

DOĞUM ORANINI ETKİLEYEN FAKTÖRLER İÇİN PANEL VERİ  
ANALİZİ ve MAKİNE ÖĞRENMESİ UYGULAMASI: TÜRKİYE ÖRNEĞİ

SOSYAL BİLİMLER ENSTİTÜSÜ

TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ

MUSTAFA ALİ CAN BERBEROĞLU

İŞLETME

YÜKSEK LİSANS TEZİ

ARALIK 2023

Bu Yüksek Lisans Tezinin Yüksek Lisans derecesi için gereken tüm koşulları yerine getirdiğini onaylarım.

---

Prof. Dr. Serdar SAYAN  
Sosyal Bilimler Enstitüsü  
Müdürü

Bu çalışmayı okuduğumu ve çalışmanın kapsam ve içerik olarak Sosyal Bilimler Enstitüsü İşletme Programı'nda bir Yüksek Lisans Tezi olabilecek yeterlilikte olduğuna kanaat getirdiğimi onaylıyorum.

Yüksek Lisans Tezi Danışmanı

Doç. Dr. Melike METERELLİYOZ  
(TOBB ETÜ, İşletme)

Yüksek Lisans Tezi Jürisi Üyeleri

Doç. Dr. Ebru YÜKSEL HALİLOĞLU  
(Gazi Üniversitesi, Endüstri Mühendisliği)

Dr. Öğr. Üyesi Salih TEKİN  
(TOBB ETÜ, Endüstri Mühendisliği)

Yüksek Lisans Tezi içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

---

Mustafa Ali Can  
BERBEROĞLU

# ÖZ

## DOĞUM ORANINI ETKİLEYEN FAKTÖRLER İÇİN PANEL VERİ ANALİZİ ve MAKİNE ÖĞRENMESİ UYGULAMASI: TÜRKİYE ÖRNEĞİ

BERBEROĞLU, Mustafa Ali Can

Yüksek Lisans, İşletme

Tez Danışmanı: Doç. Dr. Melike METERELLİYOZ

Bu araştırma, 2009-2021 yılları için Türkiye’de doğum oranını etkileyen faktörleri panel veri analizi ve makine öğrenmesi (ML) ile incelemeyi amaçlamaktadır. Bu amaçla il ve yıl bazında doğum oranının; lisans ve üzeri eğitim oranı, kadın istihdam oranı, nüfustaki kadın oranı, gayri safi yurt içi hasıla (GSYH) seviyesi, evli nüfus oranı, covid19 ve göçmen kukla değişkenleri ile ilişkisi araştırılmıştır. Panel veri analizi için havuzlandırılmış en küçük kareler (OLS), sabit etkiler ve rastgele etkiler olmak üzere üç ana model kullanılmıştır. ML için ise farklı sınıflandırma modelleri çalıştırılmış ve doğum oranı artış-azalış sınıflandırmasını en iyi öğrenen model tespit edilmiştir. ML modelleri ile elde edilen sonuçlar panel veri analizi sonuçları ile kıyaslanmış ve model değişkenleri açısından da farklılıklar incelenmiştir. Sonuçlar; sabit etkiler modeli ile panel veri analizi yapmanın anlamlı sonuç verdiği gözlemlenmiştir. Türkiye’de doğum oranı; lisans ve üzeri eğitim seviyesi, kadın istihdamı, COVID-19 ve göçmen kukla değişkeni ile negatif ilişkiliyken kadın nüfus oranı, evli oranı ve GSYH seviyesi ile pozitif ilişkili çıkmıştır. ML’de ise sınıflandırma problemine dönüştürülen doğum oranındaki değişim için, hangi modelin daha iyi performans gösterdiğine ve değişkenlerin önem düzeylerinin seviyelerine bakılmıştır. Kullanılan modellerin ikisi de doğum oranını belirleyen faktörler için etkin bir şekilde çalışabildiğini göstermiştir.

**Anahtar Kelimeler:** Doğum Oranı, Eğitim, Kadın İstihdamı, Panel Veri Analizi, Makine Öğrenmesi

## ABSTRACT

### PANEL DATA ANALYSIS AND MACHINE LEARNING PRACTICE FOR THE FACTORS THAT AFFECT THE BIRTH RATE: THE TURKIYE SAMPLE

BERBEROĞLU, Mustafa Ali Can

Master of Business Administration

Supervisor: Assoc. Prof. Melike METERELLİYOZ

This research aims to examine the factors that affect the fertility rate in Turkiye for the years 2009-2021, utilizing panel data analyses and machine learning (ML). To achieve this, the relationship between the fertility rate at the provincial and yearly levels and various factors such as the rate of people with undergraduate or higher education, women's employment rate, women's population rate, gross domestic product (GDP) level, married population ratio, COVID-19, and immigrant dummy variables has been investigated. Three main models, namely Pooled Ordinary Least Squares (OLS), Fixed Effects Model, and Random Effects Model, have been employed for panel data analyses. For ML, different classification models have been implemented to solve this problem, and the model that best learns the change of pattern in birth rate has been classified. The results obtained from ML models have been compared with panel data analysis results, and differences in terms of model variables have been examined. The findings; it has been observed that conducting panel data analysis with fixed effects model yields meaningful results. In Turkiye, the birth rate is found to be negatively correlated with the level of undergraduate and higher education, female employment, COVID-19, and the immigrant dummy variable, while it is positively correlated with the female population rate, married ratio, and GDP level. In ML, focused on transforming the change in birth rate into a classification problem, examining which model performed better and the levels of importance for the variables. Both models used have demonstrated effective performance in identifying factors that affect the fertility rate.

**Keywords:** Fertility Rate, Education, Women's Employment, Panel Data Analysis, Machine Learning



Hayatını kaybeden bütün masum çocuklara...

## TEŞEKKÜR SAYFASI

Lisans ve Lisansüstü hayatımda üzerimde ciddi emeği bulunan, her zaman anlaşıldığımı hissettiren, her daim örnek almaya gayret ettiğim, kıymetli hocam, tez danışmanım Doç. Dr. Melike Meterellioz'a içtenlikle teşekkür eder, saygılarımı sunarım.

Çocukluğumdan bugüne en büyük destekçilerim olan, sevgi ve muhabbetlerini her daim hissettiğim annem Hatice Berberoğlu, babam Mesut Berberoğlu ve kardeşlerim Sevde Gül, Rukiye Zehra ve Rana Bahar'a minnetle teşekkür ederim.

Yüksek Lisans sürecim boyunca ve beraberliğimiz süresince hayatımın en stresli ve zorlu dönemlerinde desteğini eksik etmeyen, beni hep motive eden sevgili eşim Elif Beyza Berberoğlu'na sevgiyle teşekkür ederim.

Yıllardır keyifle arkadaşlık yaptığım ve hayatımın her döneminde dostluklarımı esirgemeyen başta Fahrettin Yeşilyurt ve Faik Vural olmak üzere bütün yakın arkadaşlarıma teşekkür ederim.

Yüksek lisans eğitimim süresince TÜBİTAK 2210-A Genel Yurt İçi Yüksek Lisans Burs Programı kapsamında aldığım destek için de TÜBİTAK'a teşekkür ederim.

# İÇİNDEKİLER

İNTİHAL SAYFASI.....	iii
ÖZ .....	iv
ABSTRACT .....	v
İTHAF SAYFASI .....	vi
TEŞEKKÜR SAYFASI .....	vii
İÇİNDEKİLER .....	viii
KISALTMALAR LİSTESİ.....	xi
TABLolar LİSTESİ.....	xii
ŞEKİLLER LİSTESİ .....	xiii
BÖLÜM I.....	1
GİRİŞ .....	1
BÖLÜM II.....	9
LİTERATÜR TARAMASI.....	9
2.1. Doğum Oranını Etkileyen Faktörlere Dair Çalışmalar.....	9
2.1.a. Doğum Oranları ile Eğitim Arasındaki İlişkiye Dair Çalışmalar .....	14
2.1.b. Doğum Oranları ile Kadın İstihdamı Arasındaki İlişkiye Dair Çalışmalar .....	18
2.1.c. Doğum Oranları ile GSYH Arasındaki İlişkiye Dair Çalışmalar .....	20
2.2. Panel Veri Analizine Dair Çalışmalar .....	21
2.3. Makine Öğrenmesine Dair Çalışmalar .....	25
BÖLÜM III .....	29
VERİ .....	29
BÖLÜM IV .....	35
METODOLOJİ .....	35
4.1. Panel Veri Analizi .....	36

4.1.a. Havuzlanmış OLS Modeli .....	38
4.1.b. Sabit Etkiler Modeli.....	39
4.1.c. Rastgele Etkiler Modeli .....	40
4.1.d. Hausman Testi .....	41
4.1.e. Birim Kök Testi .....	42
4.2. Makine Öğrenmesi (ML).....	42
4.2.a Lojistik Regresyon (LR) .....	46
4.2.b. K En Yakın Komşular (KNN) .....	46
4.2.c. Çok Katmanlı Algılayıcı (MLP) .....	47
4.2.d. Destek Vektörü (SV) .....	47
4.2.e. Karar Ağaçları (DT).....	48
4.2.f. Rastgele Orman (RF).....	48
4.2.g. Gradyan Artırma (GB).....	49
4.2.h. Aşırı Gradyan Artırma (XGB).....	49
BÖLÜM V .....	51
BULGULAR.....	51
5.1. Panel Veri Analizi .....	51
5.1.a. Havuzlanmış OLS .....	51
5.1.b. Rastgele Etkiler Modeli .....	52
5.1.c. Sabit Etkiler Modeli.....	54
5.1.d. Hausman Testi .....	57
5.1.e. Birim Kök Testleri .....	57
5.2. Makine Öğrenmesi Uygulaması .....	58
5.2.a. ML İçin Model Seçimi.....	58
5.2.b. XGB Modelinde Değişkenlerin Önem Düzeyi.....	63
BÖLÜM VI.....	67
SONUÇ .....	67



## KISALTMALAR LİSTESİ

ABD	: Amerika Birleşik Devletleri
ADF	: Artırılmış Dickey-Fuller Testi (Augmented Dickey-Fuller)
AUC	: Eğrinin Altındaki Alan (Area Under the Curve)
DT	: Karar Ağacı (Decision Tree)
FN	: Yanlış-Negatif (False-Negative)
FP	: Yanlış-Pozitif (False-Positive)
GB	: Gradient Boosting
GDP	: Gayri Safi Yurt İçi Hasıla (Gross Domestic Product)
GSMH	: Gayri Safi Milli Hasıla
GSYH	: Gayri Safi Yurt İçi Hasıla
H0	: Sıfır Hipotezi (Null Hypothesis)
H1	: Alternatif Hipotez (Alternative Hypothesis)
KPSS	: Kwiatkowski-Phillips-Schmidt-Shin Testi
LR	: Lojistik Regresyon (Logistic Regression)
ML	: Makine Öğrenmesi (Machine Learning)
MLP	: Çok Katmanlı Algı (Multilayer Perception)
OECD	: Ekonomik İşbirliği ve Kalkınma Örgütü (Organization for Economic Cooperation and Development)
OLS	: En Küçük Kareler (Ordinary Least Squares)
PCSE	: Panel Doğrulama Standart Hatası (Panel Corrected Standart Error)
RF	: Rastgele Orman (Random Forest)
ROC	: Alıcı İşletim Karakteristiği (Receiver Operating Characteristic)
SV	: Destek Vektörü (Support Vector)
TFR	: Toplam Doğurganlık Oranı (Total Fertility Rate)
TN	: Doğru-Negatif (True-Negative)
TP	: Doğru-Pozitif (True-Positive)
TÜİK	: Türkiye İstatistik Kurumu
XGB	: Aşırı Gradyan Artırma (Extreme Gradient Boosting)
YSA	: Yapay Sinir Ağı

## TABLÖLAR LİSTESİ

<b>Tablo 5.1.</b> Havuzlanmış OLS Modeli Uygulama Sonuçları.....	52
<b>Tablo 5.2.</b> Rastgele Etkiler Modeli Uygulama Sonuçları .....	53
<b>Tablo 5.3.</b> Sabit Etkiler Modeli Uygulama Sonuçları .....	54
<b>Tablo 5.4.</b> Hausman Testi Uygulama Sonuçları.....	57
<b>Tablo 5.5.</b> Birim Kök Testleri Uygulama Sonuçları .....	58



## ŞEKİLLER LİSTESİ

<b>Şekil 1.1.</b> Bazı Ülkeler ve Dünya Genelindeki Doğum Oranlarının 2000, 2008 ve 2020 Yıllarındaki Değişimi (Dünya Bankası 2023).....	2
<b>Şekil 1.2.</b> 2001 ile 2021 Yılları Arasında Türkiye'deki Doğum Sayısı ve Toplam Doğurganlık Hızı (TÜİK 2023) .....	3
<b>Şekil 3.1.</b> 2009-2021 Yılları Arasında Lisansüstü Dereceye Sahip Kişi Sayısının Cinsiyet ve Yıllara Göre Değişimi (TÜİK 2023).....	31
<b>Şekil 3.2.</b> İllere Göre 13 Yıllık Doğum Oranı Ortalamaları (TÜİK 2023).....	31
<b>Şekil 3.3.</b> Lisans ve Üzeri Bir Dereceye Sahip İnsan Sayısının Şehir Nüfusuna Oranının 13 Yıllık Ortalamaları (TÜİK 2023).....	32
<b>Şekil 3.4.</b> İllerin 13 Yıllık Periyotta Kişi Başına Düşen GSYH Ortalaması (TÜİK 2023) .....	32
<b>Şekil 3.5.</b> Türkiye'nin kişi başına düşen GSYH'sinin Yıllara Göre Değişimi (TÜİK 2023) .....	33
<b>Şekil. 4.1.</b> Makine Öğrenmesi Çeşitleri (Piyush 2020) .....	43
<b>Şekil 5.1.</b> Sabit Etkiler Modeli Sonuçlarıyla Elde Edilen İl Katsayıların Haritalandırılmış Gösterimi .....	56
<b>Şekil 5.2.</b> Modellere Ait Doğruluk Oranları .....	60
<b>Şekil 5.3.</b> LR, KNN, MLP, SV, DT, RF, GB ve XGB'nin ROC Eğrisi Altında Kalan Alan Grafikleri .....	61
<b>Şekil 5.4.</b> XGB Karar Matrisi .....	62
<b>Şekil 5.5.</b> XGB Modeli Değişken Önem Düzeyi Uygulama Sonuçları.....	64

# BÖLÜM I

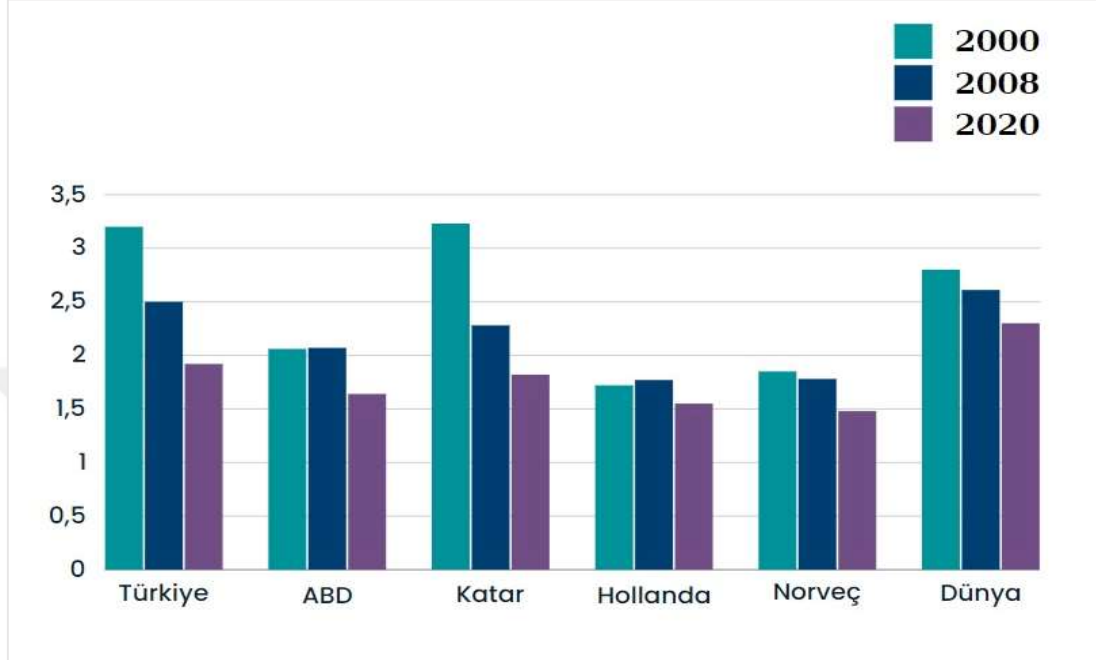
## GİRİŞ

Doğum oranı bir ülke ya da bölge için en önemli göstergelerden bir tanesidir. Bu gösterge mevcut ve gelecek politikaları belirlemede önemli bir rol oynamaktadır. Doğum oranı bir toplumdaki mevcut demografik durum ve gelecekteki nüfusun nasıl bir yaş dağılımına sahip olacağı konusunda fikir vermektedir. İş gücü ve sosyal politikaları belirleme açısından bu veriler oldukça önem arz etmektedir. Bir toplumda doğum oranının çok yüksek olması ilerleyen 20 yıl içerisinde işsizlik oranının artması; eğitim, altyapı ve sağlık hizmetlerinin yetersiz kalması ihtimallerini doğurabilmektedir. Bununla birlikte doğum oranının düşük olması da yine gelecek 20 yıl içerisinde işgücü bakımından yetersiz insan kaynağı, işgücü konusunda dışa bağımlılık, emekli nüfusu artışı nedeniyle kamu maliyetlerinin artması, sağlık sektöründe bakıma bağımlı nüfusun artması ve üretimin azalması risklerini de beraberinde getirebilmektedir.

Ülkelerin nüfuslarındaki doğum oranları gün geçtikçe düşmektedir. Şekil 1.1’de görüleceği üzere birçok ülkede ve küresel anlamda doğum oranlarının değişimine bakıldığında son 20 yılın genel bilançosu düşüş yönünde seyretmektedir. Eğer doğum oranlarındaki azalma eğilimi bu şekilde devam ederse, uzmanlara göre yüzyılın sonunda hemen hemen tüm ülkelerin nüfusunda düşüş yaşanabilir ve bu durum toplumlar üzerinde önemli etkiler meydana getirebilmektedir (Gallagher 2020). Küresel nüfusun, yüzyılın sonuna kadar zirveye ulaşması beklenmektedir. Toplumlardaki toplam doğum oranının 1,5'in altına düşmeye yönelik eğilim

göstermesi, küresel nüfusun azalmaya başladığında muhtemelen durdurulamaz bir şekilde devam edeceğini ifade etmektedir (Vollset vd. 2020, 1285).

