Tablo 5.4'te görüldüğü üzere LOGİHRACAT, LOGİTHALAT, LOGTUTUKLU ve LOGHDI değişkenleri 0,05 önem düzeyinde kritik değerlerden küçük oldukları için verilerin durağan olduğu görülmektedir. Değişkenlerin tümünde ise birinci dereceden farkları alındığında durağanlaşma gözlemlenmektedir. Seviyede durağan olmayan verilerin yapısal kırılmalardan etkilenmemesi adına tek yapısal kırılmaya izin veren Zivot-Andrews birim kök testi Tablo 5.5'te verilmektedir. Bu testte ise ADF testine ek olarak yapısal kırılmalar içsel olarak belirlenen kukla değişkenlerle modele dahil edilmektedir. Böylece yapısal kırılmaların birim kökün varlığına etkisi ortadan kaldırılmaktadır (Zivot ve Andrews 1992).

Tablo 5.5: Zivot-Andrews birim kök testi sonuçları.

0.05 p ile kritik değerler	-4,93			-5,08		
Modeller	Sabitli			Sabitli ve Trendli		
Değişkenler	t değeri	p	Kırılma Dönemi	t değeri	p	Kırılma Dönemi
LOGACENTE	-2,76	0,03	2014Q1	-3,53	0,53	2009Q4
LOGALTIN	-4,29	0,00	2013Q1	-3,9	0,00	2013Q1
LOGBASIN*	-4,65	0,00	2009Q1	-5,16	0,00	2009Q1
LOGDIŞHAT*	-6,88	0,00	2016Q2	-5,64	0,00	2016Q2
LOGGSYİH	-3,46	0,00	2008Q2	-3,48	0,02	2008Q2
LOGHCI	-2,59	0,27	2006Q3	-11,16	0,00	2010Q1
LOGKENTN	0,016	0,00	2005Q1	-1,78	0,82	2012Q2
LOGTGDR	-4,41	0,00	2007Q1	-5,23	0,10	2009Q3
LOGTERORİT*	-5,01	0,00	2015Q1	-5,42	0,00	2015Q1
LOGTUFE	-2,76	0,057	2012Q3	-3,1	0,22	2017Q2
LOGTYATAK	-2,53	0,00	2017Q1	-4,1	0,00	2014Q1
LOGYATIRIM	-2,92	0,00	2008Q4	-4,9	0,00	2008Q4

Yukarıdaki Tablo 5.5'te görüldüğü gibi LOGBASIN, LOGDIŞHAT VE LOGTERORİT değişkenleri 0,05 önem seviyesinde birim kök içermemektedir ve bu değişkenlerin durağanlığı kabul edilmektedir. Terör endeksindeki kırılma 2015 yılının ilk çeyreğinde, basın özgürlük endeksindeki kırılma 2009 yılının ilk çeyreğinde, dış hatlar uçuş trafiğindeki kırılma ise 2016 yılının ikinci çeyreğinde gerçekleşmiştir.

5.3.1.2 Durağan Olmayan Değişkenlerin Eşbütünleşme Analizleri

Elde kalan durağan olmayan değişkenler birinci dereceden farklarında durağanlaşmaktadır. Dolayısıyla bu değişkenlerin turizm geliriyle aralarındaki uzun dönem ilişkisinin varlığının saptanabilmesi için eşbütünleşme testi yapılmalıdır. Bu doğrultuda aralarındaki uzun dönem ilişkilerin tespitinde kullanılan yöntem tek yapısal

kırılmaya izin veren Gregory-Hansen eşbütünleşme testidir. Bu test ile Yöntem başlığı altında açıklandığı gibi kurulan regresyon denklemindeki artıkların birim kök analizleri incelenmektedir. Eğer artıklar yapısal kırılmalar altında birbiriyle durağanlarsa yani birim köke sahip değilse aralarında yönü belirsiz bir uzun dönem ilişki mevcuttur (Gregory ve Hansen 1996). Verilerin içerisinde yapısal kırılmaya izin veren Gregory-Hansen eşbütünleşme testi sonuçları Tablo 5.6'da gösterilmektedir. Herhangi bir modelde ADF istatistik değerleri kritik değerlerden küçükse değişkenlerin arasında eşbütünleşik bir ilişki vardır.

Tablo 5.6: Gregory-Hansen eşbütünleşme testi sonuçları.

0.05 p ile kritik ADF değerleri	-4,61	Kırılma Dönemi	-4,99	Kırılma Dönemi	4,95	Kırılma Dönemi
Modeller	C		C/T		C/S	
LOGACENTE*	-4,84	2015Q1	-4,43	2015Q1	-5,55	2015Q3
LOGALTIN	-4,46	2015Q3	-4,59	2015Q1	-4,79	2015Q3
LOGTGDR	-3,8	2010Q2	-4,17	2015Q1	-3,26	2011Q2
LOGGSYİH	-3,64	2014Q2	-4,51	2015Q1	-4,23	2015Q3
LOGHCI*	-4,76	2015Q1	-4,51	2015Q1	-5,84	2015Q3
LOGKENTN*	-4,41	2015Q1	-4,54	2015Q1	-5,94	2015Q4
LOGTUFE*	-5,72	2015Q3	-6,09	2015Q3	-6,06	2015Q3
LOGTYATAK*	-4,4	2015Q1	-5,16	2015Q3	-6,08	2015Q4
LOGYATIRIM*	-3,28	2013Q4	-4,6	2015Q1	-4,97	2015Q4

Tablo 5.6'da görüldüğü gibi istatistik değerleri kritik değerlerden küçük olan LOGACENTE, LOGHCI, LOGKENTN, LOGTUFE, LOGTYATAK ve LOGYATIRIM değişkenleri uzun dönemli olarak LOGTGLR ile ilişkilidir. Dolayısıyla bu değerlerin turizm geliriyle aralarındaki ilişkinin gerçek bir ilişki olduğu tespit edilmiştir. Kırılma dönemi incelendiğinde ise genellikle 2015 yılı göze çarpmaktadır.

5.3.1.3 Geriye Doğru Ayıklama ve Çoklu Doğrusallık Analizi

Belirlenen durağan veriler ve uzun dönemde turizm geliriyle eşbütünleşik olan veriler bir regresyon modeli içerisine sokularak geriye doğru ayıklama işlemi ile eleme yapılmıştır. İşlem sırasında t istatistiği p olasılık değeri en yüksek olan değerler her seferinde tek bir eleme yapılarak işlem anlamsız değişken kalmayınca kadar devam etmektedir. Bu sırada denklem içerisinde bulunan çoklu bağlantılar Varyans Şişirme

Faktörü (VIF) ile tespit edilerek sistem dışı edilmektedir. Yapılan geriye doğru ayıklama işlemi ve VIF analizleri sonucunda anlamlı F ve t istatistiklerine sahip ve çoklu doğrusallık barındırmayan regresyon değerleri Tablo 5.7 üzerinde görüldüğü gibidir.

Tablo 5.7: Nihai regresyon denklemi ve t istatistik değerleri gösterimi.

Bağımsız Değişkenler	Katsayılar	Std. Hatalar	İstatistik Değeri (t)	Olasılık Değeri
DIŞHATLAR	17,27206	0,779749	22,15078	0,000
TERORİT	935446,3	90220,45	10,36845	0,000
C	-3689561	548164,8	-6,730752	0,000

Tablo 5.7’de görüldüğü üzere regresyon modeli üzerinde C sabit katsayısı modeli negatif etkilerken Ters Çevrilmiş Küresel Terör Endeksi (TERORİT) ve Dış Hatlar Uçuş Trafiği (DIŞHATLAR) modelde pozitif katsayılara sahiptir. Tablo 5.8’de ise bağımsız değişkenlerin birbirleriyle aralarındaki yüksek ilişki durumu araştırılmaktadır. Varyans Şişirme Faktörü (VIF) de daha önce Yöntem başlığında anlatıldığı gibi bu ilişki için kullanılan bir analiz yöntemidir. Eğer VIF değeri 10’un üzerindeyse değişkenler aşırı ilişkili olarak değerlendirilmektedir (Büyüküysal ve Öz 2016).

Tablo 5.8: Nihai regresyon denkleminin VIF değerleri gösterimi.

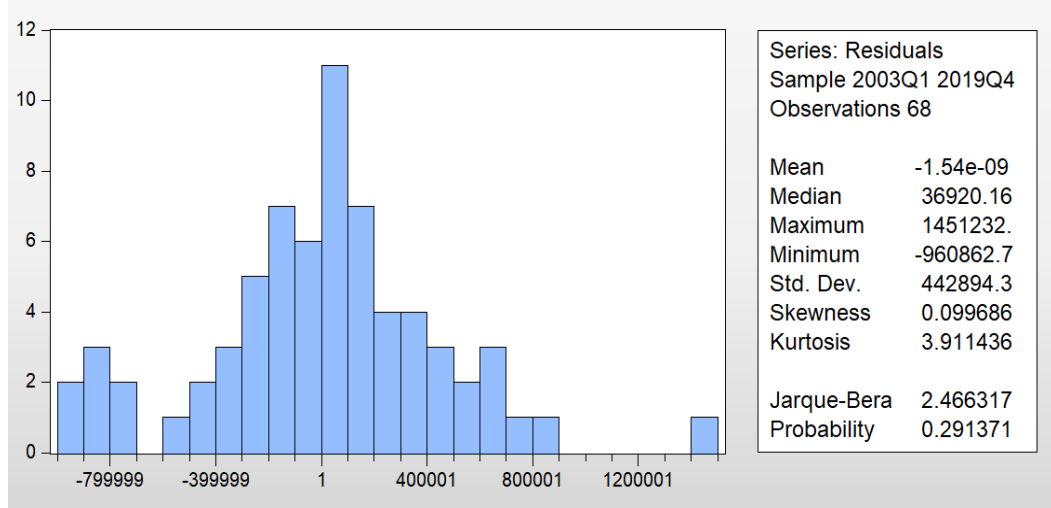
Bağımsız Değişkenler	VIF Değeri
DIŞHATLAR	1,691
TERORİT	1,691

Tablo 5.8’de de görüldüğü üzere DIŞHATLAR ve TERORİT değişkenleri için VIF değeri 10’dan küçük olduğu için değişkenler arası çoklu doğrusallık sorununun olmadığı görülmektedir.