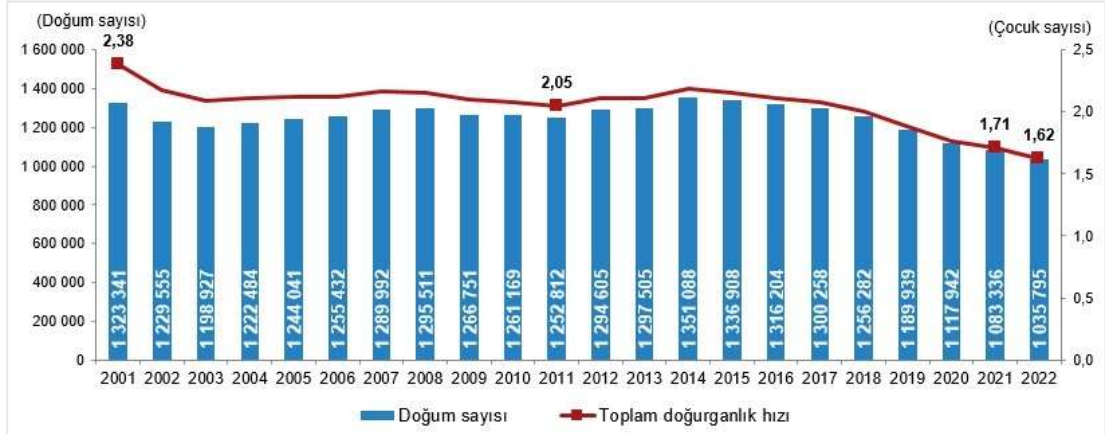
**Şekil 1.1.** Bazı Ülkeler ve Dünya Genelindeki Doğum Oranlarının 2000, 2008 ve 2020 Yıllarındaki Değişimi (Dünya Bankası 2023)



Genç nüfusun ülkelerin gelişmesi ve kalkınmasındaki öneminden kaynaklı olarak, bir ülkenin doğurganlık katsayısının yüksek olması önemli bir veridir. Bu sebeple düşüş eğiliminde olan doğum oranları ile ilgili birçok sebep araştırılmaktadır.

Türkiye’de de doğum oranları ile ilgili ciddi çalışmalar yapılmaktadır. 1950’li yıllardan itibaren Türkiye’de başlayan doğurganlık düşüşü, ilk başlarda yavaş bir tempoda ilerlemiş olup 1970’lerden sonra hızını artırarak devam etmiş ve 2000’li yıllarda kadın başına düşen toplam doğurganlık hızı, nüfusun devam edebilme seviyesi olarak bilinen 2 çocuğa kadar azalmıştır (Akadlı-Ergöçmen 1997). Günümüzde gelinen noktada ise şekil 1.2’de görüleceği üzere Türkiye’nin 2001 yılındaki toplam doğurganlık hızı 2,38 iken 2022 yılında bu oran 1,62 olmuştur (TÜİK 2023).

**Şekil 1.2.** 2001 ile 2021 Yılları Arasında Türkiye'deki Doğum Sayısı ve Toplam Doğurganlık Hızı (TÜİK 2023)



Ülkesel ve küresel bazda politikaları en çok etkileyen faktörlerden biri toplam doğurganlık hızı olduğu için politikalar oluşturulurken doğum oranını etkileyen faktörleri de derinlemesine incelemek gerekmektedir. Doğum oranını etkileyen faktörler ülkeden ülkeye, bölgeden bölgeye, kültürden kültüre hatta zaman içerisinde dahi değişebilmektedir. Doğum oranını etkileyen faktörler temelde fizyolojik/kişisel ve sosyoekonomik faktörlerdir. Fizyolojik/kişisel faktörler; genetik faktörler, geçmiş gebelik deneyimleri, sigara ve alkol gibi yaşam tarzı farklılıkları, doğum kontrolü, sağlık hizmetlerinin yeterliliği vb. şeklindedir. Sosyoekonomik faktörler ise devlet politikaları, kültürel faktörler, evlenme yaşı, kadın istihdamı, ekonomik durum/yeterlilik, göç vb. şeklindedir.

Shirahase (2000), Japonya özelinde doğurganlığın azalmasını etkileyen faktörleri inceleyen bir çalışma yaparak bu çalışmada eğitimin evlenme yaşına, evlenme yaşının ise doğum kararına güçlü bir etkisi olduğunu tespit etmiştir. Aynı zamanda artan eğitim oranı ile erkeklerin de ev işlerine katılması durumunun ortaya çıkmasına rağmen çocuk ile ilgili bakım süreçlerinin halen büyük çoğunlukla anne üzerinde hissedildiğini belirtmiştir (Shirahase 2000, 60). Bu durumun da hem eğitim ve iş hayatında aktif rol oynayan hem de çocuğunun sorumluluğunu taşıma noktasında

sorumluluğun büyük kısmını üzerinde hisseden kadın bireyler için çekimser davranıp daha az doğum yapma davranışına yol açabileceği şeklinde değerlendirilmiştir (Bernhardt 1993, 28). Osili ve Long'un Nijerya bazında inceleme yaparak analiz yaptığı çalışmada kadın eğitimi bir yıl artırmanın erken doğurganlığı 0,26 doğum azalttığı öne sürülmektedir (Osili ve Long 2008, 71). Kadın eğitimi ve doğurganlık ilişkisinin Mısır özelinde incelendiği bir çalışmada, her bir ek eğitim yılının, kadın başına düşen çocuk sayısını 0,079 azalttığını belirlenmiş ve sonuçların kadınların eğitim düzeyindeki artışın doğum oranlarını düşürme eğiliminde olduğu ortaya konulmuştur (Ali ve Gurmu 2018).

Eğitim düzeyi artan kadınların, çocuk bakımı ve sağlık hizmetlerine erişim konusunda daha bilgili olması bebek kayıplarını azaltabilmektedir. Kadın eğitim düzeyinin artması ile birlikte çocuk ölüm oranlarının azaldığı ve buna bağlı olarak doğum oranlarının da azaldığını belirten çalışmalar bulunmaktadır (Basu 2002, 1784). Bebek kayıp oranlarının yüksek olduğu koşullarda doğan bebekler için, bir kısmının öleceği varsayılarak hareket edilmektedir. Doğum istekliliğinin bulunduğu ve doğan bebeklerin yaşama olasılığının yüksek olacağı düşüncesinin hâkim olduğu koşullarda doğum sayısının da daha az olması beklenecektir.

Eğitim olanaklarına erişim imkânı daha önce kısıtlı bulunan kadınlar için, bu olanaklara sahip olduğunda eğitimi en ileri seviyeye kadar taşıma gayreti yoğun olarak görülmektedir. Örneğin, annesi eğitim imkânı bulamamış bir kız çocuğu annesi tarafından eğitimini tamamlayıp kariyer sahibi olması noktasında teşvik görmekte ve elinde bulunan bu imkânı sonuna kadar kullanarak en az lisans seviyesinde bir eğitimi tamamlayıp çalışma hayatına atılmak noktasında daha istekli davranmaktadır. Geçmişten beri eğitim konusunda ekonomik nedenler dışında pek kısıtlamaya maruz kalmayan erkek bireyler için ise eğitim olsa da olur olmazsa da bir şekilde başka bir

meslek sahibi olunabilir düşüncesi hâkimdir. Eğitimini kesintisiz tamamlama gayretindeki kadın bireyler için birinci öncelik her zaman eğitimleri ve kariyerleridir. Evlilik ve annelik ise bu aşamalar tamamlandıktan sonra geçilmesi gereken adımlar olarak görülmektedir. Dolayısıyla evlenme ve doğum yaşı da daha ileriye taşınmış olacaktır. Evlenme ve doğum yaşının ileriye taşınması ile doğurganlığı etkileyen bazı sağlık konulu faktörlerle karşılaşma ihtimali de artış göstermekte ve tıbbi sorunların tedavi edilebilme süreci kısalmaktadır. Özellikle kadın istihdamı konusunda firma/kurumlarda çocuklu kadın çalışanların bir problem şeklinde görülmesi, tercih edilmemesi çalışmak isteyen kadınların çocuk kararı alırken çekimser davranmasına sebep olmaktadır. Yıllarca eğitim ve kariyerine ciddi emek vermiş kadınlarda, çocuk sahibi olduktan sonra verdikleri emeklerin boşa gideceğine dair endişe gözlemlenmektedir. Nihayetinde, kadınların ücretli istihdamının tüm dünya genelinde toplam doğurganlık ile negatif ilişkili olduğu belirtilmektedir (Behrman ve Gonlons-Pons 2020, 707).

Kadın istihdam oranının artması kadın eğitim düzeyi ile olduğu kadar GSYH ile de doğrudan ilgili olduğu görülmektedir. Gelişen ve değişen dünyada yaşam giderlerinin artması ile orantılı olarak hane halkı geliri artış gösteremediğinde, eğitim düzeyi düşük olan kadınlar dahi özellikle hizmet sektöründe aktif olarak çalışmayı tercih etmektedir. Kadın istihdam oranının artması basit düzeyde düşünüldüğünde hane halkı gelirinin artışı şeklinde yorumlanabilir. Refah düzeyi ve çocuklara harcanabilecek kaynakların artmasının doğum oranında bir artışa sebep olmasına rağmen gelir düzeyi ile doğum oranı arasında zaman zaman pozitif zaman zaman negatif bir ilişki gözlemlenmektedir. Gelir düzeyi arttıkça kişilerin çocuklarına daha iyi olanaklar sağlayabileceğini düşünmesi ile birlikte doğum oranı artabileceği gibi doğum ve sonraki süreçlerde maddi kayıp ve işinden olma korkusu ile bu oran negatif de etkilenebilmektedir.

Özellikle enflasyonist ekonomilerde, gelecek kaygısı ve maddi belirsizlikler artış gösterdiğinden insanları yeni bir nesil konusunda daha çekimser davranmaya itebilmektedir (Lal vd. 2021, 14).

Bu faktörlerin yanı sıra kırsaldan kente göç ve sanayileşme ile de doğum oranlarının azaldığı görülmektedir. Kırsal kesimde doğum kontrol yöntemlerine erişimin kısıtlı olması, bu durumun sebeplerinden biri olarak gözlemlenmektedir. (Akça ve Ela 2012, 227-228). Kırsal kesimde çocuk büyütme için gerekli görülen imkânların kısıtlı olması nedeniyle her bir çocuğa yapılacak harcama oranı da görece yüksek değildir. Ayrıca her doğan çocuk, ileriki dönemlerde ailesi için potansiyel yardımcı konumundadır. Şehir yaşamında ise çocuk büyütme için hem maddi hem de manevi olarak görece çok daha fazla emek gerekmektedir. Köy yaşamında etraftaki insanların neredeyse tamamı akraba veya tanıdık kimseler olması nedeniyle çocukları yalnız bırakmanın kolay olmasına karşı şehir yaşamında tam tersi durum geçerlidir. Dolayısıyla bir çocuğun bakımını üstlenmek ve çocuk bakımında sosyal çevrenin yardımını almak gibi unsurlar düşünüldüğünde, şehir yaşamında bunlar için yapılan harcamaların görece büyüklüğü çocuk sahibi olmak konusunda insanların daha çekimser hareket etmelerine sebep olmuştur.

Tüm bu sayılanların neticesinde giderek azalma eğiliminde olan doğum oranı ve onu etkileyen faktörlerin; detaylıca incelenmesi, ülkeler ve nüfuslar için gündem teşkil etmesi ve buna göre politikalar üretilmesi gerekmektedir.

Bu çalışma 2009-2021 yılları arasında Türkiye’de iller bazında veri setine sahiptir. TÜİK’ten alınan veri seti kullanılarak Türkiye nezdinde doğum oranını etkileyen faktörler, panel veri analizi ve ML ile incelenecektir. Doğum oranını etkileyen faktörleri belirlerken literatürde yapılan çalışmalar ve veri setinin anlamlandırılabilir

olma durumu göz önünde bulundurulacaktır. Bu çalışmanın ilerleyen kısımlarında çalışma için faydalanılan “II. Literatür Taraması”, “III. Veri”, “IV. Metodoloji”, “V. Bulgular” ve “VI. Sonuç” bölümleri yer almaktadır.





## BÖLÜM II

### LİTERATÜR TARAMASI

Bu çalışma kapsamında öncelikle doğum oranını etkileyen faktörler incelenmiştir. Daha sonra analiz kısımlarında da kullanılacak olan eğitim, kadın istihdamı ve GSYH belirleyicilerine dair literatür araştırması yapılmıştır. Ayrıca çalışmada kullanılan analiz yöntemlerinden panel veri analizi ve ML'ye ilişkin tarama yapılmıştır. İncelenen literatürün derlemesi aşağıda sunulmaktadır.

#### 2.1. Doğum Oranını Etkileyen Faktörlere Dair Çalışmalar

Doğum oranının belirleyicilerine dair literatürde birçok çalışma bulunmaktadır. D'Addio ve D'Ercole tarafından 2005 yılında hazırlanan raporda, genel olarak Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) ülkelerindeki doğum oranlarının düşüş eğilimi ele alınmış ve özellikle doğumun ertelenmesinin etkilerine odaklanılmıştır. Her ne kadar tüm OECD ülkelerindeki doğum oranlarında belirgin bir azalma gözlemlenmişse de, bu azalmanın hızı ve düzeyi ülkelere göre farklılık göstermektedir. Araştırma, genç ve yaşlı kadınların zaman içinde farklı doğurganlık davranışları sergilediğini ve ülkelerin demografik geçiş süreçlerinin çeşitli aşamalarında olduğunu göstermektedir. İlk çocuğun geciktirilmesine rağmen, yaş ilerledikçe doğurganlıkta görülen artışın yetersiz olduğu ve bu eğilimin devam etmesi halinde birçok OECD ülkesinde doğurganlık düzeylerinin istenen seviyelere dönmesinin zor olabileceği ifade edilmektedir. Doğumun ertelenmesinin; kardeş sayısının azalması, 30'lu ve 40'lı yaşlardaki kadınlar arasında çocuksuzluğun artması ve anne ile çocukları arasındaki sağlık risklerinin artması gibi çocuk sahibi olma

sürecinde etkileri olduğu belirtilmektedir. Ayrıca, beklenen ve gerçekleşen doğum oranları arasındaki farkın arttığına da dikkat çekilmektedir. Bu bulgular, doğurganlıkla ilgili değişikliklerin ülkeler arasında farklılık gösterdiğini ve doğumun ertelenmesinin doğurganlık üzerinde önemli bir etkiye sahip olduğunu göstermektedir (D'Addio ve D'Ercole 2005).

Klasen ve Launov, 2006 yılında yaptıkları çalışmada, Çek Cumhuriyeti'ndeki ekonomik geçiş sürecinde toplam doğurganlık hızının düşüşünü incelemiştir. Yapılan analizlere göre geçişin; yüksek eğitimin daha fazla etkili hale gelmesi, iş ile çocuk sahibi olmayı zorlaştıran durumlar ve yetersiz çocuk bakım olanakları gibi nedenlerle toplam doğurganlık hızını olumsuz yönde etkilediğini göstermiştir. Ayrıca, doğurganlık düşüşünün bir sonraki doğumun gecikmesi ve doğurganlıktan daha erken vazgeçme gibi iki farklı etkenden kaynaklandığı düşünülmüştür. Eğitim seviyesi ile doğum zamanlaması arasındaki negatif ilişkinin arttığını ve yüksek eğitimli kadınların ilk çocuklarını daha uzun süre ertelediklerini, ayrıca eğitim düzeyini yükseltmek isteyenlerin tek çocuklu ailelere daha yatkın olduğu belirtilmiştir. Geçiş sürecinin, doğurganlık seviyeleri üzerinde hem zamanlama hem de miktar açısından önemli bir etkisinin olduğu vurgulanmıştır (Klasen ve Launov 2006).

2011 yılında yaptığı çalışmada, gelişmiş ve endüstrileşmiş ülkelerde düşük doğum oranlarının devlet politikalarıyla ilişkili olduğunu gösteren Rovny, aktif işgücü piyasası programlarının genellikle gençlere ve dezavantajlı gruplara yönelik iş eğitimi ve destekli istihdam sağlayarak işsizleri yeniden iş gücüne kazandırdığını belirtmiştir. Bu durum, potansiyel ve mevcut ebeveynlerin iş bulma, gelir elde etme ve çocuk yetiştirmek için mali güvenceye sahip olma olasılığını artırmaktadır. Aktif işgücü piyasası programlarını uygulayan refah sistemleri, çocuk bakımı için daha fazla maddi kaynak sağlayarak çift gelirli ailelerin potansiyelini artırmayı hedeflemektedir.

Örneğin, Kıta Avrupası ve Kuzey Avrupa ülkeleri arasında benzer işsizlik seviyeleri bulunmasına rağmen, gençlerin Kuzey Avrupa ülkelerinde daha iyi durumda olmasının bir nedeni, Kuzey Avrupa ülkelerindeki refah sistemlerinin öne çıkardığı aktif işgücü politikalarıdır. Bu politikalar, gençlerin istihdamını artırarak çocuk yetiştirme potansiyelini yükseltmekte ve dolayısıyla doğum oranlarını artırmaktadır (Rovny 2011).

Herzer, Strulik ve Vollmer, 2012 yılında yaptıkları çalışmada; doğurganlık, ölüm oranları ve gelir arasındaki ilişkiyi uzun dönemli incelemiştir. Bu inceleme sonucunda, azalan ölüm oranlarının azalan doğurganlığa sebep olduğu ve gelir artışının da doğurganlık düşüşüne katkı sağladığı tespit edilmiştir. Ancak, sadece ölüm oranlarının düşüşünün, nüfus büyümesindeki düşüşü açıklamak için yetersiz olduğu ortaya çıkmıştır. Doğurganlık değişiklikleri, ekonomik gelişmenin hem sebebi hem sonucu olarak görülmüş, bu da gelir-doğurganlık ilişkisinin demo-ekonomik gelişimde bir döngü oluşturduğunu göstermiştir. Elde edilen sonuçlar, farklı testlere karşı direnç gösterse de, sonsuz bir gelecekte gelir ile doğurganlık arasındaki lineer ilişkinin sürdürülebilir olmadığını ortaya koymuştur. Zamanla doğurganlığın gelir artışına olan tepkisinin doğrusal olmaktan çıkacağı öngörülmektedir. Şu anki verilere göre, doğurganlığın gelir artışına olan tepkisinin 20. yüzyılın sonuna kadar düzleşmeyeceği gözlemlenmektedir (Herzer, Strulik ve Vollmer 2012).

Ashraf, Weil ve Wilde, 2013 yılında yaptıkları çalışmada, doğum oranındaki azalmaların kişi başına çıktı üzerindeki etkisini nicel olarak değerlendirmektedir. Doğurganlığın büyümeyi etkileme şekli üzerine yapılan tartışmayı yeniden şekillendirmeyi ve önceki literatürde bulunan yöntemsel eksikliklerden ve anlaşılmalıklardan kaçınmayı hedeflenmiştir. Bu kapsamda, doğurganlık ve ekonomik sonuçlar arasındaki nedensel ilişkiyi belirlemek için mikroekonomik

kanıtları göz önünde bulunduran, simülasyon modelleri gibi araçlara odaklanmıştır. Model; eğitim, nüfusun büyüklüğü ve yaş yapısı, sermaye birikimi, ebeveynlerin çocuk yetiştirme sürecine harcadıkları zaman ve doğal kaynakların kullanımı gibi faktörleri dikkate almaktadır. Mikroekonomik tahminler, gelişmekte olan ülkelerdeki demografik ve doğal kaynak gelirleri ile nicel makroekonomik teorinin standart bileşenlerinin bir kombinasyonu kullanılarak oluşturulan bu model, doğurganlık değişikliğinin etkisini, Nijerya'nın doğum oranları temel alınarak incelemektedir. Bazı parametreler için 20 yıl sonra kişi başına çıktının %5,6 artacağı ve 50 yıl sonra %11,9 artacağı sonucuna varılmıştır (Ashraf, Weil ve Wilde 2013).

Müslüman çoğunluğa sahip ülkelerdeki doğurganlık seviyelerinin nelerden etkilendiğini anlamak üzerine 2015 yılında çalışma yapan Yurtseven, incelenen 33 Müslüman ülkesindeki yüksek doğum oranlarının temel sebebinin ekonomik koşullar olduğunu tespit etmiştir. Ekonomik gelişme ile doğurganlık arasında ters bir ilişki olduğu ve ülkelerin ekonomik olarak ilerlemelerinin doğum oranlarını azaltacağı öngörülmüştür. Ancak, doğurganlık kararları kültürel faktörlerin etkisi altında olduğu için bu geçişin yavaş olabileceği belirtilmiştir. Araştırma, ekonomik koşulların yanı sıra kültürel normların da doğurganlık üzerinde etkili olduğunu ve bu nedenle bu ülkelerin doğum oranlarının Batı ülkeleri ile hızlı bir uyum içinde olmayabileceğini göstermektedir (Yurtseven 2015).

Türkiye'de doğurganlık hızının, ekonomik ve sosyo-kültürel faktörlerin etkisi altında mekânsal dağılımını 2018 yılında inceleyen Aydın, Aslantaş Bostan ve Özgür, doğurganlık hızının özellikle son kırk yılda düşmüş olsa da, bölgesel farklılıkların devam ettiğini ifade etmektedir. Batı bölgelerinde düşük doğurganlık görülürken, doğu ve güneydoğu bölgeleri hâlâ yüksek doğurganlık seviyelerine sahiptir. Bulgular, Türkiye genelinde doğurganlık hızının farklı gruplara ayrıldığını ve bölgesel oto-

korelasyonun varlığını ortaya koymaktadır. Coğrafi Ağırlıklı Regresyon modeli doğurganlık hızını başarıyla açıklamaktadır (Aydın, Aslantaş Bostan ve Özgür 2018).

Rahman, Islam ve Yeasmin tarafından 2020 yılında yayınlanan çalışmada, demografik ve sağlık anket verileri kullanılarak 16 farklı ülke incelenmiştir. Çalışmanın hedefine ulaşmak için ikili lojistik regresyon rastgele etki meta-analizi ve rastgele etki meta-regresyon yöntemleri kullanılmıştır. Araştırmacılar, gelişmekte olan ülkelerde doğurganlığı etkileyen bir dizi faktörün belirlendiğini ortaya koymuştur. Kadınların yaşının doğurganlık performansındaki kritik rolünü vurgulayarak, yaşları yüksek olan kadınların evlilik süresi arttıkça daha yüksek doğurganlık sergiledikleri gösterilmiştir. Benzer şekilde, ilk evlilik yaşı da yükseldikçe düşük doğurganlığa yol açan önemli bir faktördür. Eğitim, kadınlar ve eşleri için doğurganlık performansını açıklamada kilit bir rol oynamaktadır. Ayrıca kentselleşme, kentsel alanlarda yaşayan kadınların yüksek doğurganlık riskini azaltarak doğurganlık düşüşüne neden olmaktadır. Öte yandan, kadınların servet endeksi ve vücut kitle indeksi gibi faktörlerin de doğurganlık davranışını etkilediği gözlemlenmiştir. Analiz aynı zamanda bir ülkenin kişi başına düşen sağlık harcamalarının o ülkenin doğurganlığını açıklamak için son derece kritik bir faktör olduğunu ve bu durumun diğer doğurganlık etmenlerini nasıl etkilediğini göstermiştir. Bu bağlamda, gelişmekte olan ülkelerde kadın yaşının doğurganlık performansını açıklamak için temel bir faktör olduğu ve eğitim, yerleşim yeri türü, sağlık harcamaları gibi diğer faktörlerin de doğurganlık üzerinde önemli etkileri olduğu sonucuna varılmıştır (Rahman, Islam ve Yeasmin 2020).

2021 yılında yaptıkları çalışmada, Pasifik Adası'ndaki altı ülkede doğurganlık üzerine makro düzeyde bir analize odaklanan Lal ve arkadaşları, 1990-2019 arasındaki verileri kullanarak yaptıkları analizde, doğurganlık ile kadınların güçlenmesi, sağlık,

iletişim ve yaşam maliyeti gibi faktörler arasındaki ilişkileri incelemiştir. Bulgular, doğum kontrol oranının, kadın işgücüne katılımın ve enflasyonun doğurganlıkla negatif ilişkili olduğunu göstermiştir. Kentleşmenin de negatif bir etkisi olduğu ancak daha düşük bir anlamlılık düzeyinde olduğu saptanmıştır. Gerçek GSYH ile doğurganlık arasında negatif bir ilişki gözlemlenmişse de istatistiksel olarak anlamlı değildir. Kadınların ortaöğretim kaydı, kadın nüfusu, mobil abonelik ve bebek ölüm oranı gibi değişkenlerin ise doğurganlık üzerinde sınırlı etkisi olduğu belirlenmiştir. Sonuçlar, doğurganlığı azaltmada doğurganlıkla negatif ilişkili faktörlerin ve gerçek GSYH'nin önemli rol oynayabileceğini göstermektedir (Lal vd. 2021).

#### 2.1.a. Doğum Oranları ile Eğitim Arasındaki İlişkiye Dair Çalışmalar

Amerika'daki doğum oranlarını eğitim ve yaş verileri çerçevesinde 1996 yılında inceleyen Rindfuss, Morgan ve Offutt, kadınlarda eğitim düzeyinin yükseldikçe kadınların evliliği ve doğurganlığı erteleyen bir eğilimde olduklarını göstermektedir. Bu değişimin, kadınların mesleki rollerindeki ve çocuk bakım seçeneklerindeki değişimle ilişkilendirildiği ifade edilmektedir. Ayrıca Amerika'da doğurganlık modellerinin genel olarak istikrarlı olduğu, ancak yüksek eğitimli kadınlar arasında çocuk yapma yaşlarında belirgin bir kayma olduğu belirtilmiştir. Bu değişikliğin, eğitim düzeyi yüksek olan kadınların kariyer ve annelik arasında denge kurma çabalarından kaynaklandığı vurgulanmıştır. Son olarak, eğitim düzeyi yüksek kadınların çocuk yapma yaşlarını ertelemesinin, Amerika'daki doğurganlık seviyeleri ve çocukların yaşam deneyimleri üzerindeki olası etkilerine dair bazı öngörüler sunulmaktadır (Rindfuss, Morgan ve Offutt 1996).

Eğitim harcamaları aracılığıyla kadın işgücünü, doğurganlık ve ekonomik büyüme ile ilişkilendiren Yang, 2000 yılında yaptığı çalışmasında evde öğretim süresi ve

eđitim harcamalarındaki artıřın insan sermayesi stok seviyesini artırdığı, geliřmiř ekonomilerde ise yksek eđitim harcamalarının dřk dođurgenlik, yksek kadın katılımı ve srekli byme ile bađlantılı olduđunu ifade etmektedir. Ekonomik byme srecinde kadın iřgc ile dođurgenlik arasındaki iliřkinin vurgulandıđı Tayvan rneđinde, kadın kazanlarının aile giderlerine katkı sađladıđı ve aynı zamanda aile yelerinin insan sermayesine yatırım yapılmasına destek olduđu gzlemlenmiřtir. Dřk gelirli lkelerin hane halkı eđitim harcamalarını arttırarak geliřmiř bir ařamaya geebileceđi ve sosyal, kltrel deđiřikliklerin kadın iřgc, dođurgenlik ve byme arasındaki etkileřimi řekillendirdiđi vurgulanmaktadır (Yang 2000).

Eđitim ile dođurgenlik arasındaki negatif iliřkiyi inceleyen Meisenberg 2008 yılındaki alıřmasında, 1990, 1995 ve 2000 Dnya Deđerler Arařtırması'ndan elde edilen verilerle 78 lkeyi kapsayan bir inceleme yapmıřtır. Hemen hemen tm lkelerde mevcut olan negatif korelasyonun kadınlarda erkeklere gre daha belirgin olduđunu, eđitim seviyesine gre okul sresinin daha belirgin olduđunu ve kiřisel servet seviyesinden etkilenmediđinin sonularına varmıřtır. Eđitim ile dođurgenlik arasındaki bu iliřkinin, ekonomik, sosyal ve biliřsel geliřim dzeyi dřk olan blgelerde gcl ve en geliřmiř toplumlarda ise daha zayıf olduđu gsterilmiřtir (Meisenberg 2008).

Eđitim ve dođum oranları arasındaki iliřkiyi inceleyen Eryurt ve Akadlı-Ergçmen, 2008 yılında yaptıkları alıřmalarını  temel soru erevesinde geliřtirmiřtir. İlk olarak, ailenin ocuk sayısı zerinde hangi ebeveynin eđitiminin daha etkili olduđunu belirlemeye ynelik bir soru zerinde durulmuřtur. Daha sonra, kadının dođurgenliđini etkileyen faktrleri, zellikle de annenin bydđ ailenin eđitim seviyesini, bir nceki kuřađın eđitimi ile bađlantılı olarak incelenmiřtir. Son olarak, eđitimin

doğurganlık üzerindeki etkilerinin doğum sırasına göre nasıl değiştiğine odaklanılmıştır. "2003 Türkiye Nüfus ve Sağlık Araştırması" verileri kullanılarak gerçekleştirilen bu çalışmada; annenin, babanın, annenin ebeveyninin ve diğer değişkenlerin çocuk doğurma süreçlerindeki etkileri sağkalım analizi yöntemiyle ölçülmüştür. Bulgular, ilk çocuğa geçiş aşamasında babanın eğitiminin, ardından gelen çocuklara geçiş aşamasında ise annenin eğitiminin daha etkili olduğunu belirtmektedir. İkinci ve sonraki çocuklara geçiş hızları, annenin yetiştiği ailesinin eğitim düzeyine bağlı olarak değişkenlik göstermektedir. Eğitimsiz annelerin doğum yapma olasılıklarının daha yüksek olduğu sonucu da ortaya çıkmıştır (Eryurt ve Akadlı-Ergöçmen 2008).

Becker, Cinnirella ve Woessmann, Malthusçu düşünceler döneminden modern büyümeye geçişte çocuk sayısı ile kalite arasındaki dengeyi 2010 yılındaki çalışmalarında incelemiştir. 1849'da Prusya'nın 334 ilçesine ait özel bir veri setini kullanarak 19. yüzyılda böyle bir denge olduğunu gösteren ilk kanıtları sunmuştur. Ayrıca doğurganlık ile eğitim arasındaki ilişkinin karşılıklı olduğu bulunmuştur. Eğitimin 1849'daki düzeyinin 1880-1905 arasındaki doğurganlık değişimini tahmin ettiği ortaya çıkmıştır. Gelişmiş ekonomilerin Malthusçu kısıtlamalardan kaçıp hızlı büyümeye geçiş sürecini anlamak bu noktada önem arz etmektedir. Aynı zamanda, birçok gelişmekte olan ülkenin benzer demografik değişimler yaşadığı mevcut durumda, bu konu daha da önem kazanmaktadır (Becker, Cinnirella ve Woessmann 2010).

Akça ve Ela'nın 2012 yılında yaptığı çalışmada; Türkiye'de işsizliğe etki eden faktörler arasında hızlı nüfus artışı, kırdan kente göç, yetersiz istihdam artışı, düşük eğitim seviyesi ve ekonomik krizler olduğunu belirtilmiştir. Özellikle doğurganlıkta yaşanan azalmalar, işsizlik oranlarını belirlemede önemli bir faktör olarak

değerlendirilmiştir. Bu bağlamda; kadınların işgücüne katılım oranı, gelir düzeyi, kentleşme, göç, sanayileşme ve özellikle eğitim düzeyi işsizliği etkileyen en önemli sosyoekonomik faktörler olarak öne çıkmaktadır. Türkiye'nin işsizlik sorununu çözmek için bu faktörlerin etkilerinin gerçekçi bir şekilde belirlenmesi önem arz etmektedir. Bu bağlamda eğitim en kritik faktör olarak öne çıkmaktadır (Akça ve Ela 2012).

Mısır'da kadın eğitiminin doğurganlık üzerindeki etkisini 2018 yılında inceleyen Ali ve Gurmu, ülkenin ilkokul eğitim süresinde 1988'de gerçekleşen değişikliği kullanarak kadın eğitiminde dışsal bir farklılık oluşturmayı hedeflemektedir. Bu değişiklikten etkilenen ilk grup, 1977 Ekim ayından sonra doğan bireylerdir. Bu dönüşümü incelemek ve kadınların eğitim seviyeleri ile doğurganlık durumlarını karşılaştırmak için özel bir istatistiksel analiz yöntemi kullanılmıştır. Bulgular, 5 yıllık ilkokul sürecine katılan kadınların genellikle yetişkinlik döneminde ortalama 1 yıl daha az eğitim aldıklarını göstermektedir. Ayrıca her bir ek eğitim yılının kadın başına düşen çocuk sayısını 0,079 azalttığı belirlenmiştir. Sonuçlar, kadınların eğitim düzeylerinin artmasının doğum oranlarını düşürme eğiliminde olduğunu göstermektedir (Ali ve Gurmu 2018).

Kadınların eğitim düzeyi ile doğum oranları arasındaki bağlantıya dair 2020'de çalışma yapan Cornett, yüksek eğitim alan kadınların genellikle daha az çocuk sahibi olmaları şeklindeki trende dikkat çekmektedir. Bu trend dünya genelinde çeşitli ülkeler ve kültürlerde görülmektedir. Veriler, bu ilişkinin neden-sonuç ilişkisi olduğunu gösterirken, eğitimin doğurganlığı azaltma mekanizmalarına dair çeşitli teorilerden bahsetmektedir. Bu bağlamda, demografik geçiş sürecinde olan ülkelerin, doğum oranlarını kontrol etmek amacıyla kız çocuklarının eğitimini teşvik etmeleri gerekmektedir. Kadın eğitimi, demografik geçiş modeline göre, gelişen ülkelerin

yüksek doğum ve düşük ölüm oranları döneminden geçebileceği bir evreyi işaret etmektedir. Bu süreçte doğum oranlarının kontrol altına alınması, nüfus artışını dengeleyerek iş gücünün taleplerini karşılamaya yardımcı olabilir. Ayrıca eşlerin eğitim düzeyinin doğum oranları üzerindeki etkilerini anlamak ve kız çocuklarının eğitime başlama yaşı ile doğum oranları arasındaki ilişkiyi incelemek de önemli araştırma konularıdır (Cornett 2020).

### 2.1.b. Doğum Oranları ile Kadın İstihdamı Arasındaki İlişkiye Dair Çalışmalar

Literatür verilerinin, OECD ülkelerindeki toplam doğum oranı ile kadın işgücüne katılım oranı arasındaki ilişkinin, 1980'lerin başından önce negatif bir değere sahipken sonrasında pozitif bir yöne değişmektedir. Bu bulgunun, doğurganlık ile kadın istihdamı arasındaki ilişkide zaman içinde bir değişiklik olduğuna dair kanıt şeklinde yorumlandığını söyleyen Kögel, 2004 yılındaki çalışmasında, doğurganlık ile kadın istihdamı arasındaki zamansal ilişkinin bir işaret değişikliği göstermediğini belirtmektedir. Bunun yerine, ülke arasındaki ilişkinin işaretinin tersine dönmesinin muhtemelen ölçülemeyen ülke özel faktörlerin varlığı ve doğurganlık ile kadın istihdamı arasındaki negatif zaman serisi ilişkisinin büyüklüğünde ülke bazında farklılıkların bir kombinasyonundan kaynaklandığını işaret etmektedir. Bununla birlikte 1985'ten sonra doğurganlık ile kadın istihdamı arasındaki negatif zaman serisi ilişkisinde bir azalma olduğuna dair kanıtlar da sunmaktadır (Kögel 2004).

2006 yılında yaptıkları çalışmada OECD ülkelerinde doğurganlık ile istihdam arasındaki olumlu ilişkiyi oluşturan faktörleri araştırmak amacıyla doğurganlık ve işgücü piyasasına katılım kararlarını sayısal olarak açıklayan bir teori geliştiren Da Rocha ve Luisa Fuster, işsizliğin kadınları doğumları ertelemeye yönlendirdiği ve bu durumun toplam doğum oranını azalttığını göstermiştir. Ayrıca Amerika Birleşik

Devletleri (ABD) ile İspanya arasındaki kadınların işgücüne katılımıyla ilgili farklılıkların, İspanya'daki düşük doğum oranını ABD'ye göre açıklayabileceğini belirtmiştir. Ayrıca, işgücü piyasası sorunlarının, ekonomiler arasında kadın istihdam oranları ile doğum oranları arasında olumlu bir ilişki kurabileceğini de tespit etmiştir (Da Rocha ve Luisa Fuster 2006).