5.3.1.4 Normal Dağılım Testi

Bu başlık altında ekonometrik testlerle desteklenen regresyon analizi sonucunda belirlenen açıklayıcı değişkenlerin uygulamada geçerli olan regresyon şartlarını sağlayıp sağlamadığına bakılacaktır. Böylece klasik istatistiki tahmin modellerinden olan En Küçük Kareler (EKK) yönteminin tam anlamıyla bu model için kullanılabilir bir istatistiki model olup olmayacağı tespit edilecektir. Şekil 5.4 ile

kurulan regresyon modelinin artıkları için normal dağılım testi yapılmaktadır. Testin görsel olarak çan eğrisi formuna yakın olması ve t istatistiğinin 0,05 olasılık değerinden büyük çıkması istatistiksel olarak normal dağılım göstergesidir.



Şekil 5.4: Nihai regresyon denkleminin normallik dağılımı gösterimi.

Şekil 5.4 üzerinde görüldüğü üzere kurulan denklemin artıklarıyla elde edilen normallik dağılımı testi olasılık değeri 0,05 anlamlılık değerinden büyük olduğu için denklemin artıklarının normal dağıldığından bahsedilebilir. Ayrıca grafik dağılımının çan eğrisi şekline benzerliği bu durumu doğrular niteliktedir.

5.3.1.5 Varyansların Değişkenliği Testi

Kurulan regresyon modelinin artıkları ile yapılan varyans değişkenliği testi sonuçları Tablo 5.9 ile gösterilmektedir. Varyans değişkenlik testlerinde F istatistik değerinin 0,05 olasılık değerinden büyük olması kurulan modelin sabit varyansa sahip olduğunu göstermektedir. Regresyon denkleminin gereksinimlerinden biri de artıkların sabit varyansa sahip olmasıdır. Yöntem başlığında da belirtildiği gibi varyans değişkenliği analizi için bu çalışmada Breusch-Pagan-Godfrey testi kullanılmıştır. Tablo 5.9 üzerinde bu testin istatistik değerleri gösterilmektedir.

Tablo 5.9: Nihai regresyon denkleminin varyans testinin gösterimi.

Göstergeler	Değerler	Olasılık Değerleri
F İstatistik Değeri	1,569135	0,2160
Obs*R ²	3,131901	0,2089
SS	4,165759	0,1246

Bu doğrultuda Tablo 5.9’da görüldüğü üzere Breusch-Pagan-Godfrey testine göre olasılık değerleri 0,05 değerinden büyük olduğu için varyansların sabit dağıldığından söz edilebilir.

5.3.1.6 Otokorelasyon Analizi

Otokorelasyon analizinde kullanılacak olan Wallis testi Eviews 10.0 programı üzerinde olmadığı için işlem matematiksel olarak hesaplanmıştır.

$$d_4 = \frac{\sum_{t=5}^n (c_t - c_{t-4})^2}{\sum_{t=1}^n (c_t)^2} \quad (5.1)$$

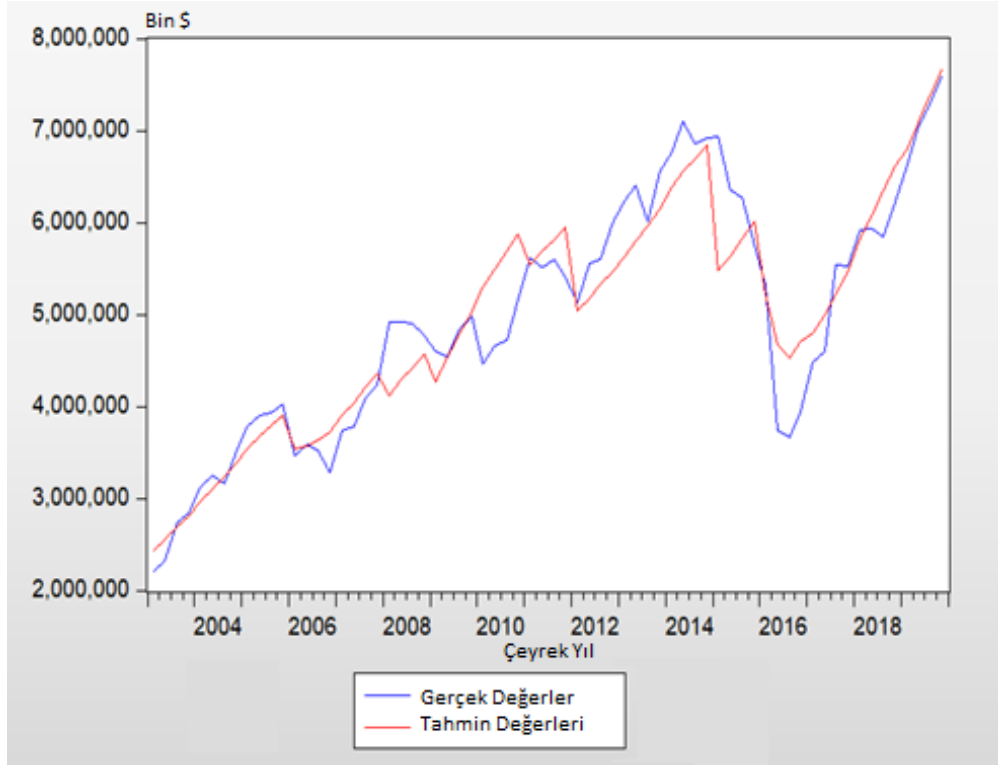
(5.1) denklemi Durbin-Watson testinin üçer aylık veriler için 4 gecikmeli versiyonu olarak uyarlanmıştır. Wallis (1972) tablosuna göre 2 değişkenli 68 verili alt ve üst kritik değerler 1,459 ve 1,577 olarak belirlenmiştir. Dolayısıyla otokorelasyonun olmadığını gösteren değer aralığı 1,577 ve (4-1,577) şeklindedir. Yapılan işlemler sonucunda d_4 değeri 1,799 olarak bulunmuştur. Bu değer aralıkta olduğu için modelde otokorelasyon probleminin olmadığı belirlenmiştir.

5.3.2 Çoklu Doğrusal Regresyon Uygulamasının Çıktıları

Modellenen ve optimum şartları sağlayan regresyon modelinde açıklayıcı değişken olarak Ters Çevrilmiş Küresel Terör Endeksi (TERORİT) ve Dış Hatlar Uçuş Trafiği (DIŞHAT) belirlenmiştir. Bu doğrultuda kurulan regresyon denklemi aşağıdaki gibidir.

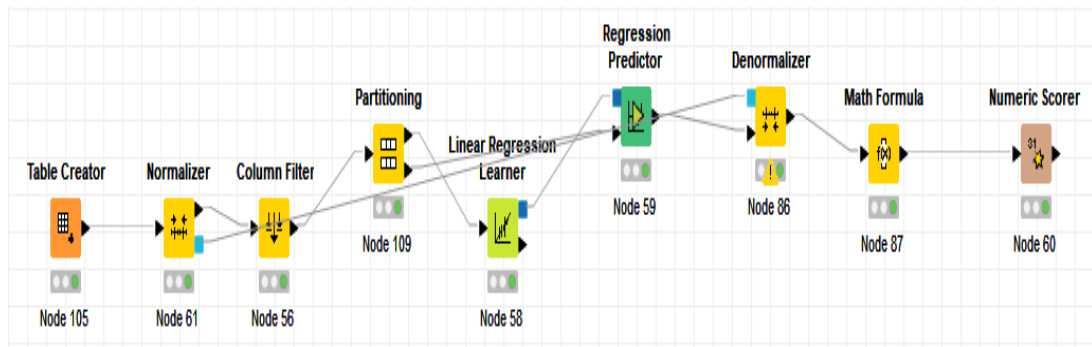
$$tgr = 17,27206dishat + 935446,3terorit - 3689561 \quad (5.2)$$

Burada TERORİT değişkeni terör olaylarının azlığını ifade ettiği için terör ile turizm geliri arasında negatif bir ilişki bulunmuştur. (5.2) denklemi ile mevsimsellikten arındırılmış gerçek turizm gelirleri ve tahmin değerleri karşılaştırıldığında Şekil 5.5 üzerindeki grafik elde edilmektedir.



Şekil 5.5: Turizm gelirinin çoklu doğrusal regresyon ile tahmin grafiği.

Sonraki başlıklarda kurulacak modellerle karşılaştırma yapabilmek adına veriler Knime 4.2.1 programı üzerinden değerlendirilmiştir. Diğer modeller gibi 0-1 değerleri arasında normalize edilmiştir. Şekil 5.6'daki görsel programın modül akışını göstermektedir. Bu konuyla ilgili detaylı bilgi sonraki bölümde anlatılacaktır.



Şekil 5.6: Çoklu doğrusal regresyon Knime modeli akış gösterimi.

5.4 Makine Öğrenmesi Modellerinin Uygulanması

Bu başlık altında daha önce çoklu doğrusal regresyon modelinde belirlenen değişkenlerle birer makine öğrenmesi yöntemi olan Esnek Geri Yayılım (RPROP) algoritmasına sahip çok katmanlı Yapay Sinir Ağları (YSA) modeli yani Çok Katmanlı Algılayıcı (MLP) yöntemi ve Gradyan Arttırmalı Regresyon Ağaçları (GBRT) algoritmasının parametre değerleri belirlenecektir. Modellerin uygulanmasında Knime 4.2.1 programından faydalanılmıştır. Veriler 0-1 değerleri arasında normalize edilmiştir. Normalizasyon işlemi min-max yöntemine göre yapılmıştır.

$$\hat{y}_i = \frac{y_i - \min(y)}{\max(y) - \min(y)} \quad (5.3)$$

(5.3) denkleminde min-max normalizasyon işlemi gösterilmektedir. Burada her bir i için y değeri y değerlerinin en küçüğünden çıkarılarak y değerlerinin en büyüğüyle en küçüğü arasındaki farka bölünür böylece y değerinin normalize edilmiş hali elde edilmiş olunur.

5.4.1 Yapay Sinir Ağları Uygulaması

Parametre seçimleri yapılırken kesin bir yöntem olmaması nedeniyle optimum parametreler belirlenip diğer parametrelerin performans olarak değişim göstermediği aralıklar analizler üzerinde parametrelerin tekrarlı olarak denenmesi sonucu tespit edilmiştir. Yapay Sinir Ağı (YSA) uygulanırken tercih edilen parametreler verilerin çeşitliliğine, veri boyutuna ve verilerin komplekslik derecesine göre değişmektedir. Uygulama için seçilen parametreler Tablo 5.10'da gösterildiği gibi kullanılmıştır.

Tablo 5.10: Yapay sinir ağı parametre gösterimi.

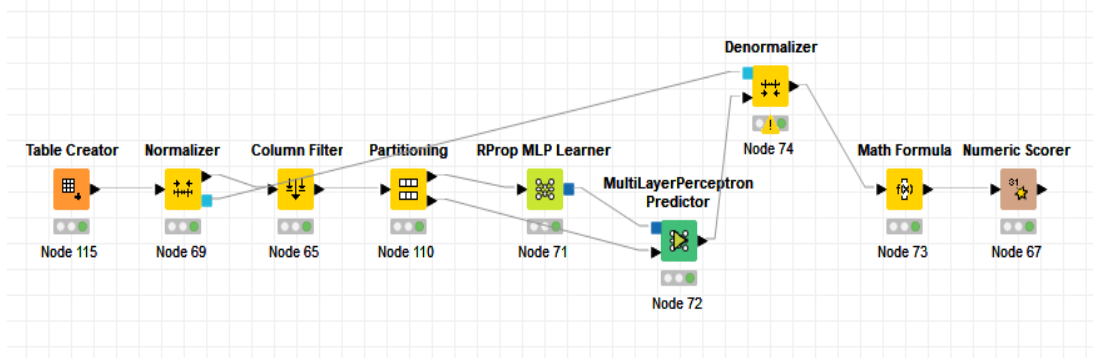
Parametre	Parametre Bilgileri
Giriş Katmanındaki Nöron Sayısı	2
Gizli Katman Sayısı	1, 2
Gizli Katmandaki Nöron Sayısı	3, 5, 7, 10
Çıktı Katmanındaki Nöron Sayısı	1
İterasyon	2500
Aktivasyon Fonksiyonu	Lojistik
Ağırlık Güncellemesi	İleri Beslemeli, Geri Yayılımlı
Kullanılan Algoritma	RPROP

Tek gizli katman sayısının zaman serilerinde yeterli olmasına karşın kompleks problemlerde katman sayısı arttırılabilmektedir. Fakat unutmamak gerekir ki katman sayısının arttırılması hem hesaplama performansı açısından hem de aşırı öğrenme sorunu açısından dezavantajlar yaratmaktadır. Genellikle gizli katmandaki nöron sayısının giriş katmanındaki nöron sayısından fazla olması daha iyi sonuçlar vermektedir (Kirk 2017). Bu nedenle gizli katman sayısı ve gizli katmandaki nöron sayısı yukarıdaki Tablo 5.10’da olduğu gibi belirlenmiştir. Nöron sayıları belirlenirken belirli bir yöntemin olmaması nedeniyle sonuçların etkilenme durumuna göre nöron sayıları test edilmiştir. Yapılan denemeler sonucu iterasyon sayısı bu uygulama için optimum 2500 olarak belirlenmiştir. Yöntemler bölümünde diğer özelliklerin gereklilikleri ve avantajları anlatılmıştır.

Aşağıdaki Şekil 5.7’de Knime 4.2.1 programında YSA’nın nasıl uygulandığı görülmektedir. Genel olarak kullanılan düğümler anlatılacak olunursa;

- “Table Creator” düğümü ile veriler girilmiş,
- “Normalizer” düğümü ile veriler 0-1 arasına çekilmiş,
- “Column Filter” düğümü ile YSA üzerinde kullanılacak değişkenler seçilmiş,
- “Partitioning” düğümü ile veri setleri eğitim ve test için ayrılmış,
- “Rprop MLP Learner” düğümü ile gerekli parametre ayarları yapılarak veriler eğitilmiş,
- “Multi Layer Perceptron Predictor” düğümü ile test için ayrılan veriler tahmin edilmiş,

- “Denormalizer” düğümü ile normalize edilen veriler eski haline getirilmiş,
- “Math Formula” düğümü ile tahmin edilen veriler eski veri aralıklarına getirilmiş,
- “Numeric Scorer” düğümü ile ise R^2 ve hata değerleri elde edilmiştir.



Şekil 5.7: Yapay sinir ağları Knime modeli akış gösterimi.

5.4.2 Gradyan Arttırmalı Regresyon Ağaçları Uygulaması

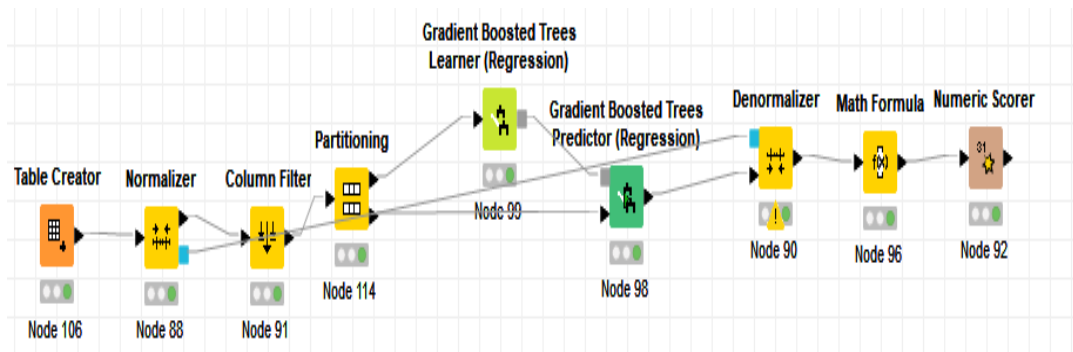
Bu aşamada parametreler belirlenirken farklı değerle yapılan analizler sonucunda ağaç derinliği hariç diğer parametrelerin uygulamada kullanılan veri kümesinde tekil ve optimum parametreler olduğu görülmüştür. Makine öğrenmesi metotlarının en büyük dezavantajlarından bir tanesi optimal bir parametre aralığına sahip olmamaları ve bu değerlerin optimum sonuçlara ulaşabilmesi için deneme yanılma yönteminin kullanılması gerekliliğidir. Tablo 5.11’de bu yöntemle belirlenen parametreler yer almaktadır.

Tablo 5.11: Gradyan arttırmalı regresyon ağaçları parametre gösterimi.

Parametre	Parametre Bilgileri
Maksimum Ağaç Derinliği	2, 3, 4, 5
Öğrenme Oranı	0,1
Maksimum Karar Ağacı Sayısı	100
Aykırı Değer Belirleme Oranı (1- Alfa)	0,05

Tablo 5.11’de gösterilen parametreler belirlenirken deneme yanılma yönteminde elde edilen optimum yöntemler kullanılmıştır. Bunun yanı sıra Yöntem başlığındaki kısıtlamalar göz önünde bulundurulmuştur.

Şekil 5.8’de Gradyan Arttırmalı Regresyon Ağaçlarının (GBRT) Knime 4.2.1 programı üzerinde nasıl uygulandığı ile ilgili genel bir akış gösterilmektedir. Bir önceki başlıkta Yapay Sinir Ağları (YSA) için akışın genel çerçevesi anlatılmıştı. Bu bölümde de işleyiş tıpkı bir önceki başlıkta anlatıldığı gibidir. “Gradient Boosted Trees Learner” düğümü içerisinde modele özgü parametrelerin ayarlamaları yapılmıştır.



Şekil 5.8: Gradyan arttırmalı regresyon ağacı Knime modeli akış gösterimi.

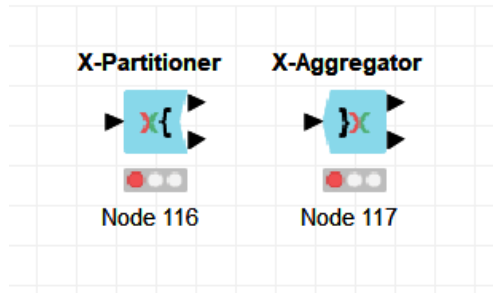
5.5 Modellerin Karşılaştırılması

Bu bölümde ilk olarak regresyon modeli ile belirlenen değişkenler makine öğrenmesi yöntemleriyle karşılaştırılacaktır. Daha sonra tüm değişkenler için makine öğrenmesi yöntemleri ayrı olarak tekrar karşılaştırılıp sonuçlar değerlendirilecektir.