Senegal'in kırsal bölgelerinde kadın istihdamının doğurganlık üzerindeki etkisini araştıran Van Den Broeck ve Maertens, 2015 yılında yaptıkları çalışmada iş bulan kadınların evlenme ve ilk çocuk sahibi olma yaşlarının arttığını ve daha az çocuk sahibi olduklarını göstermektedir. Kadın istihdamının, doğurganlık üzerindeki etkisinin yüksek olduğu ve bu durumun yoksul kadınlar için de geçerli olduğu belirtilmektedir. Bulgular, kırsal kesimdeki kadınları güçlendirmede etkili bir araç olarak kadın istihdamının önemini vurgulamaktadır. Ayrıca kadın istihdamının arttığı bölgelere veya doğrudan kadın çalışanlara yönelik aile planlaması programlarının etkinliğinin artabileceği üzerinde durulmaktadır. Doğurganlık tercihlerindeki değişimin sadece kültürel değişimle değil, ani ekonomik fırsat değişiklikleriyle de meydana gelebileceği gösterilmektedir (Van Den Broeck ve Maertens 2015).

Kadın istihdamı ile doğurganlık arasındaki küresel ilişkiyi anlamak için Behrman ve Gonalons-Pons, 2020 yılında yaptıkları çalışmada; Uluslararası Çalışma Örgütü'nden kadınların ülke düzeyinde ücretli istihdamına dair verileri, Birleşmiş Milletler'den doğurganlık ve üreme sağlığı ölçümlerini, Birleşmiş Milletler Eğitim, Bilim ve Kültür Örgütü, OECD ve Dünya Bankası'ndan ek bilgileri içeren büyük bir veri seti ile 1960-2015 dönemini kapsayan bir analiz yapmıştır. Kadınların ücretli istihdamının dünya genelinde toplam doğurganlık ile negatif ilişkilendiğini göstermiştir. Ancak bu ilişkilerin tarım dışı istihdamla sınırlı olduğu belirtilmiştir (Behrman ve Gonalons-Pons 2020).

Kadın istihdamı ile doğum oranı arasındaki ilişkiyi 2022 yılında eşbütünleşme analizi ile inceleyen Yeşilkaya, İsveç ve ABD'nin 1991-2020 verileri üzerinden kadın istihdamı ile doğum oranı arasındaki uzun vadeli ilişkiyi ele almıştır. Bulgular, her iki ülkede doğum oranı ile kadın istihdamı arasında olumlu bir bağlantının bulunduğunu göstermektedir. Ancak İsveç'teki bu ilişkinin ABD'ye kıyasla daha güçlü olduğu belirtilmiştir. İsveç'in iş-yaşam dengesi ve cinsiyet eşitliği konularında daha etkili önlemler alması ile doğum sonrası kadınların işgücü piyasasına dönüşünün ABD'ye göre daha hızlı ve etkili bir şekilde teşvik ettiği sonucuna varılmıştır. Kadın istihdamı ile doğum oranı arasındaki ilişkinin, ülkelerin sosyoekonomik yapısı ve işgücü piyasasındaki dinamiklere bağlı olarak çeşitlilik gösterebileceği vurgulanmıştır (Yeşilkaya 2022).

#### 2.1.c. Doğum Oranları ile GSYH Arasındaki İlişkiye Dair Çalışmalar

Kişi başına düşen çıktı ve ekonomik büyüme performansındaki uluslararası farkları açıklamak amacıyla Galor ve Zang tarafından 1997 yılında yapılan çalışmada, aile büyüklüğü, gelir dağılımı ve doğurganlık gibi faktörler teorik ve ampirik olarak değerlendirilmiştir. Ödünç alma kısıtlamaları altındaki ülkelerde, daha küçük aile büyüklüğü ve gelirin daha eşit dağılımının kişi başına düşen çıktıyı artırma eğiliminde olduğu ortaya konulmuştur. Diğer yandan neoklasik büyüme modelinin iş gücü veya nüfus büyüme hızının ekonomik büyümeye olan etkileri önemsiz bulunurken aile büyüklüğü ve gelir dağılımının birleşik etkisinin daha belirgin olduğu gösterilmiştir. Büyük aileler ve eşitsiz gelir dağılımına sahip ülkelerin kalkınma tuzağına düşme riskine işaret ederek, aile büyüklüğünün azaltılması ve eğitime yönelik destek gibi politika önerileri vurgulanmıştır. Ampirik bulgular, demografik geçişin ekonomik

büyümedeki rolüne dair yeni bir araştırma perspektifi sunmaktadır (Galor ve Zang 1997).

Karra, Canning ve Wilde 2017'de yayınlanan çalışmada, Nijerya'daki doğurganlık düşüşünün ekonomik büyüme üzerindeki etkisini değerlendirmiştir. Kadın başına düşen doğum oranının bir çocuk azalmasıyla 2060'a kadar kişi başına gelirin neredeyse iki katına çıkacağı ortaya konulmuştur. Ayrıca doğurganlık düşüşünün tek başına yeterli olmadığını ve ekonomik kalkınmanın bir dizi faktöre bağlı olduğunu vurgulayarak bu konuda gerçekçi beklentilere dikkat çekilmektedir (Karra, Canning ve Wilde 2017).

Koppes 2022 yılında yaptığı çalışmada, bir ülkenin doğum oranı ile kişi başına düşen GSYH arasındaki ilişkiyi araştırmaktadır. Bulgular, bir ülkenin doğum oranını etkileyen faktörleri incelemek için yapılan regresyon analizi sonucunda ölüm oranları ile doğum oranları arasında pozitif bir ilişki olduğunu; kadın eğitim seviyeleri ile ise negatif bir ilişki olduğunu göstermektedir. Ancak kişi başına düşen GSYH ile doğum oranı arasındaki ilişki belirsiz veya anlamsızdır. Bu çalışmanın eksik veriler nedeniyle bazı ülkelerin dışlanması, kullanılan modelin yazar tarafından oluşturulması gibi bazı sınırlamaları bulunmaktadır. Bu durum elde edilen sonuçların kesinliği konusunda şüpheleri beraberinde getirmektedir (Koppes 2022).

## **2.2. Panel Veri Analizine Dair Çalışmalar**

Panel veri analizi birçok araştırmada olduğu gibi doğum oranı ile ilgilenen araştırmalarda da yaygın olarak kullanılmaktadır. Doğum oranı ile diğer verilerin ilişkisini ortaya koymak açısından bu analiz yöntemine çokça başvurulmuş ve birçok anlamlı sonuçlar elde edilebilmiştir. Literatürde hem panel veri analizinin kendisinin hem de doğum oranı ile ilişkisinin çokça yer edindiği gözlemlenmektedir.

Yaffee, 2003 yılında yayınladığı çalışmasında, panel veri analizinin sosyal ve davranışsal bilim araştırmacıları arasında giderek popüler hale geldiğini belirtmektedir. Çalışmasında, panel verileri analizindeki temel modelleri özetlemekte ve bunların göreceli avantajları ve dezavantajları hakkında bilgi vermektedir. Ayrıca sabit veya rastgele etkiler modellerini kullanmak için bir testi ele almaktadır (Yaffee 2003).

Hsiao, 2005 yılında yayınlanan çalışmasında, panel verilerinin birçok avantajı olmasına rağmen bunların her durumda kesin bir çözüm sunmadığına dikkat çekmiştir. Panel verilerinin belirli eylemlerin, tedavilerin veya daha genel politikaların etkilerini izole etme gücü, istatistiksel araçların varsayımlarının veri üretim süreciyle uyumluluğuna bağlı olduğunu belirtmiştir. Bir panelin zenginliğini ve spesifik özelliklerini kullanmak için uygun bir yöntem seçerken, panel verilerinin ekonomik konularda araştırma yapmada sunduğu avantajları, panel verilerinin ve bu verileri analiz etmek için önerilen ekonometrik yöntemlerin sınırlarını, istatistiksel çıkarım prosedürlerinin ve veri üretim sürecinin varsayımlarının uyumlu olup olmadığını ve parametre tahmincilerinin etkinliğini artırmanın yollarını düşünmenin önemli olduğuna dikkat çekilmiştir (Hsiao 2005).

Akın'ın 2014 yılında yapmış olduğu çalışmada, 1980 ile 1998 yılları arasındaki 14 Orta Doğu ülkesinde yetersiz büyüme ve yüksek doğurganlık arasındaki ilişki panel veri analizi ile incelenmiştir. Bu sorunu çözmek için öncelikle eğitim düzeylerini artırmak ve kadınların işgücüne katılımını teşvik etmenin aile planlaması hizmetleriyle birleştirilmesinin önemli olduğu sonucuna ulaşılmıştır (Akin 2005).

Nargeleçekenler'in 2009 yılında hazırladığı çalışmasında, ekonometrik açıdan panel veri analizini ele almış ve bu alandaki yenilikleri analiz etmiştir. Panel veriler, hem birim hem de zaman boyutlarını içerdiği için yatay kesit ve zaman serisi

verilerinden daha kapsamlı bilgiler sağlamaktadır. Özellikle makroekonomik ve finansal serilerin analizinde panel verilerin kullanımı, daha güçlü sonuçlar elde etmeyi mümkün kılmaktadır. Çalışma, geleneksel panel veri modelleri ve son yıllarda geliştirilen birim kök ve eşbütünleşme (cointegration) analizlerini kapsamakta ve teorik konuların makroekonomik veya finansal beş farklı alana nasıl uygulandığını detaylı bir şekilde ele almaktadır (Nargeleçekenler 2009, 1-265).

Güneydoğu ve Güney Asya'daki doğum oranlarının belirleyicileri üzerine 2009 yılında panel veri analizi yapan Teguh, hızlı bir şekilde gerçekleşen sosyal ve ekonomik gelişme, kadınların ikincil okula kayıt oranındaki artış ve kentsel nüfusun artışı gibi faktörlerle doğum oranlarının belirgin bir şekilde azaldığını göstermektedir. Bölgedeki yüksek doğum oranını etkileyen temel faktörlerden biri, beş yaş altı çocuk ölüm oranıdır. Yüksek ölüm oranı, çocuk kaybı riskini artırarak ebeveynleri daha fazla çocuk yapmaya yönlendirebilmektedir. Şaşırtıcı bir şekilde doğum oranı, gelir artışıyla yükselmektedir. Doğum oranını düşürmenin en etkili yolu, sağlık hizmetlerine erişimi artırmak ve bebek ölüm oranını azaltmak olduğu belirtilmiştir (Teguh 2009).

2015 yılında yayınlanan çalışmasında panel verilerini kullanıp zaman ve kesit serilerini birleştirdiği çeşitli regresyon modellerini inceleyen Uyar, iki farklı uygulama örneği üzerinde analizler yapmıştır. Bunlardan ilki Türkiye'deki kadın işgücüne katılım oranları, ikincisi ise doğrudan yabancı yatırımların belirleyicileridir. Elde edilen sonuçlar, anlamlı katsayıların bağımlı değişken üzerinde önemli etkilere sahip olduğunu ve rassal etkili lojistik modellerin panel verisini içermesi gerektiğini ortaya koymaktadır. Bulgular, özellikle mikro panellerde model varsayımlarının detaylı bir şekilde incelenmesi ve farklı dağılımlara sahip yanıt değişkenlerini içeren modellere odaklanılması önerilerinde bulunmaktadır (Uyar 2015).

2002-2012 yılları arasında İran'da boşanma, evlenme, kentleşme ve işsizlik gibi bazı makro sosyo-ekonomik faktörlerin toplam doğurganlık hızı (TFR) üzerindeki etkilerini inceleyen çalışmada, panel veri ile analiz yapılmıştır ve boşanma oranının TFR'yi olumsuz etkilediğini, evlenme oranı ile TFR arasında ise olumlu ve anlamlı bir ilişki olduğunu ifade etmiştir. İran'da azalan doğurganlığın sadece sosyo-ekonomik faktörlere değil, aynı zamanda kültürel ve bağlamsal değerlere de bağlı olduğunu belirtmektedir (Jafari vd. 2016).

Rayhan, Akter ve Islam'ın 2018 yılında yaptıkları çalışmada, Güney Asya ülkelerindeki doğum oranı düşüşünün makroekonomik belirleyicileri incelenmiştir. 1990-2015 dönemine ait veriler; Bangladeş, Hindistan, Sri Lanka, Nepal, Pakistan, Bhutan ve Maldivler gibi yedi Güney Asya ülkesinden alınmıştır. Yapılan analizler; kentselleşme, bebek ölüm oranı, eğitim, kişi başına Gayrisafi Milli Hasıla (GSMH) ve kadın işgücüne katılım oranının doğum oranını anlamlı düzeyde etkilediğini ortaya koymaktadır. Panel Doğrulama Standart Hatası (PCSE) modelinin sonuçları, kişi başına GSMH, kadın işgücüne katılım oranı, eğitim ve kentselleşme artışının doğum oranını azaltacağını, bebek ölüm oranının azalmasının da doğum oranının azalmasına yol açacağını göstermektedir. Bu sonuçlara göre Rayhan, Akter ve Islam nüfus artış hızını kontrol etmeyi amaçlayan politika yapıcılara, özellikle bu faktörlere odaklanmalarını önermektedir (Rayhan, Akter ve Islam 2018).

Iwasaki ve Kumo, 2005-2015 dönemine ait panel verilerini kullanarak Rusya'nın doğum oranının bölgesel belirleyicilerini deneysel olarak incelemiş ve yapmış oldukları sistem sonucunda ekonomik büyüme ve istihdam oranı verilerinin doğurganlık hızına olumlu; göç, yoksulluk gibi verilerin ise olumsuz etki yaptığını saptamıştır (Iwasaki ve Kumo 2020, 201-202).

2016 yılında yayınlanan çalışmasında, kadın istihdamının doğum oranı üzerindeki etkilerini inceleyen Emara, 1990-2011 arasındaki dönemi kapsayan 29 gelişmekte olan ülkenin verilerini kullanarak panel veri analizi yapmıştır. Ülkeler arası farkları yakalamak için üç önemli aile politikası değişkenini içeren bir aile politikası endeksi oluşturmuştur. Bu değişkenler; anneler için ücretli izin süresi, ücretli izin ücretinin yüzdesi ve emzirme kapsamının süresidir. Ayrıca sabit ve zaman etkilerini dikkate almak için coğrafi konum ve zaman etkilerini eklemiştir. Ampirik sonuçlar, kadın işgücü katılımındaki artışın doğum oranını azalttığını, ancak bu negatif etkinin zaman içinde azaldığını göstermektedir. Öte yandan çalışmanın bulguları, daha esnek aile planlaması politikalarının doğurganlığı artırıcı bir etkisi olduğunu da ortaya koymaktadır. Son olarak zamanın, kadın işgücü katılımı ile doğurganlık arasındaki negatif ilişki üzerinde etkili olduğunu ve bu negatif etkinin zamanla azaldığı tespit edilmiştir (Emara 2016).

### **2.3. Makine Öğrenmesine Dair Çalışmalar**

Al-Jarrah ve arkadaşlarının 2015 yılında yaptığı çalışmaya göre yaklaşan teknolojiler ve bu teknolojilere bağlı cihazlarla birlikte, önümüzdeki birkaç yıl içinde büyük miktarlarda veri oluşması öngörülmektedir. Mevcut durumda, bilgisayar mühendislerinin çevreye minimal etki ile bilgisayarları ve alt sistemleri etkili bir şekilde tasarlamasını inceleyen sürdürülebilir bilişim alanında, mevcut yapay zekâ ve ML sistemleri genellikle performansa odaklanmaktadır. Ancak, bu sistemler özellikle büyük veri kümelerinde öğrenme görevi olduğunda genellikle yüksek hesaplama maliyetine ihtiyaç duymaktadır. Bu çalışmada, büyük veri yoğun alanlarda model verimliliği, öğrenme için hesaplama gereksinimleri ve düşük hesaplama maliyeti

sağlayan yeni algoritmik yaklaşımları incelenmiş ve zorluklar ele alınmıştır (Al-Jarrah vd. 2015).

L'Heureux ve arkadaşları 2017 yılında yayınlanan çalışmalarında; büyük veri devrimi ile işlemlerin optimize edilmesinin, içgörü keşfinin güçlendirilmesinin ve karar verme süreçlerinin iyileştirilmesinin vaat edilerek yaşam tarzımız, çalışma şeklimiz ve düşünce biçimimizin dönüştürülmesinin hedeflendiğini belirtmektedir. Bu potansiyeli gerçekleştirmek için büyük veriden değer çıkarma yeteneği, veri analitiği aracılığıyla sağlanmaktadır. Bu noktada; ML, veriden öğrenme ve veri odaklı içgörüler, kararlar ve tahminler sağlama yeteneği nedeniyle önemli bir rol oynamaktadır. Ancak, geleneksel ML yöntemleri, farklı bir dönemde geliştirilmiş olup bu yeni bağlamda geçerli olmayan bir dizi varsayıma dayanmaktadır. Bu kırık varsayımlar, büyük verinin özellikleriyle birleşerek geleneksel tekniklere engel oluşturmaktadır. Bu çalışmada, büyük veri ile ilgili ML zorlukları derlenip özetlenmiştir. Bu zorluklar, neden-sonuç ilişkisine vurgu yapılarak düzenlenmiştir. Ayrıca, çeşitli zorluklarla başa çıkma yeteneğine sahip yeni ML yaklaşımları ve teknikleri, uygulayıcılara kendi kullanım durumları için uygun çözümleri seçmelerine yardımcı olacak şekilde tartışılmıştır. Son olarak zorluklar ve yaklaşımlar arasındaki ilişkiyi gösteren bir matris sunulmuş ve büyük veri ile ML alanındaki ileri araştırmalar için güçlü bir temel oluşturmuştur (L'Heureux vd. 2017).

El Bouchefry ve De Souza tarafından "Knowledge Discovery in Big Data from Astronomy and Earth Observation" kitabının bir bölümü olarak 2020'de yayınlanan çalışma, ML tekniklerine giriş yapmakta ve birkaçını değerlendirmektedir. En olgun ML yaklaşımlarından biri olan induktif öğrenime odaklanmaktadır ve denetimli, denetimsiz, yarı denetimli ve takviyeli öğrenme türlerini tanımlamaktadır. ML tabanlı yöntemler, bu sistemleri incelemek için geleneksel istatistiksel yöntemlere göre

avantaj sağlar. Çünkü öncekiler gibi gerçekçi olmayan varsayımları dayatmaz, eksik verileri çıkarabilir ve uzman açıklaması yükünü azaltabilir. Bu nedenle yer bilimi alanında ML yöntemlerinin daha yaygın bir şekilde benimsenmesi, bilimin hızını ve kalitesini büyük ölçüde artırma potansiyeline sahiptir. Ancak bu avantajlara rağmen ML teknikleri yer bilimi alanında geniş çapta benimsenmemiştir. Bunun nedenleri; ML araştırma topluluğu ve doğal bilimciler arasındaki iletişim ve işbirliği eksikliği, kolay erişilebilir araç ve hizmetlerin eksikliği ve sağlam eğitim ve test veri kümelerinin gerekliliğidir. Bu engellerin, işbirliğine finansal destek sağlanması ve ML kullanımını kolaylaştıran araç ve hizmetlerin geliştirilmesi yoluyla aşılabileceği söylenmektedir (El Boucheffy ve De Souza 2020).

Gradient boosting (GB) makineleri, geniş bir yelpazede uygulamalarda önemli başarılar gösteren güçlü ML teknikleridir. Uygulamanın özel ihtiyaçlarına göre yüksek seviyede özelleştirilebilirler. Örneğin farklı kayıp fonksiyonlarına göre öğrenilebilirler. Natekin ve Knoll'un 2013'te yayınlanan çalışması, gradient boosting yöntemlerinin metodolojisine odaklanarak ML modellenmesinin tüm aşamalarını kapsayan tanımlayıcı örneklerle teorik bilgi sunmaktadır. Model karmaşıklığı ile ilgili konular ele alınmıştır. Üç pratik GB uygulaması sunulmuş ve kapsamlı bir şekilde analiz edilmiştir. Hem teorik çerçeve hem de tasarım seçenekleri açıklanmıştır. GB makinelerinin yetenekleri, gerçek dünya uygulamaları üzerinde araştırılmış ve her durumda, GB makineleri doğruluk ve genelleme açısından mükemmel sonuçlar vermiştir. Ayrıca GB makineleri, sonuçta oluşan model tasarımına ek bir bakış açısı sunarak modelleme etkilerinin daha derinlemesine incelenmesine ve analiz edilmesine olanak tanımıştır (Natekin ve Knoll 2013).

Nyoni, Tatenda ve Nyoni tarafından 2021'de yayınlanan çalışmada ABD'deki toplam doğum oranlarını analiz etmek için yapay sinir ağı yöntemi kullanmıştır.

ABD'nin, mevcut durumda nüfus artışını sürdürebilecek seviyenin altında doğum oranlarına sahip olduğunu belirtilmiştir. Bu durumun yakın gelecekte genç iş gücü bulunurluğu, sağlık harcamalarındaki artış ve sosyal güvenlik maliyetleri gibi önemli olumsuz etkileri olabileceğine vurgu yapılmıştır. Çalışmada kullanılan yıllık veriler 1960-2018 dönemini kapsamakta olup dış örneklem dönemi 2019-2030'u içermektedir. Uygulanan modelin artıkları ve tahmin değerlendirme kriterleri, modelin ABD'deki toplam doğum oranını tahmin etme konusunda istikrarlı olduğunu göstermektedir. Araştırmanın sonuçları, Amerika'daki yıllık toplam doğum oranlarının dış örnekleme dönemi boyunca muhtemelen kadın başına ortalama 1,9 doğum olacağını ortaya koymaktadır. Bu nedenle, ABD hükümetinin özellikle 18-30 yaş aralığındaki kadınları doğurganlığı teşvik etmek amacıyla teşvikler sunmaya ve çocuk yetiştirmenin maliyetini düşürmeye yönelik önlemler alması önerilmektedir (Nyoni, Tatenda ve Nyoni 2021).

Literatürde ülkelere ait ekonomik, sosyal ve fiziksel özelliklerin doğum oranına etkilerinin panel veri analizi, makine öğrenmesi ve çeşitli yöntemlerle incelendiği araştırmalar bulunmaktadır. Ancak, hem bu iki yöntemi birlikte inceleyen hem de ayrı ayrı bu iki yöntemle Türkiye'de 2009-2021 yılları arasında bu konuda inceleme yapan bir araştırmaya rastlanmamıştır. Bu araştırmada diğer çalışmalardan farklı olarak, 2009 - 2021 yılları arasında Türkiye'de iller bazında doğum oranını etkileyen faktörler panel veri analizi ve makine öğrenmesi ile incelenecek, bu iki farklı metodun çıktıları karşılaştırılacaktır.

## BÖLÜM III

### VERİ

Bu tezde kullanılan veriler ağırlıklı olarak demografik özellikler taşımakla beraber bazı makroekonomik verilerden de yararlanılmıştır. Aşağıda verileri tanıtırken veri kaynakları ve değişkenlerin nasıl hesaplandığıyla ilgili bilgiler yer almaktadır.

- **Doğum\_oranı:** Modeldeki bağımlı değişkendir. Literatürde genel olarak toplam doğurganlık oranı şeklinde ifade edilmekte ve  $TFR = \sum (\text{Doğurganlık Hızı}) \times \text{Katsayı}$  formülü ile hesaplanmaktadır (Teguh 2009). Bu çalışmada ise bir ilde aynı yıl içerisinde toplam doğum sayısının toplam nüfusa oranınının 100 ile çarpılması sonucu hesaplanmaktadır.

$$\text{Doğum Oranı} = (\text{Toplam Doğum Sayısı} / \text{Toplam Nüfus}) \times 100$$

Bu formülde, toplam doğum sayısı belirli bir dönem ve ildeki toplam doğum sayısını, toplam nüfus ise aynı dönem ve ildeki toplam nüfusu temsil eder. Bu oran 100 ile çarpılarak 100 kişi başına düşen doğum sayısını ifade eder.

- **Ln doğum oranı:** Doğum oranının daha durağan bir seri haline getirilmesi amacıyla doğal logaritmasının alınmış halidir.
- **Kadın oranı:** Bir ilde aynı yıl içerisindeki kadın nüfusunun toplam nüfusa oranlanmasıyla hesaplanır.

$$\text{Kadın Oranı} = \text{Kadın Nüfusu} / \text{Toplam Nüfus}$$

- **Evli oranı:** Bir ilde aynı yıl içerisindeki evli birey sayısının toplam nüfusa oranlanmasıyla hesaplanır.

$$\text{Evli Oranı} = \text{Evli Birey Sayısı} / \text{Toplam Nüfus}$$

- **Lisans ve üzeri oranı:** Bir ilde aynı yıl içerisindeki lisans veya üzeri mezuniyete sahip birey sayısının toplam nüfusa oranlanmasıyla hesaplanır.

Lisans ve Üzeri Oranı = Lisans veya Üzeri Mezuniyete Sahip Birey Sayısı /  
Toplam Nüfus

- **Kadın istihdam oranı:** Bir ilde aynı yıl içerisindeki çalışan kadın sayısının kadın nüfusuna oranlanmasıyla hesaplanır.

Kadın İstihdam Oranı = Çalışan Kadın Sayısı / Kadın Nüfusu

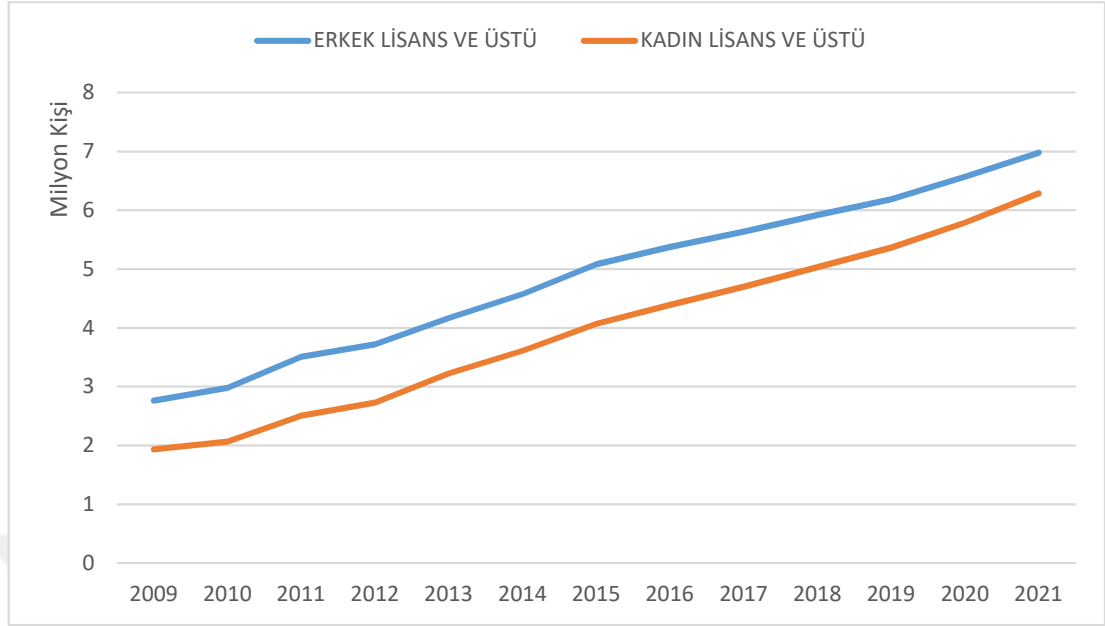
- **GSYH:** Bir ilde, ilgili yıl içerisinde kişi başına düşen dolar bazında reel GSYH değeridir.

- **Ln GSYH:** Kişi başına düşen GSYH değerinin daha durağan bir seri haline getirilmesi amacıyla doğal logaritmasının alınmış halidir.

- **Covid Kuklası:** Bir ilde; ilgili yılın 2020 veya 2021 olduğu durumlarda 1, tersinin geçerli olduğu durumlarda 0 şeklinde kodlanmasıyla hesaplanır.

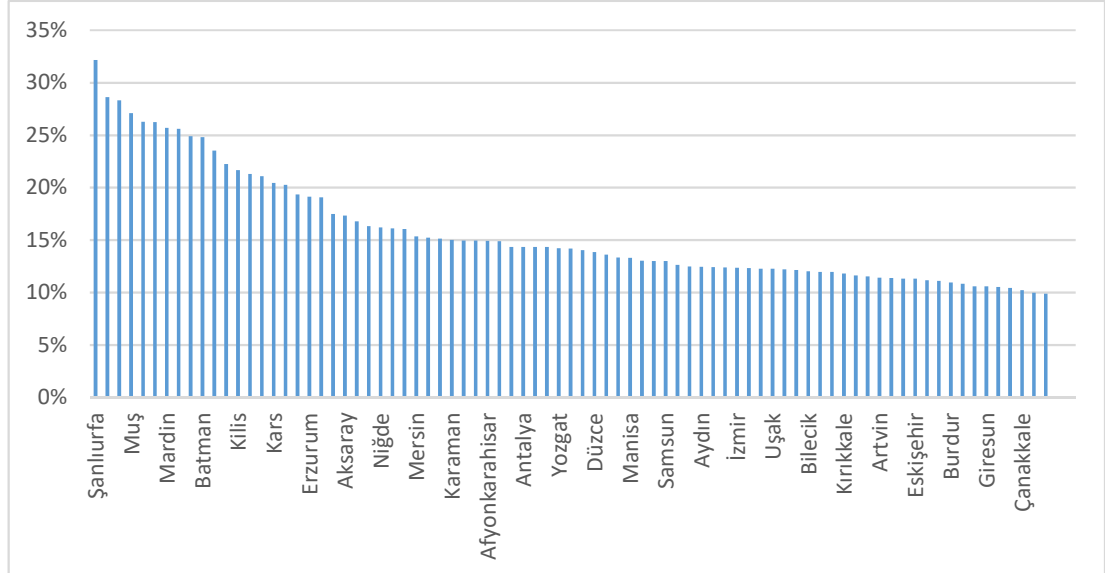
- **Göçmen Kuklası:** Bir ilde; ilgili yıl içerisinde geçici koruma kapsamında gelen göçmen sayısının toplam nüfusa oranının %5'ten büyük olduğu durumlarda 1, tersinin geçerli olduğu durumlarda 0 şeklinde kodlanmasıyla hesaplanır.

**Şekil 3.1.** 2009-2021 Yılları Arasında Lisansüstü Dereceye Sahip Kişi Sayısının Cinsiyet ve Yıllara Göre Değişimi (TÜİK 2023)



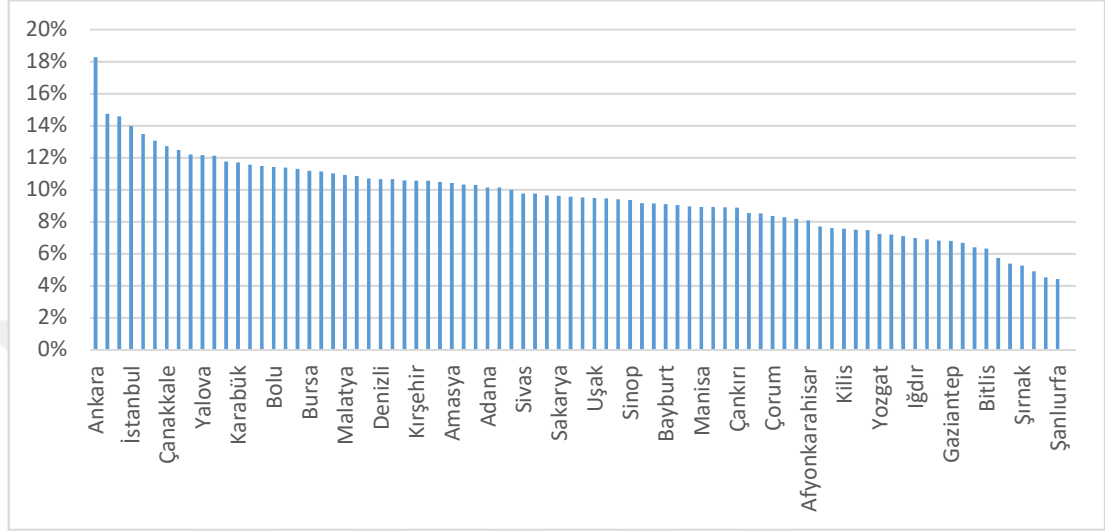
Şekil 3.1’de ise lisansüstü dereceye sahip insanların cinsiyet ve yıllara göre değişimi görülmektedir. 2021 yılsonu itibarıyla herhangi bir lisansüstü dereceye sahip kişi sayısının 12 yıl içerisinde yaklaşık 5 milyondan yaklaşık 13 milyona ulaştığı görülmüştür.

**Şekil 3.2.** İllere Göre 13 Yıllık Doğum Oranı Ortalamaları (TÜİK 2023)



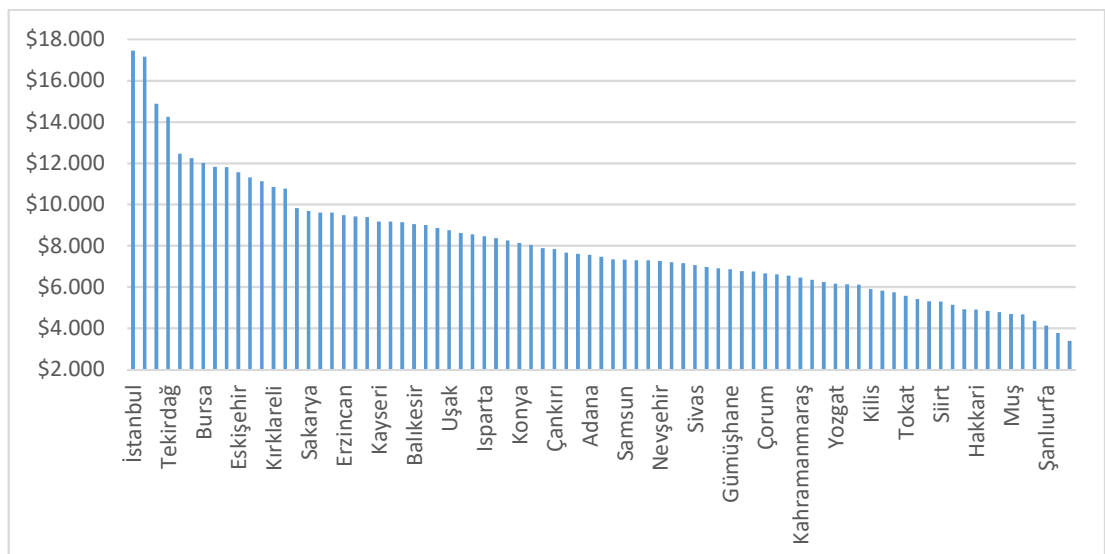
Şekil 3.2’de illere göre 13 yıllık doğum oranı ortalamalarına yer verilmiştir. Doğum oranı en yüksek 3 il Şanlıurfa, Şırnak ve Ağrı iken doğum oranı en düşük 3 il ise Çanakkale, Kırklareli ve Edirne olmuştur.