5.5.1 Eğitim ve Test Veri Setlerinin Belirlenmesi

Veri setleri üzerinde çalışılırken verilerin aşırı öğrenme sorunuyla karşılaşmaması ve tatmin edici bir veri seti üzerinde test edilebilmesi için eğitim verileri ve test verileri optimum şekilde bölümlenmelidir. Kurulan modellerde eğitim verilerinin ve test verilerinin belirlenmesi için 5 ayrı veri bölümlenme parametresi seçilmiştir. Bu veri setleri oluşturulurken veriler rassal olarak, lineer sabit bir sırayla veya ilk n veri, k grubuna ait veriler gibi spesifik şekillerde seçilebilir. Standart

bölümleme işleminde veri verilen oranda eğitim ve test olarak ayrılmaktadır. K-katlamalı çapraz doğrulama seçiminde ise veriler k adet parçaya bölünür ve birinci parçadan k. parçaya kadar sırayla test verisi olarak belirlenirler. Tüm bu sonuçlar birleştirilip bir performans değeri ortaya konulur. Şekil 5.9 Knime 4.2.1 programı üzerinde katlamalı çapraz doğrulama yönteminin kullanılabilmesi için gerekli modülleri göstermektedir. 'X-Partitioner' akışta 'Partitioning' modülünün yerine konularak katlama sayısı ve yöntemi belirlenmektedir. Daha sonra 'X-Aggregator' modülü ile elde edilen veriler toplanmaktadır.



Şekil 5.9: K-katlamalı çapraz doğrulama Knime modülleri.

Büyük verilerde standart bölümleme yöntemi işe yarar sonuçlar verebilmektedir. Fakat küçük veri setlerinde yeterli doyuma ulaşmayacaktır. Bunun yanı sıra k-katlamalı çapraz doğrulama modeli ise iki şekilde de daha iyi çalışabilecekken büyük veri setlerinde kaynak kullanımını yavaşlatmaktadır. Aşağıdaki Tablo 5.12'de uygulama için kullanılacak bölümleme yöntemleri gösterilmektedir.

Tablo 5.12: Bölümleme parametreleri gösterimi.

Bölümleme Parametreleri	Lineer Örneklem (LN)	Rassal Örneklem (RND)
Standart Bölümleme (Eğitim Seti)	%70	%70
K-Katlamalı Çapraz Doğrulama	K=10, K=17	K=10

Genellikle standart bölümleme için verilen optimum oran %70 eğitim %30 test bölümlemesidir. K-katlamalı çapraz doğrulama için ise veriler genel olarak 10 parçaya ayrılarak her defasında test seti birer kaydırılarak verilerin 9/10'u eğitim, 1/10'u test için kullanılmaktadır. Verilerin yıllık olarak (4 çeyrek veri) tahmin edebilmesi adına 17 değeri de (4 çeyrek ve 17 yıllık toplam 68 dönemlik veri) gözlemler içerisine dahil edilmiştir. Bunun dışında rassal veri seçimi performansları arttırsa da eşit koşullarda

bir gözlem yapabilmek yani aynı eğitim ve test setleri için konuşabilmek adına lineer yöntemler sisteme dahil edilmiştir.

5.5.2 Performans Kriterlerinin Belirlenmesi ve Karşılaştırılması

Modellerin performanslarının ölçülmesinde Ortalama Mutlak Yüzde Hata (MAPE) kriteri kullanılmıştır. MAPE hesaplamasında her bir hata değerinin beklenen değere bölünmesiyle orantısal bir çıkarım yapılmaktadır. Her bir değer için bahsedilen bu oran Mutlak Yüzde Hata (APE) olarak adlandırılmaktadır. MAPE değeri ise bu değerlerin toplamının toplam değer sayısına bölümü ile bulunmaktadır. MAPE birim farklılıklarının oluşturabileceği sorunların önüne geçmekte ve yapısal kırılmaların yarattığı aşırı değişimlerden etkilenmemektir. Kısacası çok büyük bir değer ile çok küçük bir değer hata değerini karşılaştırabilir hale getirmektedir. Bu nedenle endeks, uçuş kapasitesi ve para birimi gibi farklı değerlerin karşılaştırılması için en uygun kriter MAPE olarak seçilmiştir. MAPE kriterinin formülü aşağıda verilmiştir.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \text{ Formülde;} \quad (5.4)$$

y_i = i. beklenen değeri,

\hat{y}_i = i. tahmin değerini,

n = toplam değer sayısını ifade etmektedir.

Aşağıda Tablo 5.13'te daha önce denklemleri kurulan çoklu doğrusal regresyon modeli test edilmiştir. Bu testte her bir bölümlenme parametresi ve bunların ortalaması için R^2 ve MAPE değerleri karşılaştırılmıştır.

Tablo 5.13: Çoklu doğrusal lineer regresyon test sonuçları.

LINEER	RND(k=10)		LN(k=10)		LN(k=17)		%70 RND		%70 LN		ORTALAMA	
	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE
DEĞER	0,876	%7,1	0,856	%7,8	0,838	%8,0	0,825	%6,6	0,862	%6,1	0,851	%7,1

Yapılan testler sonucunda en verimli parametrenin lineer olarak dağılan ve verilerin %70'i eğitim %30'u ise test için kullanılan senaryo sonucunda elde edildiği

görülmüştür. Bu senaryoda bulunan R^2 değerinin (0,86), matematiksel olarak hesaplanan denklemin R^2 değerinden (0,88) sadece %2'lik bir farkla tahmini ortaya koyduğu görülmektedir. Bu doğrultuda uygulamada kurulan makine öğrenmesi metotlarının da optimuma yakın açıklayıcılık değerleri içerdiği dikkat çekmektedir.

Tablo 5.14'te ise Gradyan Arttırmalı Regresyon Ağacı (GBRT) modeli için elde edilen sonuçlar gösterilmektedir. GBRT modelinde az sayıda değişkenin (2) olması ve 4 ağaç derinliğinden sonra sonuçların giderek düşük sonuçlar vermesi nedeniyle maksimum 5 ağaç derinliği kullanılmıştır.

Tablo 5.14: Gradyan arttırmalı regresyon ağaçları test sonuçları.

GBRT Ağaç Derinliği	RND(k=10)		LN(k=10)		LN(k=17)		%70 RND		%70 LN		ORTALAMA	
	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE
2	0,896	%6,0	0,796	%10,8	0,796	%11,0	0,816	%7,7	0,892	%5,7	0,8392	%8,2
3	0,906	%6,1	0,759	%11,1	0,777	%11,2	0,841	%6,8	0,891	%5,5	0,8348	%8,1
4**	0,893	%6,5	0,731	%11,5	0,774	%11,1	0,837	%6,7	0,899	%5,3	0,8268	%8,2
5	0,856	%6,9	0,704	%11,8	0,76	%11,5	0,803	%7,5	0,896	%5,5	0,8038	%8,6

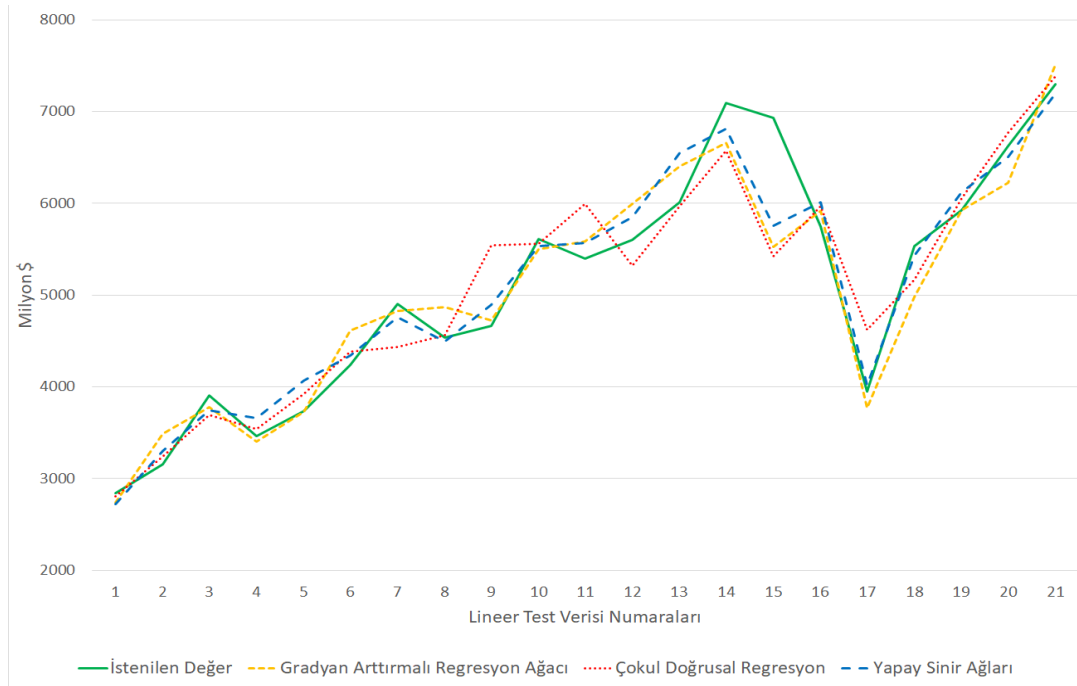
Yukarıdaki Tablo 5.14'te görüldüğü gibi GBRT modeli çoklu doğrusal regresyon modeline göre daha iyi sonuçlar vermektedir. Aynı parametreler altında verilerin %70'inin eğitim için kullanıldığı durumda en iyi modelin 4 ağaç derinliğiyle kurulan model olduğunu ve MAPE değerinin %5,3 ve R^2 değerinin ise 0,899 olduğu görülmektedir.

Tablo 5.15'te ise Yapay Sınır Ağları (YSA) modellerinden biri olan Çok Katmanlı Algılayıcı (MLP) yöntemi kullanılmıştır. Tek gizli katman sayısının yeterli olmasına karşın iki gizli katmanın doğurabileceği ekstrem sonuçları kaçırmamak adına test parametreleri iki katman sayısı olarak da denenmiştir.

Tablo 5.15: Yapay sinir ağıları test sonuçları.