**Şekil 3.3.** Lisans ve Üzeri Bir Dereceye Sahip İnsan Sayısının Şehir Nüfusuna Oranının 13 Yıllık Ortalamaları (TÜİK 2023)



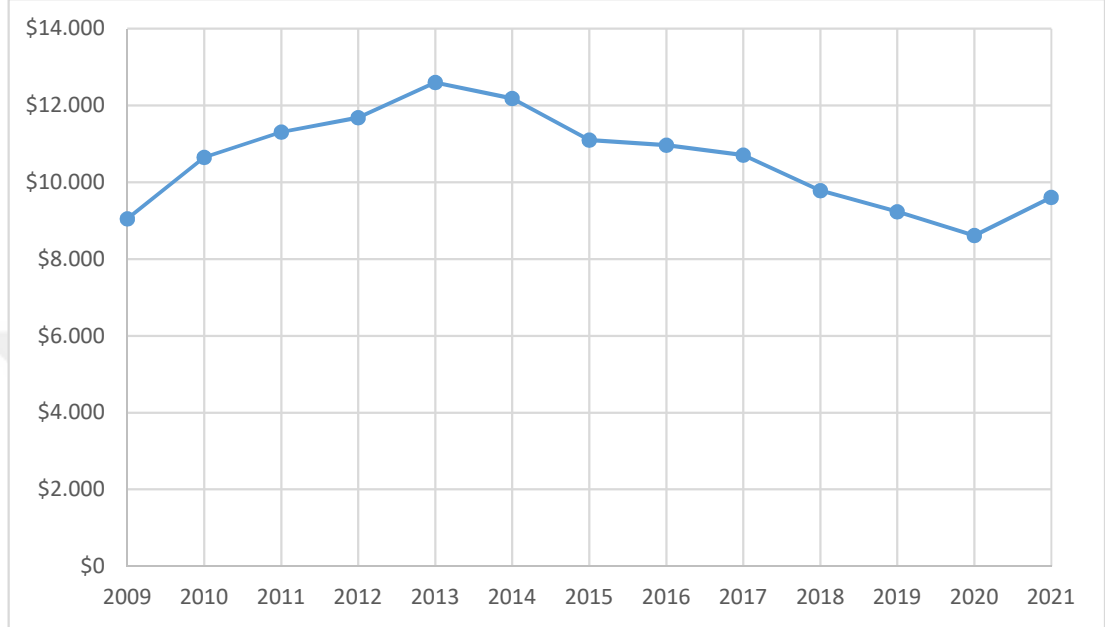
Şekil 3.3’te ise lisans ve üzeri herhangi bir dereceye sahip insan sayısının şehir nüfusuna oranının 13 yıllık ortalamaları gösterilmektedir. Akademik eğitimi yüksek nüfus oranını gösteren bu grafiğe göre ortalama eğitim seviyesi en yüksek 3 il Ankara (%18.2), Eskişehir ve İzmir olurken en düşük 3 il sırasıyla Muş, Ağrı ve Şanlıurfa (%4.4)’dır.

**Şekil 3.4.** İllerin 13 Yıllık Periyotta Kişi Başına Düşen GSYH Ortalaması (TÜİK 2023)



Şekil 3.4'te illerin 13 yıllık periyotta ABD Doları cinsinden alınmış kişi başına düşen GSYH'lerinin ortalamasını göstermektedir. Bu grafiğe göre en yüksek 3 il İstanbul, Kocaeli ve Ankara iken en düşük 3 il Şanlıurfa, Van ve Ağrı'dır.

**Şekil 3.5.** Türkiye'nin kişi başına düşen GSYH'sinin Yıllara Göre Değişimi (TÜİK 2023)



Şekil 3.5'te Türkiye'nin kişi başına düşen GSYH'sinin ABD Doları cinsinden yıllara göre değişimi gösterilmektedir. Reel olarak değişimleri görebilmek adına bu değişimleri ABD Doları cinsinden değerlendirmek daha faydalı olacaktır.



## BÖLÜM IV

### METODOLOJİ

Bu tezde doğum oranını etkileyen faktörlerin tespit edilmesi adına bazı makroekonomik verilerle beraber çoğunlukla demografik veriler kullanılarak öncelikle panel veri analizi ve bunun yanında ML incelemeleri yapılmıştır.

Kullanılan veri, panel veri formatında olduğu için veri setine uygun 3 ana model kullanılmıştır. Modellerde bağımlı değişken doğum oranı; bağımsız değişkenler ise lisans ve üzeri eğitim oranı, kadın istihdam oranı, nüfustaki kadın oranı, gayri safi yurt içi hasıla (GSYH) seviyesi, evli nüfus oranı, covid19 ve göçmen kukla değişkenleri şeklindedir. Kullanılan modeller havuzlanmış OLS, sabit etkiler modeli ve rastgele etkiler modelidir. Hausman testine bakılarak hangi modelin kullanılmasının daha doğru olduğuna karar verilmiştir. Panel veri analizine uygunluğu açısından her bir değişkene birim kök yani durağanlık testi yapılmıştır. Ayrıca şehirlerin kendine ait etkilerini görebilmek adına şehir kukla değişkeni kullanılarak bir haritalandırma yapılmıştır. Kullanılan ekonometrik modeller ve testler için Stata15 programı kullanılmıştır (Torres-Reyna, 2007).

ML için ise şehirlerdeki doğum oranı değişimini yıldan yıla hesaplayıp eğer artış varsa 1 azalış varsa 0 şeklinde ifade ederek yeni bir araştırma konusu ortaya konulmuştur. Artış ve azalış durumları bir sınıflandırma problemi olarak ortaya atılmış ve bu problemin çözümü için yeni bir yöntem olan ML modelleri kullanılmak istenmiştir. Farklı ML sınıflandırma modelleri çalıştırılmış ve doğum oranındaki değişim durumunu en iyi öğrenen model tespit edilmiştir.

ML modelleri ile elde edilen sonuçlar panel veri analizi sonuçları ile kıyaslanmış ve model değişkenleri açısından da farklılıklar incelenmiştir. ML modelleri için Python3.8 programı kullanılmıştır (Brownlee, 2016).

Sıradaki başlıklarda öncelikle panel veri analizinden sonrasında ML uygulamasından bahsedilecektir.

#### 4.1. Panel Veri Analizi

Verinin hızla önem kazandığı dünyada veriyi analiz etmenin de önemi her geçen gün artmaktadır. Fakat bu verilerden anlamlı sonuçlar çıkararak işe yarar şekilde kullanabilmek için bazı yöntemlere ihtiyaç duyulmaktadır. Panel veri analizi de ekonometri alanında kullanılan önemli veri anlamlandırma yöntemlerinden bir tanesidir. Bu yöntem birçok bilimsel çalışmanın temelini oluşturmakta ve yaygın olarak kullanılmaktadır. Panel veri analizi zaman ve kesite bağlı olarak elde edilen veriler ile anlamlı bir sonuç elde edilmesine olanak sağlayan ekonometrik bir analiz yöntemidir. Yani zamana göre çapraz kesitsel verilerin birleşiminden oluşan verilere panel veri; bu tür verileri analiz etmek için kullanılan ekonometrik-istatistiksel analizlerin birleşimine de panel veri analizi denilmektedir. Genel bir panel regresyon modeli aşağıdaki gibi temsil edilmektedir (Tatoğlu 2016, 4):

$$Y_{it} = \alpha_{it} + \beta_{it}X_{it} + \mu_{it} \quad i = 1, \dots, N; t = 1, \dots, T$$

Bu denklemde:

- Y: Bağımlı değişken
- X: Bağımsız değişken
- $\alpha$ : Sabit parametre
- $\beta$ : Eğim parametreleri
- $\mu$ : Hata terimi

- $i$ : Birim indisi,  $t$ : Zaman indisini göstermektedir.

Yaffee'ye göre panel veri analizi, sosyal ve davranış bilimleri arařtırmacıları arasında popüler hale gelen bir uzunlamasına veri analizi yöntemidir. Bu yöntem, mekânsal ve zamansal boyutlarda karmařık bir regresyon analizi ailesi olarak tanınır. Farklı yerlerden gelen veriler veya zaman serilerinin ayrı ayrı analiz için çok kısa olması durumunda, panel veri analizi, verilerin uzunlamasına analiz edilmesinin en etkili yollarından biridir. Analiz için seriler yeterince uzun olsa bile, panel veri analizi, belirli bir çapraz kesit birimine özgü zaman içindeki deęişikliklere odaklanmak için geniş bir teknik yelpaze sunmaktadır. Bu nedenle, sosyal veya davranıřsal arařtırmacılar için önemli bir araç olarak deęerlendirilebilmektedir (Yaffee 2003, 3).

Hsiao'nun çalıřması panel veri kullanmanın avantajları hakkında fikir vermektedir. Bu çalıřmaya göre panel veri; çapraz kesit verilerinden daha fazla serbestlik derecesi içerdięinden ekonometrik tahminlerin etkinlięini artırmakta, özellikle insan faktörünün olduęu analizlerde tek bir çapraz kesit verisine göre daha iyi bir yakalama kapasitesi sunmakta, hesaplama ve çıkarımı basitleřtirmekte ve analiz hatalarını en aza indirmektedir (Hsiao 2005, 146-149).

Panel veri setleri genellikle farklı bireylere, firmalara, ülkelere, şehirlere veya başka bir kesite birçok zaman noktasında toplanan verilerden oluşmaktadır (Pesaran, 2015). Bu analiz türü, farklı farklı alanlarda, özellikle ekonometri, demografi, iktisat, sosyoloji ve saęlık bilimleri gibi alanlarda çokça kullanılmaktadır.

Panel veri analizi belli bařlı bileřenlere sahiptir. Bunlar bu çalıřmadaki veri detaylarına göre örneklendirerek anlatılacaktır. Çapraz kesit verileri; illerin aynı yıl içerisindeki deęerleri ve oranları, zaman serisi verileri ise yıllara göre farklılařmasını deęiřkenler bazında (eęitim, kadın istihdamı vb.) ifade etmektedir.

Panel veri modelleri genellikle sabit etkiler ve rastgele etkiler gibi özel yapılar içerir ve bu modellere ait parametreler çeşitli ekonometrik yöntemlerle tahmin edilir. Panel veri analizi için havuzlanmış OLS, sabit etkiler, rastgele etkiler, hata bileşenleri, en küçük kareler, fark ve farklar arasındaki fark şeklinde çeşitli yöntemler kullanılmaktadır (Nargeleçekenler 2009, 8-59).

Bu çalışmada panel veri analizi için havuzlanmış OLS, sabit etkiler ve rastgele etkiler modelleri kullanılmış ayrıca hausman ve birim kök testleri yapılmıştır.

#### 4.1.a. Havuzlanmış OLS Modeli

Havuzlanmış veri tabanı, kesitsel ve zaman serisi verilerinin özelliklerini bir araya getirir. Kesitsel veri gibi, birçok bireyi tanımlar; zaman serisi veri gibi ise her bir bireyi zaman içinde tanımlar. Bu nedenle, analist için önemlidir çünkü araştırılan varlıkların zamansal dinamikleri ve bireysel özellikleri hakkında bilgi içermektedir (Dielman 1983, 111).

Havuzlanmış OLS modeli, panel verilerine OLS yöntemini uygular. Bu model, gözlemlenen varlıklara özgü etkilerin olmadığını varsayar. Yani, veri setindeki tüm varlıkların aynı temel özelliklere sahip olduğu düşünülmektedir. Bu nedenle, bireyler arasında etkilerini göz ardı eder. Birey grupları arasında bağımlılık olmadığını varsayarak, sabit etki regresyonu varsayımlarının birinci varsayımı sağlanır. Bu durum, hata teriminin bağımsız değişkenlerle korelasyon göstermediğini ifade eder. Havuzlanmış OLS modelini tahmin etmek, diğer OLS modeli varsayımları karşılandığı sürece regresyon katsayılarının yanlı olmayan tahminlerini elde eder.

Model, farklı zaman dilimlerindeki gözlemleri bir havuzda birleştirir ve bunların belirli firmalara ait olduğunu göz ardı eder. Bu şekilde, tüm gözlemler tek bir gruptan gelmiş gibi ele alınır ve panel verileri etkili bir şekilde kesitsel bir veri setinde

birleştirilir. Model katsayıları daha sonra OLS kullanılarak tahmin edilebilir (Tilburg Science Hub 2023).

#### 4.1.b. Sabit Etkiler Modeli

Sabit etkiler modeli şu şekilde temsil edilmektedir:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mu_{it} \quad i = 1, \dots, N; t = 1, \dots, T$$

Bu modelde aşağıdaki varsayımlarda bulunmaktadır:

1. Sıfır koşullu ortalama:  $E(U_{it} | X_{i1}, \dots, X_{iT}) = 0$ .

Bu ifadedeki E sembolü, bir rassal değişkenin beklenen değerini temsil eder ve hata terimlerinin (U) bağımsız değişkenlerin (X) değerleriyle herhangi bir ilişki taşımadığını belirtir. Başka bir deyişle, belirli bağımsız değişken değerleri altında hata terimi ortalamasının sıfır olduğunu ifade eder. Bu nokta önemlidir, çünkü bu varsayım, regresyon modelinin tahmin gücünü ve geçerliliğini etkileyen bir faktördür. Eğer bu şart sağlanmazsa, modelin tahminleri yanıltıcı olabilir ve istatistiksel sonuçlar güvenilir olmayabilir.

2. Varlıklar arasındaki gözlemler aynı türdendir. Ortak dağılımlarından yararlanır.
3. Büyük aykırı değerlerin olması muhtemel değildir.
4. Mükemmel bir doğrusallık yoktur.

Birden fazla bağımsız değişken olduğunda,  $X_{it}$ ;  $X_{1,it}$ ,  $X_{2,it}$ , ...,  $X_{k,it}$  ile değiştirilir (Tilburg Science Hub, 2023).

Sabit etki modeli bağlamında, analizde yer alan tüm çalışmaların temelinde yatan tek bir gerçek etki büyüklüğünün var olduğu varsayılmakta ve gözlemlenen etki farklılıklarının tamamının örneklem hatasından kaynaklandığı düşünülmektedir. Bu modele "sabit etki modeli" adının konulmasına rağmen, daha açıklayıcı bir terim

olarak "ortak etki modeli" terimini kullanmak da mümkündür (Borenstein vd. 2010, 97-111).

Sabit etki modelini kullanmak için genelde iki göstergeye bakılmaktadır. Tüm çalışmaların işlevsel olarak aynı olduğuna dair sağlam bir neden bulunmalı ve ortak etki büyüklüğünü hesaplanmak istenmelidir. Analize dâhil edilen popülasyonun dışına genelleştirilmemelidir (Borenstein vd. 2010, 97-111).

#### 4.1.c. Rastgele Etkiler Modeli

Model aşağıdaki gibi temsil edilebilir:

- $Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + \epsilon_{it} + \mu_{it}$
- Panel verinin genel formasyonunda açıklandığı gibi denklemden  $Y_{it}$  modele ait bağımlı değişkenleri,  $X_{it}$  bağımsız değişkenleri,  $\beta_0$  başlangıç eğim parametresini göstermektedir. Hata terimi olarak;  $\epsilon_{it}$  terimi bilinen hataları ifade ederken,  $\mu_{it}$  terimi birim hatayı, yani zamanla değişmeyen birimler arasındaki farkı göstermektedir (Tatoğlu 2016, 103).

Rastgele etkiler modeli çerçevesinde, gerçek etki büyüklüklerinin çeşitlilik göstermesine izin verilmektedir. Tüm çalışmaların aynı etki büyüklüğünü paylaştığı bir senaryo varsayılabilir, ancak etki büyüklüklerinin değişebildiği senaryoların da mümkün olduğu ve çalışılması gerektiği gözlemlenmektedir. Örneğin, katılımcıların diğer çalışmalara göre daha yaşlı, daha eğitilmiş veya daha yüksek gelirli olduğu durumlarda, ya da müdahalenin daha yoğun bir varyantının kullanıldığı durumlarda etki büyüklüğü daha yüksek (veya daha düşük) olabilir. Çalışmalar, katılımcı karışımları ve müdahale uygulamalarındaki farklılıklar gibi sebeplerle farklı etki büyüklüklerine sahip olabilirler. Eğer inceleme için belirlenen dâhil etme kriterlerine dayanarak sınırsız sayıda çalışma yapmak mümkün olsaydı, bu

çalışmaların gerçek etki büyüklüklerinin belirli bir ortalama etrafında dağıldığı gözlemlenirdi. Rastgele etki modelinde, gerçekleştirilen çalışmalardaki etki büyüklükleri, bu etki büyüklüklerinin belirli bir dağılımından rastgele bir örnekleme temsil etmektedir (Borenstein vd. 2010, 97-111).

Birçok araştırma yönteminde, sabit etki varsayımı gerçekçi değildir. Bir grup çalışmayı meta-analize dâhil etme durumu olduğu zaman, bunların bilgiyi sentezlemek açısından yeterince ortak özellikler içerdiği düşünülse bile aslında çoğu çalışmada gerçek etki büyüklüğünün birebir aynı olduğunu varsaymak için bir temel bulunmamaktadır (Borenstein vd. 2010, 97-111).

#### 4.1.d. Hausman Testi

Panel veri analizinde farklı tahmin yöntemleri arasında seçim yapma konusunda rehberlik eden istatistiksel testlerden bir tanesi de Hausman testidir. Bu test, genellikle sabit etkiler ve rastgele etkiler modelleri arasında hangi tahmin yönteminin tercih edileceğini belirlemek amacıyla kullanılır. Hausman testinin temel mantığı, sabit etkiler modeli (ki birey veya birim özgü sabit etkiler içerir) ile rastgele etkiler modeli (ki birey veya birim özgü rastgele etkiler içerir) arasında yapılan iki farklı tahminin karşılaştırılmasına dayanır. Eğer bu iki tahmin arasında istatistiksel olarak anlamlı bir fark varsa, Hausman Testi, hangi modelin kullanılacağına dair bilgi sağlar (Nargeleçekenler 2009, 48-51). Hausman testi, şu hipotezleri test eder (Dağlı ve Ayaydın 2012, 45-65):

- H<sub>0</sub>: Açıklayıcı değişkenler ve birim (zaman) etki arasında ilişki yoktur.
- H<sub>1</sub>: Açıklayıcı değişkenler ve birim (zaman) etki ilişkilidir.

Yapılan test sonucunda temel hipotez olan H<sub>0</sub> hipotezi kabul edilirse sabit ve rastgele etkilerin tahmincileri arasında fark çok ufaktır. Yani rastgele etkiler modelini

kullanılması daha doğru bir seçenektir.  $H_0$  hipotezinin reddi yani diğer  $H_1$ 'in kabul edilmesiyle ise, rastgele etkiler modelinin tahminçileri ciddi farklılık sergilemektedir, bu yüzden sabit etkiler modelinin kullanılması daha doğru bir seçenektir (Tatoğlu 2016, 185).

#### 4.1.e. Birim Kök Testi

Birim kök testi, bir zaman serisinin durağanlık özelliğini test etmek amacıyla kullanılan istatistiksel bir testtir. Zaman serisi, durağan (stationary) ya da durağan olmayan (non-stationary) şeklinde kategorize edilmektedir. Durağanlık, bir zaman serisinin istatistiksel özelliklerinin zaman içinde sabit kaldığı anlamına gelir. Birim kök testi, genellikle Augmented Dickey-Fuller (ADF) testi veya Kwiatkowski-Phillips-Schmidt-Shin (KPSS) testi gibi test istatistikleri kullanılarak gerçekleştirilir (Phillips ve Xiao 1998, 423-470).