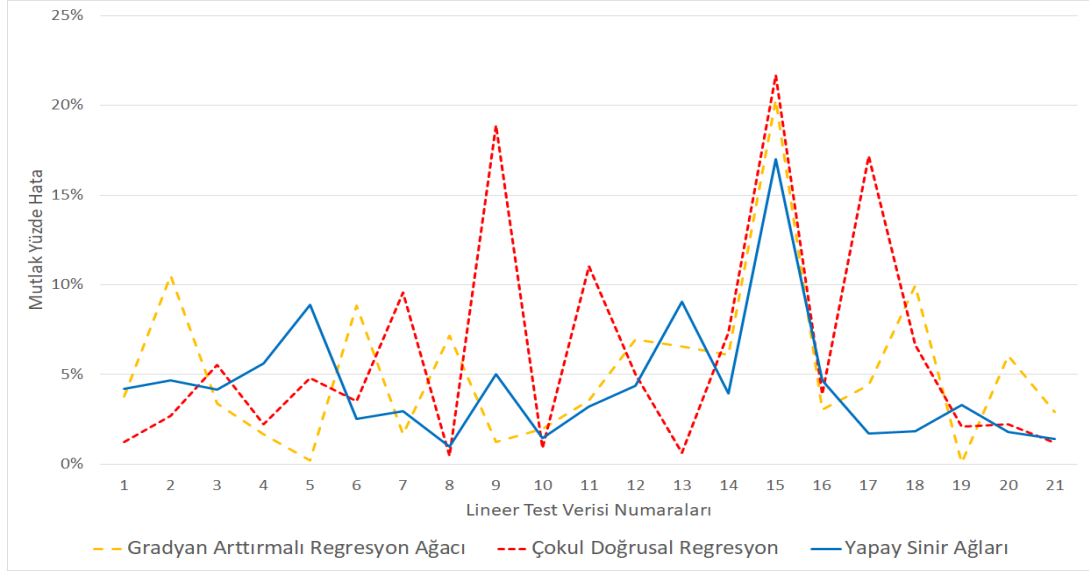
MLP	RND(k=10)		LN(k=10)		LN(k=17)		%70 RND		%70 LN		ORTALAMA	
	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE	R ²	MAPE
YSA(2,3,1)	0,903	%5,8	0,86	%7,9	0,893	%6,8	0,859	%8,8	0,919	%4,9	0,886	%6,8
YSA(2,5,1)***	0,942	%5,1	0,891	%7,3	0,875	%7,6	0,929	%5,0	0,933	%4,4	0,914	%5,8
YSA(2,7,1)	0,92	%5,6	0,84	%8,6	0,744	%10,1	0,918	%5,4	0,922	%4,7	0,868	%6,8
YSA(2,10,1)	0,933	%5,1	0,781	%9,4	0,809	%9,5	0,918	%6,4	0,911	%5,3	0,870	%7,1
YSA(2,3,3,1)	0,928	%5,1	0,884	%8,5	0,745	%9,6	0,928	%5,0	0,923	%5,1	0,881	%6,6
YSA(2,5,5,1)	0,914	%5,7	0,829	%10,0	0,795	%9,8	0,859	%8,6	0,913	%5,2	0,862	%7,8
YSA(2,7,7,1)	0,927	%5,1	0,68	%12,4	0,467	%13,4	0,889	%6,5	0,915	%5,6	0,775	%8,6
YSA(2,10,10,1)	0,889	%5,9	0,75	%10,7	0,667	%10,9	0,913	%6,7	0,936	%4,4	0,831	%7,7

Yukarıdaki Tablo 5.15'e göre ise en iyi parametrelerin verilerin lineer dağıtıldığı %70 eğitim oranıyla ve tek gizli katmanlı, gizli katmandaki nöron sayısı 5 olan senaryo olduğu görülmektedir. Bu sonuçlara göre karşılaştırmalı modelde en iyi sonuç veren model %4,4 MAPE ve 0,933 R² değeri ile [2:5:1] Esnek Geri Yayılım (RPROP) algoritmalı MLP yöntemiyle elde edilmiştir. Daha önce de bahsedildiği gibi çok kompleks, doğrusal olmayan modeller dışında genel olarak tek gizli katman kullanılması daha yüksek performanslı sonuçlar elde edilmesini sağlamaktadır. Şekil 5.10'da eğitim verileri sonucunda kurulan modeller test verilerinde tahmin edilerek istenilen yani gerçek verilerle karşılaştırılmıştır.



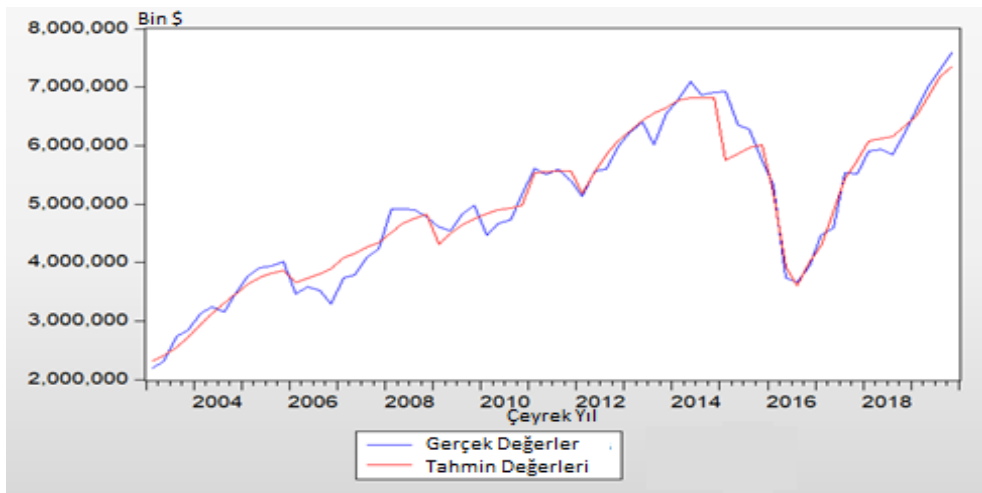
Şekil 5.10: Modellerin tahmin değerlerinin karşılaştırılması grafiği.

Şekil 5.10'da yapılan karşılaştırmada makine öğrenmesi modellerinin regresyon modeline göre istenilen değerler eğrisini daha iyi takip ettiği görülmektedir. Fakat her bir test verisi için kurulan modellerin Mutlak Yüzde Hata (APE) değerleri görsel olarak daha iyi bir karşılaştırma imkanı sağlayacağından Şekil 5.11'de modellerin APE değerleri karşılaştırılmıştır.



Şekil 5.11: Modellerin mutlak yüzde hatalarının karşılaştırılması grafiği.

Şekil 5.11'de de görüldüğü gibi kurulan Yapay Sinir Ağları (YSA) modeli diğer modellere göre daha az hatalı tahminler yapmıştır. En iyi model olarak saptanan RPROP algoritmalı YSA [2:5:1] modeli için tüm serilerin tahminlemesi ise Şekil 5.12 üzerinde gösterilmiştir.



Şekil 5.12: Turizm gelirinin en iyi model ile tahmin grafiği.

Bilindiği üzere makine öğrenmesi algoritmaları doğrusallık, normallik, varyans homojenliği vb. şartlar gerektirmeden çalışabilen yöntemlerdir. Bu doğrultuda Tablo 5.16’da çoklu doğrusal regresyondaki gereklilikler yüzünden kullanılmayan değişkenlerin tamamı kullanılarak Esnek Geri Yayılım (RPROP) algoritmali Çok Katmanlı Algılayıcı (MLP) ve Gradyan Arttırmalı Regresyon Ağaçları (GBRT) yöntemleri daha önceki tüm modellerde en başarılı sonucu veren bölümlene yöntemi %70 lineer bölümlene metoduyla karşılaştırılmıştır.

Tablo 5.16: YSA ve GBRT modelleri test sonuçları.

MODEL	R ²	MAPE
GBRT(2 Derinlik)	0,872	%7,1
GBRT(3 Derinlik)	0,863	%7,1
GBRT(4 Derinlik)	0,884	%7,0
GBRT(5 Derinlik)	0,931	%5,8
GBRT(6 Derinlik)	0,944	%5,0
GBRT(7 Derinlik)	0,949	%4,3
GBRT(8 Derinlik)	0,935	%4,8
GBRT(9 Derinlik)	0,934	%4,8
YSA(2,3,1)	0,89	%5,7
YSA(2,5,1)	0,944	%4,5
YSA(2,7,1)	0,932	%5,8
YSA(2,10,1)	0,918	%4,8
YSA(2,3,3,1)	0,947	%5,0
YSA(2,5,5,1)	0,849	%8,1
YSA(2,7,7,1)	0,856	%7,1
YSA(2,10,10,1)	0,86	%6,9

Yukarıdaki Tablo 5.16’da görüldüğü üzere en iyi sonuç veren makine öğrenmesi yöntemi 7 ağaç derinliğine sahip GBRT modelidir. Modelde MAPE değeri %4,3 ve R² değeri 0,949 olarak bulunmuştur. Çoklu doğrusal regresyon modelinin de dahil olduğu bir önceki modeller ile karşılaştırıldığında 2 parametre kullanılan modellemede RPROP algoritmali MLP daha iyi sonuçlar verirken tüm parametrelerin dahil edildiği senaryoda GBRT yöntemi daha iyi sonuçlar vermiştir. Ayrıca ana karşılaştırma kriterimiz olan MAPE değerlerine bakıldığında değişken sayısının arttırılmasının sonucun performansını dikkate değer bir şekilde etkilemediği görülmektedir.

Buradan çıkarılabilecek sonuçlar ise belirli gereksinimlere ihtiyaç duyan ve klasik matematiksel bir denklemle çalışan En Küçük Kareler (EKK) yönteminin

kaynakların daha efektif kullanılabilmesi ve spesifik belirlemeler yapılabilmesi adına makine öğrenmesi yöntemlerinde kullanılacak parametre seçimlerinde daha anlamlı ve optimum sonuçlar verebileceğidir. Kısacası sağlam temellere ve şartlara dayandırılan istatistiki yöntemlerin yüksek performanslı tahmin sonuçları veren makine öğrenmesi yöntemleriyle beraber kullanılması daha spesifik ve daha verimli sonuçlar doğurabilecektir. Bunun dışında Gradyan Arttırımlı Regresyon Ağaçları (GBRT) modellerinin yüksek parametrelerle Yapay Sinir Ağlarından (YSA) daha iyi sonuçlar çıkarabileceği düşünülmektedir. Bu nedenle mevcut literatürde sıklıkla kullanılan YSA modeline alternatif olarak özellikle yatay kesit olarak ilerleyen durumlarda iyi performans sergileyebileceği görülmüştür.

Literatürde eğer MAPE değeri %50' den daha büyükse bu tür modellerin yanlış kurulmuş olabileceği, %20 MAPE değerindeki modellerin kabul edilebilir modeller oldukları ve %10 MAPE değerine sahip modellerinse gayet iyi performanslı değerler oldukları bilinmektedir (Çuhadar 2020). Genel olarak bakıldığında tüm sonuçların oldukça tatmin edici performans değerlerine sahip oldukları açıkça ortadadır. Ayrıca iki değişkenli senaryoda en iyi model olan %4,4 MAPE ve 0,933 R² değeri ile [2:5:1] RPROP algoritmalı MLP yöntemi ve yirmi değişkenin de kullanıldığı MAPE değeri %4,3 ve R² değeri 0,949 olan GBRT yönteminin verdiği sonuçların oldukça tatmin edici oldukları açıkça görülmektedir.

6. TARTIŞMA VE ÖNERİLER

Turizm her açıdan ülkelerin gelişimine katkı sağlayan çok boyutlu getirileri olan bir sektördür. Ancak en büyük etkilerinden bir tanesi şüphesiz ki ülkelerin ekonomik gelişimine sağladığı katkı olmuştur. Ülkelerin tüm üretim ve hizmet ihracatlarına, istihdam kanallarına bakıldığında turizm önde gelen sektörler arasında üst sıralarda yer almaktadır. Turizm sektörünün her yıl katlanarak büyümeye devam etmesi ve ülkeler açısından çok büyük ekonomik etkilere sahip olması sektör içerisinde kıyasıya bir rekabeti de beraberinde getirmektedir. Ülkemiz de turizme elverişli ve yabancı ziyaretçi sayısı bakımından başarılı ülkeler içerisinde yer almaktadır. Fakat bu başarı diğer ülkelere kıyasla aynı oranda turizm gelirine yansıyamamıştır. Dolayısıyla Türkiye örneğinde de açıkça görüldüğü gibi turizmle ilgili performans ölçümü yapılırken turist sayılarına değil turizm gelirlerine bakılması, planlamaların, tahminlerin ve geliştirilen stratejilerin bu yöne odaklanması daha doğru olacaktır. Bu doğrultuda, çalışmada turizm gelirini etkileyen faktörleri bulmak ve bu faktörlerin turizm gelirini açıklama konusunda performanslarını tam anlamıyla değerlendirebilmek adına ekonomi, turizm kapasitesi, ülke güvenliği ve özgürlüğü, ülke gelişmişliği başlıklarında 20 adet değişken üzerinde 2003-2019 yılları arası üçer aylık verilerle analizler yapılmıştır. Bu veri analizleri ekonometrik modellerle desteklenmiş çoklu doğrusal regresyon, Yapay Sinir Ağları (YSA) ve Gradyan Arttırımlı Regresyon Ağaçları (GBRT) yöntemleri kullanılarak gerçekleştirilmiştir.

Çalışmanın ilk aşamasında çok fazla gereksinimle çalışan klasik regresyon modeli ile bir denklem kurulmaya çalışılmıştır. Kurulan bu denklemde birim kök analizi, eşbütünleşme analizi gibi ekonometrik yöntemlerden de faydalanılmıştır. Bu analizler sonucunda seçilen 20 adet değişken içerisinde turizm gelirlerini anlamlı bir şekilde ve yüksek oranlarda etkileyen 2 adet değişken bulunmuştur. Bu değişkenler yaklaşık 0,88 R² açıklayıcılık oranıyla dış hatlar uçuş trafiği ve küresel terör endeksidir. Bunun dışında eşbütünleşme analizleriyle ise turizm gideri, Gayri Safi Yurtiçi Hasıla (GSYİH), işsizlik oranı gibi ekonomik göstergelerin uzun dönemde anlamlı olarak turizm gelirlerini açıklamakta yeterli olmadıkları görülmüştür. Turizm gelirlerini etkileyen faktörlerin ise çoğunlukla turizm kapasitesi ve ülke güvenliğiyle ilgili değişkenler olduğu saptanmıştır.

Uygulamanın ikinci aşamasında bulunan bu çoklu regresyon denklemi makine öğrenmesi yöntemleriyle karşılaştırılmıştır. İlk olarak teoride bulunan ve yüksek R^2 değerine sahip olan iki değişkenli doğrusal regresyon denkleminin, pratikte doğruluğunu koruyup koruyamayacağına bakmak için Knime 4.2.1 programı üzerinde aynı veriler eğitim ve test serilerine ayrılıp incelenmiştir. Test verilerinin denenmesi sonucunda R^2 değerinin yaklaşık 0,86 ve MAPE değerinin yaklaşık %6 olduğu görülmüştür. Bu da göstermektedir ki kurulan çoklu doğrusal regresyon modeli ile bir önceki aşamada bulunan istatistiki sonuçlara çok yakın ve performans kriterleri açısından gayet başarılı bir sonuç elde edilmektedir. Daha sonra bağımsız değişkenleri terör endeksi ve dış hatlar uçuş trafiği olarak belirlenen bu model performans açısından makine öğrenmesi modelleriyle karşılaştırılmıştır. Sonuç olarak 2 değişkenle turizm gelirini en iyi açıklayan model, verilerin %70 eğitim için ayrıldığı parametre sonuçlarıyla tek gizli katmanın olduğu ve gizli katman nöron sayısının 5 olduğu çok katmanlı Esnek Geri Yayılım (RPROP) algoritmasına sahip YSA olmuştur. Bu modelle elde edilen MAPE değeri %4,4 ve R^2 değeri 0,933 olarak bulunmuştur. Eğitim ve test seti değiştirildiğinde ortalamalara bakıldığında ise yine aynı modelin %91,4 R^2 ve %5,8 MAPE değeriyle en başarılı model olduğu görülmüştür. Daha sonra regresyon modellerindeki gereksinimlere sahip olmayan makine öğrenmesi modelleri 20 değişkenin hepsi dahil edilerek karşılaştırılmış ve en iyi modelin ağaç derinliği 7 olan GBRT modeli olduğu görülmüştür. Bu modelle ise elde edilen en iyi sonucun 0,949 R^2 değeri ve %4,3 MAPE hata değeri olduğu saptanmıştır ve değerlerin 2 değişkenle kurulan YSA modelinden çok da uzak olmadığı fark edilmiştir.

Sonuç olarak yöntemler için genel bir değerlendirme yapmak gerekirse, ekonometrik yöntemler, regresyon ve makine öğrenmesi yöntemleri birleştirildiğinde gayet iyi çözümlerin ortaya çıkabileceği gözlemlenmiştir. Makine öğrenmesi modellerinde kullanılacak değişkenleri belirlerken regresyon ve ekonometrik modeller gibi klasik istatistiki yöntemlerden detaylı bir şekilde yararlanmak, hem daha etkin kaynak kullanımı açısından, hem de klasik istatistiksel modellerin getirdiği sağlam dayanakların makine öğrenmesi yöntemlerine yansıtılması açısından çok faydalı olabilecektir. Bunun yanı sıra makine öğrenmesi modelleri arasında yapılan karşılaştırma ise bize yatay kesiti fazla olan veri setlerinde Gradyan Arttırımlı Regresyon Ağaçları (GBRT) modellerinin, az olan veri setlerinde ise Yapay Sinir Ağları (YSA) modellerinin daha performanslı sonuçlar verebileceğini göstermiştir.

Dolayısıyla turizm gelirleri açısından ileride yapılacak çalışmalarda panel verilere dayanacak çoklu ülke karşılaştırmalarında GBRT modellerinin, açıklayıcı değişkeni bulunmayan zaman serisi modellerinde ise YSA modellerinin kullanılmasının faydalı olacağı düşünülmektedir.

Turizm gelirini etkileyen faktörler açısından genel bir değerlendirme yapmak gerekirse, küresel ölçekte daha güvenli ve kolay ulaşım imkanlarına sahip bir ülke imajı çizmenin turizm gelirini büyük ölçekli etkileyebileceği fark edilmiştir. Zannedilenin aksine ekonomik faktörlerin ise turizm gelirlerine önemli bir etkide bulunmadığı görülmüştür. Kısacası ekonomi turizmi değil turizm ekonomiyi etkilemektedir. Turizmi etkilediği düşünülen kısa dönemli ekonomik etkilerin ise ülkemize gelen orta gelir düzeyine sahip turist sayısının çokluğundan kaynaklandığı düşünülmektedir. Bu konuda ise turist profillerinin değişmesinin turizm gelirin nasıl etkileri olabileceğiyle ilgili ve yüksek gelirli turist akışının sağlanması için hangi çalışmaların etkili olabileceğiyle ilgili araştırmaların fayda sağlayabileceği düşünülmektedir.

7. KAYNAKLAR

Abar, H., “XGBoost ve Mars yöntemleriyle altın fiyatlarının kestirimi”, *Ekev Akademi Dergisi*, 24(83), 427-445, (2020).