Bu testler, bir zaman serisinin birim kök hipotezini test eder. Birim kök hipotezi, bir zaman serisi birim köke sahiptir ve dolayısıyla durağan değildir şeklindedir. Bunun karşı hipotezi ise bir zaman serisi birim köke sahip değildir ve dolayısıyla durağandır şeklindedir (Nargeleçekenler 2009, 86-97). Eğer birim kök hipotezi reddedilirse, zaman serisinin durağan olmadığı kabul edilir. Bu durumda, serinin durağan hale getirilmesi için fark alma, doğal logaritmik fonksiyon alma veya farklı değişimler yapılmaktadır.

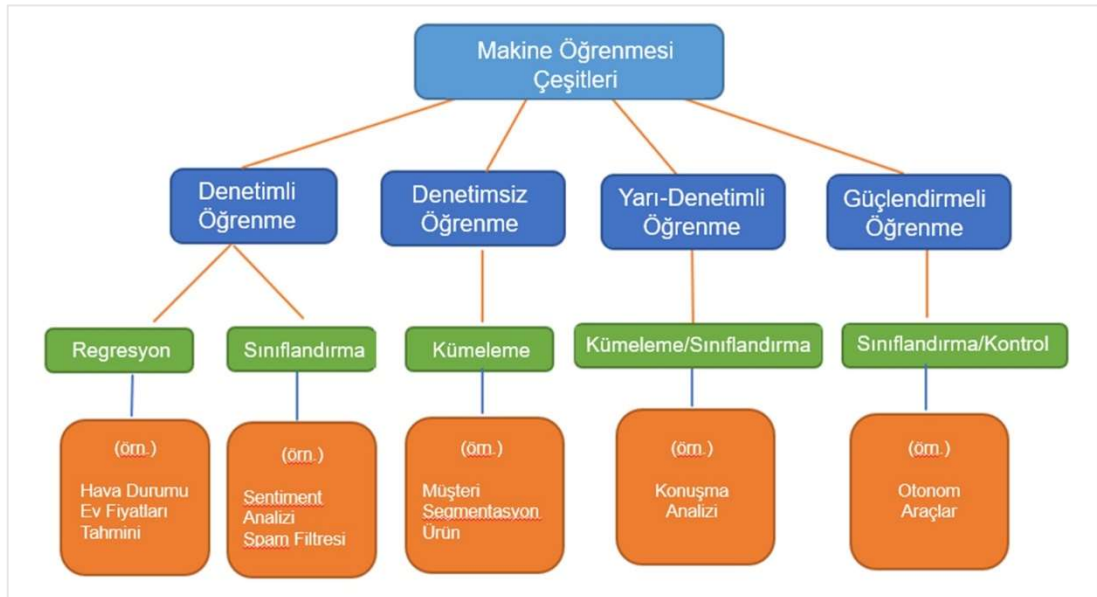
#### **4.2. Makine Öğrenmesi (ML)**

ML modelleri verileri anlamlandırmak için her geçen gün daha fazla kullanılmaya başlanmıştır. Bu tezde ise doğum oranının yıllar içerisindeki değişimi ile ilgili, etiketli sınıflandırma modellerinden yararlanılacaktır. ML, temelde bilgisayarların

öğrenmelerini sağlayan algoritmalar ve tekniklerle ilgilenmektedir. ML yaklaşımı, veri madenciliği, zor programlanabilir uygulamalar ve yazılım uygulamaları gibi ana alanları kapsamaktadır. Çeşitli algoritmaların total koleksiyonu denilebilmekte ve çok değişkenli, nonlinear, nonparametrik regresyon veya sınıflandırma sağlayabilen yöntemleri içermektedir (El Bouchefry ve de Souza 2020, 225-249).

ML, bilgisayarların verilerden öğrenerek görevleri bağımsız olarak gerçekleştirmesine olanak tanıyan yapay zekânın bir alt alanıdır. Esasen ML, verilerden desen çıkarmayı otomatikleştiren bir süreci ifade eder. ML'nin amacı, makinelerin deneyimden öğrenme ve problemleri çözme yeteneğini kazanmasıdır. Bunun temel nedeni, adım adım bir modeli izleyen bir algoritmanın her problemi modelleyip çözmemesidir. Örneğin, kedi ve köpekleri resimlerinden ayırt etmek, bir makine için karmaşık bir görev olabilirken insanlar için basittir. Bu tür durumlar için standart algoritmaların uygulanamaması nedeniyle önceki örneklerden ve deneyimlerden öğrenerek bu tür problemleri çözmek ve performansı artırmak için ML yöntemleri kullanılmaktadır (Mitchell 1997).

Şekil. 4.1. Makine Öğrenmesi Çeşitleri (Piyush 2020)



ML modelleri, verilerden bir model veya kural öğrenerek çalışır. Bu model veya kural, yeni bilgileri tahmin etmek veya sınıflandırmak için kullanılabilir. Şekil 4.1'de görüldüğü üzere ML modelleri, verinin niteliğine ve istenilen sonuca bağlı olarak dört ana kategoriye ayrılmaktadır:

- **Denetimli Öğrenme:** Bu modeller etiketli veriler kullanılarak eğitilir. Etiketli veriler, her bir bilginin doğru mu yanlış mı, yüksek mi yoksa düşük mü olduğunu vb. öğrenerek göstermektedir. Denetimli ML modelleri sınıflandırma, regresyon, tahmin ve optimizasyon gibi çeşitli görevleri gerçekleştirmek için kullanılabilir. Sınıflandırma modelleri bir nesnenin hangi kategoriye ait olduğunu belirlerken, regresyon modelleri ise sürekliliği olan çıktılar tahmin etmektedir.
- **Denetimsiz Öğrenme:** Bu modeller etiketli veriler üzerinde eğitilmezler. Etiketli olmayan verilerdeki gizli kalıpları ve ilişkileri keşfetmek için kullanılır. Kümeleme, aykırı değer tespiti ve boyutluluk azaltma gibi çeşitli görevleri gerçekleştirmek için kullanılabilir (Michie, Spiegelhalter ve Taylor,1994).
- **Yarı denetimli Öğrenme:** Hem etiketli hem de etiketsiz veriler üzerinde eğitilmiştir. Yarı denetimli ML modelleri, denetimli ML ve denetimsiz ML modellerinin avantajlarını birleştirmektedir.
- **Güçlendirilmeli Öğrenme:** ML modelleri ödül veya ceza sistemi ile eğitilebilir. Bu gelişmiş ML modelleri, karmaşık ortamlardaki etkinlikleri planlamak ve yürütmek için kullanılmaktadır.

ML modelleri çeşitli endüstrilerde ve uygulamalarda kullanılmaktadır. Örneğin, ML modelleri bankacılıkta kredi riskini değerlendirmek, sağlık koşullarını teşhis etmek, perakende satışta müşteri davranışını tahmin etmek ve endüstrideki üretim hatlarını optimize etmek için kullanılabilir. Son zamanlarda ML teknikleri;

astronomi, jeobilim ve uzaktan algılama gibi alanlarda dahi birçok uygulama bulmuştur. ML'nin geniş uygulanabilirliğini sağlayan faktörlerden biri, karmaşık ve gerçek dünya koşullarında çok iyi performans gösterebilen çeşitli ML algoritmalarının bulunmasıdır (El Bouchefry ve de Souza 2020, 225-249).

ML modelleri için çeşitli gruplamalar yapılmaktadır. Regresyon modelleri, verilerdeki bir değişken ile başka bir değişken arasındaki ilişkiyi tahmin etmek için kullanılır. Örneğin, bir öğrencinin final notunun test puanlarına ve devam durumuna göre nasıl değişeceğini tahmin etmek için bir regresyon modeli kullanılmaktadır. Kümeleme modelleri, verileri benzer özelliklere sahip gruplara veya kümelere ayırmak için kullanılır. Her küme yapısı için belirlenmiş bir uygunluk ölçütü mevcuttur. Daha yüksek bir uygunluk değeri, daha etkili bir küme yapısını ifade eder. Örneğin, müşterileri demografik özelliklere veya satın alma davranışlarına göre gruplara ayırmak için bir grup modeli kullanılmaktadır. Öneri motorları, kullanıcılara ilgilerini çekebilecek ürün veya hizmetleri önermek için kullanılır. Örneğin bir çevrimiçi alışveriş sitesinde, kullanıcının önceki satın alma geçmişine ve arama geçmişine dayalı olarak ürünler önermek için bir öneri motoru kullanılmaktadır. Sınıflandırma modelleri, verileri belirli sınıflara veya kategorilere ayırmak için kullanılır. Basitçe ifade edildiğinde; fotoğraftaki nesnenin araba mı, köpek mi yoksa kedi mi veya iki ihtimalli olasılıklarda doğru mu yanlış mı şeklinde bir sınıflandırma yapabilmek için kullanılmaktadır.

Bu çalışmada doğum oranının bir önceki yıla göre artış/azalışının sınıflandırılması sebebiyle literatürde sıkça kullanılan etiketli sınıflandırma ML modellerinden faydalanılmıştır. Aşağıda bu modellerden bahsedilecektir.

#### 4.2.a Lojistik Regresyon (LR)

Lojistik regresyon, kesikli deęişken arasındaki ilişki nedeniyle kullanılmaktadır. Regresyonun sayısal veri deęerlerinde daha iyi sonuç verdięi bilinmekle birlikte, sürekli ve kesikli tahmincilerin karışık deęerleri ile kesikli deęişkenlerin tahminine izin verir. Lojistik regresyonun, lineer regresyon ile arasındaki en büyük farkı iki sınıfı birbirinden ayıracak çizgiyi nasıl uyguladıęıdır. Lineer regresyon, en uygun çizgiyi çizmek için “En Küçük Kareler Yöntemi” (Least Squares) kullanırken, lojistik regresyon “Maksimum Olabilirlik” (Maximum Likelihood) kullanır. LR’ de sigmoid fonksiyonu kullanılır. Temelde, verilerimizi 0 ile 1 arasına sıkıştırmak için kullanılan bir fonksiyondur. Bu fonksiyon sayesinde sınıflandırma işlemi gerçekleştirebiliriz (Akca 2021).

#### 4.2.b. K En Yakın Komşular (KNN)

KNN, sürekli ya da kesikli hedef fonksiyonlarını tahmin etmek için kullanılan bir örnek tabanlı algoritmadır ve bu algortmada örneklerin n-boyutlu öklidyen uzayındaki noktalara karşılık geldięi varsayılmaktadır. Yeni bir sorgu için hedef fonksiyon deęeri, k, en yakın eğitim örneğinin bilinen deęerlerinden tahmin edilir. Yerel ağırlıklı regresyon yöntemleri, KNN’nin genelleştirilmiş bir versiyonudur ve her sorgu örneęi için hedef fonksiyona özgü bir yerel yaklaşım oluşturulur. Bu yerel yaklaşım, sabit, lineer veya karesel fonksiyonlar veya uzamsal olarak yerelleştirilmiş çekirdek fonksiyonları gibi çeşitli fonksiyonel formlara dayanabilir. (Mitchell 1997, 258). KNN, bir veri kümesinde yeni bir veriyi sınıflandırmak için, eğitim verilerindeki en yakın k komşularına bakar. Komşuların mesafesi, mesafe ölçüsü kullanılarak belirlenir. KNN eğitim verilerininin mesafe ölçüsü kullanılarak sınıflandırılır. Yeni bir

veri, eğitim verilerindeki en yakın  $k$  komşusuna göre sınıflandırılır. Euclidean mesafe, minkowski mesafe, mahalanobis mesafe gibi çeşitli mesafe ölçüleri de kullanılabilir.

#### 4.2.c. Çok Katmanlı Algılayıcı (MLP)

MLP, yapay sinir ağları kullanan bir ML yöntemidir. MLP'ler, giriş katmanı, gizli katmanlar ve çıkış katmanı olmak üzere birden fazla katmandan oluşur. Her katmandaki her bir nöron, bir dizi ağırlığa sahiptir. MLP denetimli öğrenme problemlerini çözmek için etkili ve güçlü bir algoritmadır. Yapay sinir hücreleri, araştırma alanında birçok ilginç ve kullanışlı özelliğe sahip biyolojik sinir hücrelerinin basitleştirilmiş bir versiyonunu temsil eder (El Bouchefry ve de Souza 2020, 225-249).

MLP'ler, eğitim verilerinden ağırlıkları öğrenerek çalışmaktadır. MLP ise eğitim verileri, giriş katmanına verildikten sonra giriş katmanı, nöronların ağırlıklarını kullanarak gizli katmana sinyaller gönderir. Gizli katman da nöronların ağırlıklarını kullanarak çıkış katmanına sinyaller gönderir. Çıkış katmanı, sinyalleri kullanarak sınıflandırma sonuçlarını üretir. MLP'ler, karmaşık kalıpları öğrenmek için etkili olabilmektedir. Ancak, eğitim ve tahmin için gereken hesaplama maliyeti yüksektir.

#### 4.2.d. Destek Vektörü (SV)

SV için genellikle sınıflandırma problemlerinde uygulanan bir gözetimli öğrenme metodu denilebilir. Noktaları bir düzlem üzerinde ayırmak için bir doğru çizer ve bu doğrunun iki sınıfın noktaları arasında maksimum uzaklıkta olmasını amaçlamaktadır. Bu metot, karmaşık ancak küçük ve orta seviyeli veri setleri için daha uygundur. Sınıflandırma problemlerindeki esas hedef eldeki verinin hangi sınıfta yer alacağını belirleyebilmektir. SV'de sınıflandırmayı yapabilmek için verileri iki sınıfı ayıran bir

dođru çizilir ve bu dođrunun  $\pm 1$ 'i arasında bulunan kısma kenarlık (margin) adı verilir. Kenarlık genişledikçe sınıfa ayırma işlemi daha iyi yapılabilmektedir (Akca 2020).

#### 4.2.e. Karar Ağaçları (DT)

DT'ler, verileri sınıflandırmak için kullanılan bir başka basit ML yöntemidir. Karar ağaçları, bir dizi karar kuralı kullanarak verileri sınıflandırır. Veriler, karar ağaçlarında her yaprağın belirli bir kavrama işaret ettiği ve bu kavramın olasılıksal bir açıklamasını içeren hiyerarşik bir ağaçla gösterilir. Etiketlenmemiş verileri temsil etmek için çeşitli algoritmalar, sınıflandırma ağaçları oluşturur (Rokach ve Maimon 2005). Karar ağacı metodu eğitim verileri, en önemli özelliğın seçilmesi için kullanılır. En önemli özellik seçildikten sonra, özellik değerine göre veriler iki alt kümeye ayrılır. Bu işlem, alt kümelerin boyutu yeterli olana kadar devam eder. Her alt küme için, bir sonraki en önemli özellik seçilir ve işlem tekrarlanır. Karar ağaçları, anlaşılması ve yorumlanması kolaydır. Ancak, veri kümesi küçük veya karmaşıkça aşırı uyuma eğilimindedir.

#### 4.2.f. Rastgele Orman (RF)

RF, Breiman (2001) tarafından geliştirilen bir topluluk sınıflandırma yöntemidir ve birden fazla DT'nin veriyi bölmek amacıyla çoğunluk oyu kullanarak sınıfları tahmin etmesini sağlar. RF, birden çok ağaç büyütürken, her ağacın düğümünde rastgele belirlenen bir sayıda değişkeni seçer (Breiman 2001). Her bir DT'yi, eğitim verilerinden rastgele seçilen bir alt küme üzerinde eğitir. Bu, her bir DT'nin aşırı uyum olasılığını azaltır. RF ise eğitim verileri, birden fazla alt kümeyle rastgele bölünür. Her bir alt küme, bir DT eğitmek için kullanılır. Sonuç olarak, birden fazla DT elde edilir. Yeni bir veri, bu DT'ler tarafından sınıflandırılır. RF, DT'lere göre daha dođru sonuçlar verebilmektedir. Çünkü rastgele bölünen alt kümeler karar ağaçlarını eğitmek için kullanıldığından sonuçlar daha optimize olacaktır.

#### 4.2.g. Gradyan Artırma (GB)

GB, birden fazla karar ağacı kullanan bir başka ML yöntemidir. GB makineleri, çeşitli pratik uygulamalarda önemli başarılar elde etmiş güçlü bir ML tekniği ailesidir. Bu makineler, uygulamanın özel gereksinimlerine yüksek derecede adapte edilebilmektedir (Natekin ve Knoll 2013). GB, her bir DT'yi, önceki ağaçlardan gelen hatalara göre eğitir. Bu, her bir DT'nin önceki ağaçların hatalarını düzeltmeye odaklanmasını sağlar. GB, eğitim verileri, bir DT eğitmek için kullanılır. DT'nin hataları hesaplanır. İkinci bir DT, önceki ağacın hatalarını düzeltmek için eğitilir. Bu işlem, gerekli doğruluk elde edilene kadar devam eder. GB, DT'lere göre daha doğru sonuçlar verebilmektedir. Çünkü önce DT'ler oluşturulup daha sonra bu DT'lerin hataları düzeltildiği için daha verimli bir DT mekanizması oluşturulmaktadır.

#### 4.2.h. Aşırı Gradyan Artırma (XGB)

XGB, sınıflandırma problemleri için geliştirilmiş olan yeni bir topluluk öğrenme yöntemidir. XGB, GB algoritmasının optimize edilmiş bir sürümüdür. Başka bir ifade ile verimli bir şekilde dış bellek öğrenimi ve seyreklik farkındalığı sunan bir gradyan ağaç artırma algoritmasıdır. Dış bellek öğrenimi, tek bir bilgisayarın belleğine sığmayan ancak yerel bir sabit diske benzer bazı veri depolama alanlarına rahatça sığabilen verileri işleyen bir dizi algoritma anlamına gelir. XGB'nin çalışma prensibi, GB algoritmasının çalışma prensibi ile aynıdır. Ancak, XGB, daha verimli bir ağaç öğrenme algoritması kullanır ve GB'yi daha düzenli şekle dönüştürmek için bir dizi teknik kullanmaktadır. Bu, XGB'yi daha doğru ve ölçeklenebilir bir yöntem haline getirmektedir. Bu sebeple, XGB, talep tahmini sorunu gibi konular için oldukça uygun bir denetimli öğrenme algoritması olarak öne çıkmaktadır (Chen ve Guestrin 2016).