Abdualgalil, B. S. and Abraham, S., “Tourist prediction using machine learning algorithms”, *Researchgate Publication*, <https://www.researchgate.net/publication/339927009>, 111-121, (2020).

Ağyar, Z., “Yapay sinir ağlarının kullanım alanları ve bir uygulama”, *Mühendis ve Makine*, 56(662), 22-23, (2015).

Aktaş, C., “Türkiye'nin turizm gelirini etkileyen değişkenler için en uygun regresyon denkleminin belirlenmesi”, *Doğuş Üniversitesi Dergisi*, 6 (2), 163-174, (2005).

Arabacı, H., “Turizm sektörünün ekonomik büyümeye etkisi üzerine teorik bir inceleme”, *Balkan ve Yakın Doğu Sosyal Bilimler Dergisi*, 04 (03), 104-109, (2018).

Arı, E. and Yıldız, A., “Examination of affecting variables for youth unemployment with cointegration analysis”, *Alphanumeric Journal*, 5 (2), 309-316, (2017).

Ataseven, B., “Yapay sinir ağları ile öngörü modellemesi”, *Öneri Dergisi*, 10 (39), 101-115, (2013).

Aydin, A., Darici, B. ve Taşçı, H., “Uluslararası turizm talebini etkileyen ekonomik faktörler: Türkiye üzerine bir uygulama”, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (45), 143-177, (2015).

Bahar, O., “Turizm sektörü- ekonomi ilişkisi”, (ed: M. Kozak), *Turizm Ekonomisi*, Eskişehir: Anadolu Üniversitesi Yayınları, 2-26, (2013).

Bauman, Z., *Küreselleşme*, 2. Baskı, İstanbul: Ayrıntı Yayınları, (2006).

Bayram, O., “Türkiye için turizm talep denkleminin tahmini: ekonomik, sosyal ve siyasal etkenler”, Yüksek Lisans Tezi, *İzmir Katip Çelebi Üniversitesi Sosyal Bilimler Enstitüsü, Finansal Ekonomi Anabilim Dalı*, İzmir, (2018).

Bento, J. P., “Tourism and economic growth in Portugal: An empirical investigation of causal links”, *Tourism and Management Studies*, 12(1), 164–171, (2016).

Bilgiçli, İ. ve Altınkaynak, F., “turizm endüstrisinin Türkiye ekonomisi içindeki yeri ve önemi; ekonomi paradigmasıyla yaklaşım”, *Uluslararası Yönetim İktisat ve İşletme Dergisi*, ICAFR 16 Özel Sayısı, 560-580, (2016).

Büyükuysal, M. Ç. ve Öz, İ. İ., “Çoklu doğrusal bağıntı varlığında en küçük karelere alternatif yaklaşım: ridge regresyon” *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, 6 (2), 110-114, (2016).

Cardenas-Garcia, P. J. and Pulido-Fernandez J. I., “Tourism as an economic development tool. Key factors”, *Current Issues in Tourism*, 22(17), 2082-2108, (2019).

Cerovic, S., Vukadinovic, P. and Knezevic, M., “The influence of globalization on tourism and impact of tourism on other activities with an emphasis on greenfield investments in tourism”, *Singidunum International Tourism Conference*, Serbia, 47-52. (2015).

Cevahir, E., *SPSS ile nicel veri analizi rehberi*, İstanbul:Kıbele, (2020).

Civelek, G., “Küreselleşmenin turizm ekonomisi üzerine etkisi: Türkiye analizi”, Yüksek Lisans Tezi, *T.C. Necmettin Erbakan Üniversitesi Sosyal Bilimler Enstitüsü*, İktisat Anabilim Dalı, Konya ,(2020).

Çayıroğlu, İ., “Yapay sinir ağları[online]”, (27.12.2020), http://www.ibrahimcayiroglu.com/Dokumanlar/GoruntuIsleme/Goruntu_Isleme_Ders_Notlari-11.Hafta.pdf, (2015).

Çeken, H., “Küreselleşme eğilimlerinin uluslararası turizm hareketlerine etkisi ve Türkiye”, *Muğla Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (12), 2-11, (2004).

Çeken, H., Dalgın, T. ve Karadağ, L., “Küreselleşme ve uluslararası turizm arasındaki ilişki”, *Muğla Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (22), 21-36, (2009).

Çelik Uğuz, S., “Turizm ve suç arasındaki nedensellik ilişkisi”, *Uluslararası Ekonomik Araştırmalar Dergisi*, 3(3), 383-397, (2017).

Çımat, A. ve Bahar, O., “Turizm sektörünün Türkiye ekonomisi içindeki yeri ve önemi üzerine bir değerlendirme”, *Akdeniz İİBF Dergisi*, (6), 1-18, (2003).

Çuhadar, M., Güngör, İ. ve Göksu, A., “Turizm talebinin yapay sinir ağları ile tahmini ve zaman serisi yöntemleri ile karşılaştırmalı analizi: Antalya iline yönelik bir uygulama”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14 (1), 99-114, (2009).

Çuhadar, M., “Türkiye’ye yönelik dış turizm talebinin MLP, RBF ve TDNN yapay sinir ağı mimarileri ile modellenmesi ve tahmini: karşılaştırmalı bir analiz”, *Journal of Yasar University*, 8 (31), 5274-5295, (2013).

Çuhadar, M., “Türkiye’nin dış aktif turizm gelirlerinin alternatif yaklaşımlarla modellenmesi ve tahmini”, *Ankara Hacı Bayram Veli Üniversitesi Turizm Fakültesi Dergisi*, 23(1), 115-141, (2020).

Çullu Kaygısız, N., “Son 10 yılda dünyada ve Türkiye’de turizm (2008-2017)”, *Iğdır Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (19), 607-626, (2019).

Dabiri, S. and Abbas, M., “Evaluation of the gradient boosting of regression trees method on estimating car-following behavior”, *Transportation Research Record*, 2672(45), 136-146, (2018).

Demir, R., “Yapay sinir ağlarıyla şirket birleşmelerinin kestirimi”, Doktora Tezi, İstanbul Teknik Üniversitesi *Fen Bilimleri Enstitüsü*, İstanbul, (2007).

Deniz, S. S., “Veri madenciliği araçları kullanılarak Türkiye’nin turizm gelirlerinin aylara göre yapay sinir ağları ile tahminlenmesi”, *Yüzyüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 241-255, (2019).

Dickey, D. A. and Fuller, W. A., “Distribution of the estimators for autoregressive time series with a unit root”, *Journal of The American Statistical Association*, 74(366), 427-431, (1979).

Dickey, D. A. and Fuller, W. A., “Likelihood ratio statistics for autoregressive time series with a unit root”, *Econometrica*, 49(4), 1057-1072, (1981).

Doğan, S., “Günlük buharlaşma miktarının yapay sinir ağları metotları ve klasik yöntemlerle tahmini”, Yüksek Lisans Tezi, İskenderun Teknik Üniversitesi, *Mühendislik ve Fen Bilimleri Enstitüsü*, Hatay, (2020).

Dritsakis, N., “Cointegration analysis of German and British tourism demand for Greece”, *Tourism Management*, 25(1), 111-119, (2004).

Engle, R. F. And Granger, C. W. J., "Co-integration and error correction: representation, estimation, and testing", *Econometrica*, 55(2), 251-276, (1987).

Eşidir, K. A., “Yapay sinir ağları ile metodu ile Türkiye tavukçuluk sektörü ihracat tahmini uygulaması”, Doktora Tezi, Fırat Üniversitesi, *Sosyal Bilimler Enstitüsü*, Elazığ, (2020).

Fareed, Z., Meo, M. S., Zulfiqar, B., Shahzad, F. and Wang, N., “Nexus of tourism, terrorism, and economic growth in Thailand: new evidence from

asymmetric ARDL cointegration approach”, *Asia Pacific Journal of Tourism Research*, 23(12), 1129-1141, (2018).

Friedman, J. H., “Greedy function approximation: A stochastic boosting machine”, Technical Report, *Department of Statistics Stanford University*, (1999).

Gökler, S. H., “Kan bankalarında talep tahmini ve stokastik stok yönetimi”, Doktora Tezi, *Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Anabilim Dalı*, Sakarya, (2020).

Granger, C. W. J., “Spurious regressions in econometrics”, (ed: B. H. Baltagi), *A Companion to Theoretical Econometrics*, Oxford: Blackwell, 557-561, (2001).

Granger, C. W. J. and Newbold P., “Spurious regressions in econometrics”, *Journal of Econometrics*, 111-120, (1974).

Gregory, A. W. and Hansen, B. E., “Residual-based tests for cointegration in models with regime shifts”, *Journal of Econometrics*, 70, 99-126, (1996).

Grover, P., “Gradient Boosting from scratch[online], (05.01.2021), <https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d>, (2017).

Hastie, T., Tibshirani, R. and Friedman, J., “Boosting and additive trees”, *The Elements of Statistical Learning*, New York:Springer, 337-387, (2009).

Höpken, W., Eberle, T., Fuchs, M., and Lexhagen, M., “Improving tourist arrival prediction: a big data and artificial neural network approach”, *Journal of Travel Research*, 1-20, doi: 10.1177/0047287520921244, (2020).

Işık, N., Acar, N. ve Işık, H. B., “Enflasyon ve döviz kuru ilişkisi: bir eşbütünleşme analizi”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 9(2), 325-340, (2004).

Kang, C. and Gu, J., “Analysis of tourist flow forecasting model based on multiple additive regression tree”, *IOP Publishing*, doi:10.1088/1757-899X/490/4/042001, (2019).

Karakullukçu, E., “Yanık görüntülerinin çok değişkenli istatistiksel yöntemler ve derin öğrenme yaklaşımları ile analizi”, Doktora Tezi, *Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstatistik ve Bilgisayar Bilimleri Anabilim Dalı*, Trabzon, (2020).

Kavzođlu, T., řahin, E. K. ve řölkesen, İ., “Heyelan duyarlılıđının incelenmesinde regresyon ađaçlarının kullanımı: Trabzon örneđi”, *Harita Dergisi*, 147(3), 21-33, (2012).

Keleş, M. B., Keleş A. ve Keleş A., “Makine öđrenmesi yöntemleri ile uçuş fiyatlarının tahmini”, *Eurosia Journal of Mathematics*, 7(11), 72-78, (2020).

Keskin, H. İ., “Using the seemingly unrelated regression model in the estimation of tourism demand of Turkey”, *Journal of Tourism Theory and Research*, 5(2), 182-190, (2019).

Khan, S. A. R., Qianli, D., SongBo, W., Zaman, K. and Zhang, Y., “Travel and tourism competitiveness index: The impact of air transportation, railways transportation, travel and transport services on international inbound and outbound tourism”, *Journal of Air Transport Management*, 58, 125-134, (2017).

Kirk, M., *Thoughtful Machine Learning with Python: A Test-Driven Approach*, USA: O’reilly, (2017).

Koutras, A., Panagopoulos, A. ve Nikas, I. A., “Forecasting tourism demand using linear and nonlinear prediction models”, *Academica Turistica-Tourism and Innovation Journal*, 9(1), 85-98, (2016).

Kozak, M., “Turizm Sektörünün Ulusal Ekonomiler Üzerindeki Etkileri”, (ed: M. Kozak), *Turizm Ekonomisi*, Eskişehir: Anadolu Üniversitesi Yayınları, 172-201, (2013).

Mohan, A., Chen, Z. ve Weinberger, K., “Web-search ranking with initialized gradient boosted regression trees”, (eds: O. Chapelle, Y. Chang, T. Liu, *Yahoo learning to rank challenge*, St. Louis, JMLR, 77-89, (2011).

Natekin, A. ve Knoll, A., “Gradient boosting machines, a tutorial”, *Frontiers in neurobotics*, doi: 10.3389/fnbot.2013.00021, (2013).

OECD, “Megatrends shaping the future of tourism[online]”, (19.12.2020), <https://www.oecd-ilibrary.org/docserver/tour-2018-6-en.pdf?expires=1608408843&id=id&accname=guest&checksum=832F472CD8BE991D9261E6680C780045>, (2018).

Uçan, O. ve Saraç, M., “İthalat, döviz kuru ve ekonomik büyüme arasındaki ilişkinin zaman serisi analizi: Türkiye örneđi (1994-2018)”, Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 13(4), 696-708, (2020).

Özkan, K., “Sınıflandırma ve regresyon ağacı tekniği (SRAT) ile ekolojik verinin modellenmesi”, *SDÜ Orman Fakültesi Dergisi*, 13, 1-4, (2012).

Öztemel, E., *Yapay Sinir Ağları*, İstanbul, Ankara, İzmir, Adana: Papatya Yayıncılık, (2003).

Öztürk, K. ve Şahin, M. E., “Yapay sinir ağları ve yapay zekâ’ya genel bir bakış”, *Takvim-i Vekayî*, 6(2), 25-36. (2018).

Paksoy, S., Yenipazarlı, A. ve Bulut, E., “Turizmin toplam istihdam üzerindeki etkisi; var analizi”, *Sosyal ve Beşeri Bilimler Araştırmaları Dergisi*, 19 (41), 35-50, (2018).

Payne, J. E. ve Mervar, A., “A note on modelling tourism revenues in Croatia”, *Tourism Economics*, 8(1), 103-109, (2002).

Pekmezci, A. ve Bozkurt, K., “Döviz kuru ve ekonomik büyüme: türk turizm sektörü için bir analiz”, *SDÜ Sosyal Bilimler Enstitüsü Dergisi*, 2(24), 98-110, (2016).

Perron, P., “The great crash, the oil price shock and the unit root hypothesis” *Econometrica*, 57, 1361-1401, (1989).

Raza, S. A., and Jawaid, S. T., “Terrorism and tourism: A conjunction and ramification in Pakistan”, *Economic Modelling*, 33, 65-70, (2013).

Riedmiller, M. ve Braun, H., “A direct adaptive method for faster backpropagation learning: The RPROP algorithm”. *International Conference On Neural Networks*, IEEE, San Francisco, 586-591, (1993).

Rojas, R., *Neural Networks-A Systematic Introduction*, Berlin: Springer-Verlag, (1996).

Soylu, K., “Kredi kartı sahte işlem tespiti”, Yüksek Lisans Tezi, *Ankara Üniversitesi Fen Bilimleri Enstitüsü*, Bilgisayar Mühendisliği Anabilim Dalı, Ankara, (2018).

Schwert, G. W., "Tests for unit roots: a monte carlo investigation," *Journal of Business and Economic Statistics*, 7, 147-160, (1989).

Şen, A. ve Şit, M., “Turizm gelirlerinin Türkiye ekonomisindeki rolü”, *Dicle Üniversitesi İİBF Dergisi*, 5 (8), 30-45, (2015).

T.C. Kültür ve Turizm Bakanlığı, “Turizm İstatistikleri Genel Değerlendirme Raporu[online]”, (21.12.2020),

<https://yigm.ktb.gov.tr/Eklenti/69320,turizmistatistikleri2019-4pdf.pdf?0>, (2019).

Tabachnick, B.G. and Fidell, L.S., *Using Multivariate Statistics*, Boston:Pearson, (2013).

Tengilimođlu, E. ve Kuzucu, S. C., “Döviz kuru oynaklıkları turist başına düşen ortalama harcamayı etkiler mi? 2003-2018 yılları arasında bir nedensellik analizi”, *International Human and Civilization Congress, From Past to Future*, Ines, Alanya, 148-156, (2019).

Turanlı, M. ve Güneren, E., “Turizm sektöründe talep tahmin modellemesi”, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 2(3), 1-13, (2003).

TÜİK, “ İstatiksel Tablolardan ve MEDAS Veritabanından Derlenmiş bilgiler[online]”, (30.12.2020), <https://www.tuik.gov.tr/>, (2020).

Uğurlu, E., “Durađanlık ve birim kök sınamaları”, *Researchgate*, 1-17, doi: 10.13140/rg.2.1.3262.2561, (2009).

UNWTO (UN World Tourism Organization), “UNWTO world tourism barometer[online]”, (19.12.2020), <https://www.unwto.org/unwto-tourism-dashboardbarometer>, (2017).

UNWTO (UN World Tourism Organization), “A compilation of data on inbound tourism by country[online]”, (21.12.2020), <https://www.unwto.org/country-profile-inbound-tourism>, (2020).

Uysal, M. and El Roubi, M. S., “Artificial neural networks versus multiple regression in tourism demand analysis” *Journal of Travel Research*, 38(2), 111-118, (1999).

Uzgören, N. ve Uzgören, E., “Zaman serilerinde sahte regresyon sorunu ve reel kamu harcamalarına yönelik bir ekonometrik model uygulaması”, *Akademik Bakış*, (5), 1-14, (2005).

Ünlü, H. ve Bozdađ, H., “Yapısal kırılma durumunda imalat sanayi ihracatı ve ekonomik büyüme ilişkisi: Türkiye üzerine bir uygulama”, *International Journal of Social Sciences and Education Research*, 2(1), 93-102, (2016).

Wallis, K. F., “Testing for fourth order autocorrelation in quarterly regression equations”, *Econometrica: Journal of the Econometric Society*, 40(4), 617-636, (1972).

WTTC (World Travel and Tourism Council), “Economic Impact Reports[online]”, (15.12.2020), <https://wttc.org/Research/Economic-Impact>, (2019).

Yavuz, S., “Hataları ardışık bağımlı (otokorelasyonlu) olan regresyon modellerinin tahmin edilmesi”, *Ataturk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 23(3), 123-139, (2009).

Yıldırım Tıraşoğlu, B., “Yapısal kırılmalı birim kök testleri ile OECD ülkelerinde satın alma gücü paritesi geçerliliğinin testi”, *Ekonometri ve İstatistik Dergisi*, (20), 68-87, (2014).

Yin, Y. C., Lee, C. S. and Wong, Y. P., “Demand prediction of bicycle sharing systems”, [http://cs229.stanford.edu/proj2014/YuchunYin, ChiShuenLee, Yu-PoWong, DemandPredictionofBicycleSharingSystems. pdf](http://cs229.stanford.edu/proj2014/YuchunYin,ChiShuenLee,Yu-PoWong,DemandPredictionofBicycleSharingSystems.pdf), (2012).

Yonar, H., “Yapısal kırılmalı zaman serileri analizi ile durağanlığın incelenmesi ve bir uygulama”, Yüksek Lisans Tezi, *Fırat Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı*, Elazığ, (2012).

Zaman, T., “Çoklu doğrusal regresyon analizinde çakı tahmin yöntemi ve test problemlerine katkılar”, Doktora Tezi, *Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı*, Samsun, (2017).

Zivot, E. and Andrews, D. W. K., “Further evidence of the great crash, the oil price shock and the unit root hypothesis”, *Journal of Business and Economic Statistics*, 10, 251-270, (1992).

Zhou, C. Yu, H., Ding, Y., Guo, F. and Gong, X-J., “Multi-scale encoding of amino acid sequences for predicting protein interactions using gradient boosting decision tree”, *Plos One*, 12(8), 1-18, (2017).

8. ÖZGEÇMİŞ

Adı Soyadı : Özgür GÜLER
Doğum Yeri ve Tarihi : MERSİN 12/09/1994
Lisans Üniversite : PAMUKKALE ÜNİVERSİTESİ
Elektronik posta : ozgurgler@hotmail.com
İletişim Adresi : TOROSLAR/MERSİN