## BÖLÜM V

### BULGULAR

Bu bölümde; doğum oranını etkileyen faktörler belirlenirken literatürde yapılan çalışmalar ve veri setinin anlamlandırılabilir olma durumu göz önünde bulundurulmaktadır. Tüm bu gözlemler neticesinde Türkiye'nin illere ve yıllara göre belirlenmiş doğum oranlarının; lisans ve üzeri eğitim oranı, kadın istihdam oranı, nüfustaki kadın oranı, gayri safi yurt içi hasıla (GSYH) seviyesi, evli nüfus oranı, göçmen ve covid19 kukla değişkenleri ile ilişkisi panel veri analizi ve ML modelleriyle incelenecektir.

#### 5.1. Panel Veri Analizi

Çalışmada kullanılan analizlerde, veri seti panel veri olması sebebiyle bu tarz verilere uygun modeller seçilerek ilerlenmektedir. Modeller kurulduktan sonra en iyisinin seçilmesi ve gerekli testlerin yapılması, araştırmacının doğru ve tutarlı sonuçlar elde etmesini sağlamaktadır. Bu nedenle aşağıda aynı panel veri seti ile analiz edilen; havuzlanmış OLS modeli, rastgele etkiler modeli, sabit etkiler modeli, hausman testi ve birim kök testine ilişkin sonuçlar yer almaktadır.

##### 5.1.a. Havuzlanmış OLS

Havuzlanmış model, tüm birimlerin aynı olduğunu varsaydığı için en basit modeldir. Bu modelde, birim etkileri (sabit etkiler ve zaman etkileri) modele dahil edilmemektedir. Havuzlanmış modelin en büyük avantajı, tahminlerinin kolay yorumlanmasıdır. Ancak, bu modelin bazı önemli dezavantajları da vardır: Birim

etkileri modele dahil edilmediği için, açıklayıcı değişkenler ile birim etkileri arasında korelasyon varsa, bu korelasyon hatalı tahminlere yol açabilmektedir.

Linear regression		Number of obs	=	1,053	
		F(7, 1045)	=	181.35	
		Prob > F	=	0.0000	
		R-squared	=	0.6074	
		Root MSE	=	.1989	
ln_dogum_orani	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
evli_oran	.6398028	.0828426	7.72	0.000	.4772459 .8023596
kadin_oran	-5.284955	1.136019	-4.65	0.000	-7.514094 -3.055816
ln_GSYH	-.2877113	.0249237	-11.54	0.000	-.3366174 -.2388051
k_istihdam	-.0145496	.0018779	-7.75	0.000	-.0182344 -.0108648
lisans_ve_üstü	-.0034781	.0003231	-10.77	0.000	-.004112 -.0028441
1.gocmen	.1732848	.0180663	9.59	0.000	.1378344 .2087352
1.covid19	-.1385267	.0234787	-5.90	0.000	-.1845974 -.0924559
_cons	8.347011	.5450599	15.31	0.000	7.277475 9.416548

**Tablo 5.1.** Havuzlanmış OLS Modeli Uygulama Sonuçları

İlk model olarak Tablo 5.1’de sonuçları ifade edilen Havuzlanmış OLS modeli uygulanmıştır. Modeldeki değişkenlerin doğum oranı değişimini istatistiksel olarak açıklayabilme gücü yaklaşık %60 olarak bulunmuştur. Değişkenlerin hepsi  $\alpha=0,05$  düzeyinde istatistiksel olarak anlamlı çıkmıştır. Bu tez kapsamında anlamlılık düzeyi sürekli 0,05 olarak kabul edilmiştir. Evli oranı ve göçmen kukla değişkenine bakıldığında doğum oranı ile pozitif korelasyona sahiptir. Lisans ve üstü eğitim oranı, GSYH seviyesi, kadın istihdam oranı, kadın oranı ve covid19 kukla değişkenlerinde ise doğum oranı ile negatif bir korelasyon gözükmemektedir. Genel olarak anlamlı bir analiz olmasının yanında nüfustaki kadın oranının beklenenin ve bilinenin tam aksine doğum oranı ile negatif korelasyona sahip olduğu gözlemlenmektedir.

### 5.1.b. Rastgele Etkiler Modeli

Rastgele etkiler modeli, birim etkilerinin sabit etkilerden bağımsız olduğunu varsaydığı için havuzlanmış modelden daha esnektir. Bu modelde, birim etkileri

modele dahil edilir ve açıklayıcı değişkenler ile birim etkileri arasında korelasyon var olsa bile, bu korelasyon hatalı tahminlere yol açmaz. Rastgele etkiler modelinin en büyük avantajı, havuzlanmış modele göre daha etkin tahminler üretmesidir. Ancak, bu modelin bazı önemli dezavantajları da vardır: Birim etkilerinin sabit etkilerden bağımsız olduğunu varsayması, gerçekte her zaman doğru olmamaktadır.

Random-effects GLS regression		Number of obs = 1,053				
Group variable: plaka		Number of groups = 81				
R-sq:		Obs per group:				
within = 0.8168		min = 13				
between = 0.1352		avg = 13.0				
overall = 0.1700		max = 13				
corr(u_i, x) = 0 (assumed)		Wald chi2(7) = 3851.87				
		Prob > chi2 = 0.0000				
-----						
ln_dogum_orani	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
evli_oran	.1263701	.0417115	3.03	0.002	.044617	.2081232
kadın_oran	2.807061	.4135493	6.79	0.000	1.996519	3.617603
ln_GSYH	.070217	.0131076	5.36	0.000	.0445265	.0959074
k_istihdam	-.0029626	.0006809	-4.35	0.000	-.0042971	-.0016281
lisans_ve_üstü	-.0020041	.0001044	-19.20	0.000	-.0022087	-.0017994
1.gocmen	-.0157273	.0087056	-1.81	0.071	-.0327901	.0013354
1.covid19	-.1210125	.0065298	-18.53	0.000	-.1338106	-.1082144
_cons	.9196671	.2327121	3.95	0.000	.4635599	1.375774
-----						
sigma_u	.17761834					
sigma_e	.04674758					
rho	.93521785	(fraction of variance due to u_i)				

**Tablo 5.2.** Rastgele Etkiler Modeli Uygulama Sonuçları

Tablo 5.2’de rastgele etkiler modeline ait sonuçlar yer almaktadır. Panel veri ile çalışmak modele bazı farklılıklar getirmiştir. Modeldeki değişkenlerin doğum oranı değişimini istatistiksel olarak açıklayabilme gücü yaklaşık %81 olarak bulunmuştur. Göçmen kukla değişkeni hariç değişkenlerin hepsi  $\alpha=0,05$  düzeyinde istatistiksel olarak anlamlı çıkmıştır. Evli oranı, kadın oranı ve GSYH seviyesi değişkenlerine bakıldığında doğum oranı ile pozitif korelasyona sahiptir. Lisans ve üstü eğitim oranı, kadın istihdam oranı ve covid19 kukla değişkenlerinde ise doğum oranı ile negatif bir korelasyon gözükmemektedir. Havuzlanmış OLS yapıldığında negatif çıkan kadın oranı

katsayısı rastgele etkiler modelinde beklenildiği gibi pozitif olmuştur. Sadece rastgele etkiler modeline bakılarak bir sonuca varmak mümkün değildir. Panel veri analizlerinde hem rastgele hem de sabit etkiler modeli uygulanarak hausman testi sonucuyla hangisinin daha doğru olduğuna karar verilmektedir.

### 5.1.c. Sabit Etkiler Modeli

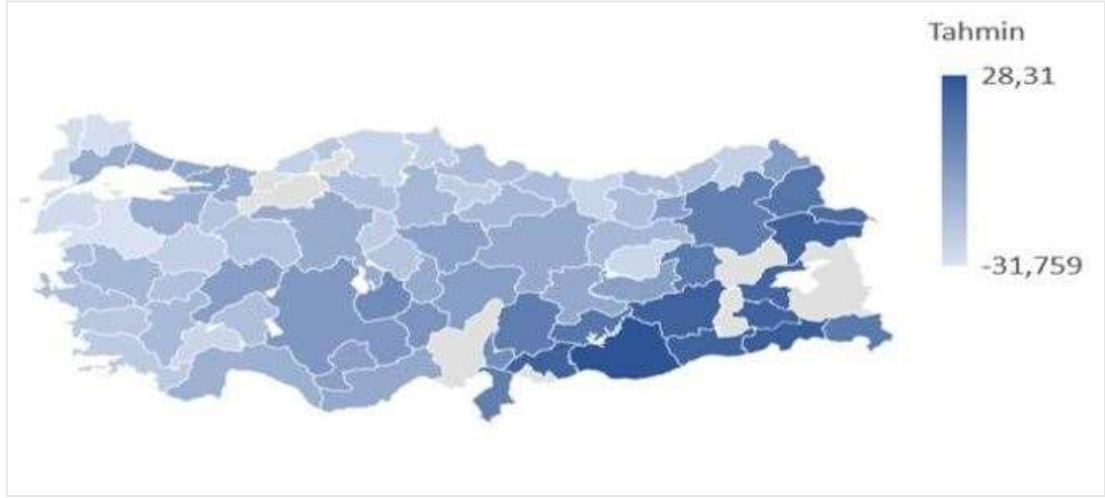
Sabit etkiler modeli, birim etkilerinin sabit olduğunu varsaydığı için rastgele etkiler modelinden daha katı bir modeldir. Bu modelde, birim etkileri modele dahil edilir ve açıklayıcı değişkenler ile birim etkileri arasında korelasyon varsa, bu korelasyon tahminlere dahil edilir. Sabit etkiler modelinin en büyük avantajı, birim etkilerinin gerçekte sabit olduğu durumlarda daha tutarlı tahminler üretmesidir. Ancak, bu modelin bazı dezavantajları da vardır. Çünkü birim etkilerinin sabit olduğunu varsayması, her zaman doğru olmamaktadır.

Fixed-effects (within) regression		Number of obs	=	1,053		
Group variable: plaka		Number of groups	=	81		
R-sq:		Obs per group:				
within	= 0.8180	min	=	13		
between	= 0.0239	avg	=	13.0		
overall	= 0.1110	max	=	13		
corr(u_i, Xb) = 0.0079		F(7,965)	=	619.71		
		Prob > F	=	0.0000		
-----						
ln_dogum_orani		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
evli_oran		.128201	.0395865	3.24	0.001	.0505155 .2058864
kadın_oran		3.158797	.3957854	7.98	0.000	2.382098 3.935496
ln_GSYH		.0960497	.0126177	7.61	0.000	.0712884 .120811
k_istihdam		-.0025467	.0006437	-3.96	0.000	-.0038098 -.0012835
lisans_ve_üstü		-.0019694	.0000987	-19.96	0.000	-.002163 -.0017758
1.gocmen		-.0204642	.0082196	-2.49	0.013	-.0365946 -.0043338
1.covid19		-.1182118	.0061597	-19.19	0.000	-.1302997 -.1061238
_cons		.5009307	.2226061	2.25	0.025	.0640829 .9377786
-----						
sigma_u		.29664794				
sigma_e		.04674758				
rho		.97576839	(fraction of variance due to u_i)			
-----						
F test that all u_i=0: F(80, 965) = 224.40				Prob > F = 0.0000		

**Tablo 5.3.** Sabit Etkiler Modeli Uygulama Sonuçları

Sabit etkiler modeli sonuçları Tablo 5.3'te yer almaktadır. Modeldeki deęişkenlerin doğum oranı deęişimini istatistiksel olarak açıklayabilme gücü yaklaşık %82 olarak bulunmuştur. Deęişkenlerin hepsi  $\alpha=0,05$  düzeyinde istatistiksel olarak anlamlı çıkmıştır. Evli oranı, kadın oranı ve GSYH seviyesi deęişkenlerine bakıldığında doğum oranı ile pozitif korelasyona sahiptir. Lisans ve üstü eğitim oranı, kadın istihdam oranı, göçmen ve covid19 kukla deęişkenlerinde ise doğum oranı ile negatif bir korelasyon gözükmemektedir. Sabit etkiler modelinde iki deęişken için tezin başında düşünölenin ve beklenilenin aksine sonuçlar elde edilmiştir. Göçmen kukla deęişkeni için nüfusunun %5'i üzerinde geçici koruma kapsamında göç alan iller için doğum oranını artıracakğı yönündeki beklenti sabit etkiler modeline göre tam tersi azaltıcı bir etkiye sahip çıkmaktadır. Diğer deęişken ise GSYH seviyesidir. Literatürde genel olarak GSYH yükseldikçe doğum oranında bir azalma beklenmektedir (Rayhan vd., 2018). Ancak hep tek düze sonuçlar yoktur. Bazı çalışmalar pozitif veya negatif korelasyon tespit edemezken özellikle gelişmekte olan ölkeler için yapılan bazı çalışmalarda ise GSYH seviyesi yükseldikçe doğum oranını azaldığı şeklinde çıkarımlar gözlemlenmektedir (Koppes 2022). Türkiye için yapılan bu analizde de boş vakit ve maaş arasındaki ilişki gibi GSYH'nin de doğum oranına pozitif etki yapacakğı düşünölmesi halinde bu durum sıkıntı teşkil etmemektedir. Türkiye'de refah seviyesindeki artış, çocuk bakımı ile ilgili insanların yaşadığı maddi sorunların çözümö için yardımcı olup doğum oranında yükselmeye sebep olduğı şeklinde bir yorum yapılabileceğı görölmektedir.

**Şekil 5.1.** Sabit Etkiler Modeli Sonuçlarıyla Elde Edilen İl Katsayıların Haritalandırılmış Gösterimi



Şekil 5.1’de illere göre sabit etkiler modeli kurgulanıp sonuçlarına göre il il elde edilen katsayıların haritalandırılmış gösterimi yer almaktadır. Haritada gri olarak kalan iller (Adana hariç) excelden kaynaklı il tanınmadığı için öyle kalmıştır. Adana ise sabit etkideki referans şehir olduğu için verisi gözükmemektedir. Adana’ya kıyasla etkilere bakıldığında Doğu ve Güney Doğu Anadolu bölgelerinde görece oranlar yükselirken batı kıyıları ve Batı Marmara’da doğum oranları hayli düşük gözükmemektedir.

Sabit etkiler ve rastgele etkiler modellerinin çıktılarını incelendikten sonra aşağıda hangi modele karar verileceği ile ilgili Hausman test sonucuna bakmak faydalı olacaktır.

#### 5.1.d. Hausman Testi

	---- Coefficients ----			
	(b) fe	(B) re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
evli_oran	.128201	.1263701	.0018309	.0054706
kadın_oran	3.158797	2.807061	.3517358	.0767088
ln_GSYH	.0960497	.070217	.0258327	.0028265
k_istihdam	-.0025467	-.0029626	.0004159	.0000655
lisans_ve_~ü	-.0019694	-.0020041	.0000346	9.66e-06
1.gocmen	-.0204642	-.0157273	-.0047369	.0007161
1.covid19	-.1182118	-.1210125	.0028008	.0004598

b = consistent under Ho and Ha; obtained from xtreg  
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(6) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)  
= 125.00  
Prob>chi2 = 0.0000  
(V\_b-V\_B is not positive definite)

Tablo 5.4. Hausman Testi Uygulama Sonuçları

Tablo 5.4'te görüldüğü üzere yapılan hausman test sonucunun olasılık değeri 0 çıkmaktadır. Yani H0 hipotezi reddedilmekte ve katsayılar arasında sistematik bir farklılık olduğu gözlemlenmektedir. Test sonucuna göre sabit etkiler modelini tercih etmenin daha doğru olduğu anlaşılmaktadır.

#### 5.1.e. Birim Kök Testleri

Panel verilerle ekonometrik analizler için bir diğer önemli test olan birim kök testleri de tüm değişkenler için yapılmıştır. Birim kök testleri model değişkenlerinin durağan olup olmadığını göstermekle beraber modellerin durağan yapıdaki değişkenlerle kurulmasına yardımcı olan bir testtir.

Birim Kök Testleri		
Ho: Panels contain unit roots	Number of panels = 81	
Ha: Panels are stationary	Number of periods = 13	
	Adjusted t*	p-value
ln_dogum_orani	-7.3072	0.000
evli_oran	-8.6871	0.000
kadın oran	-22.7957	0.000
lisans_ve_üstü	-10.1237	0.000
ln_GSYH	-4.9197	0.000
k_istihdam	-28.6111	0.000

**Tablo 5.5.** Birim Kök Testleri Uygulama Sonuçları

Tablo 5.5'te de görüldüğü gibi elde edilen birim kök test sonuçlarına göre tüm değişkenler için H0 hipotezi reddedilmiştir. Yani panel hiçbir değişken için birim köke sahip değildir, hepsi durağandır sonucu çıkmaktadır.

## 5.2. Makine Öğrenmesi Uygulaması

Bu bölümde, doğum oranındaki artışı 1, azalışı ise 0 olacak şekilde sınıflandırılıp doğum oranı, iki sonuçlu bir bağımlı değişkene dönüştürülmektedir. Doğum oranındaki değişim, panel veri analizlerinde kullanılan bağımsız değişkenlerle eğitilip sonuçlar incelenecektir. Önce hangi ML etiketli sınıflandırma modelinin bu konuda daha iyi sonuçlar verdiğine karar verilecek, daha sonra modelde doğum oranındaki değişim ile ilgili değişkenlerin önem düzeyleri analiz edilecektir.

### 5.2.a. ML İçin Model Seçimi

ML modeli seçimi, bir projenin başarısı açısından önem arz etmektedir. Veri setinin özellikleri, proje hedefleri, veri boyutu, hesaplama gücü, genelleme yeteneği ve açıklanabilirlik gibi çeşitli faktörler, doğru modelin belirlenmesini etkilemektedir. Veri setine uygun, hedeflere uygun, hesaplama gücü ve zaman açısından uygun,

genelleme yeteneđi yüksek ve gerektiđinde açıklanabilir modellerin seçilmesi kritiktir. Yanlış tercihler, modelin performansını olumsuz etkileyebilir ve projenin başarısını tehlikeye atabilmektedir. Bu nedenle, model seçimi, ML projelerinde özenle ele alınması gereken kilit bir aşamadır. ML modellerinden en iyi sonucu veren modeli seçmek için aşağıdaki faktörler göz önünde bulundurulmalıdır:

**Dođruluk:** Doğruluk, modelin doğru olarak sınıflandırma yaptığı veri noktalarının yüzdesidir. Yani doğru pozitifler(TP) ve doğru negatiflerin(TN) toplamının; doğru pozitif, doğru negatif, yanlış pozitif(FP) ve yanlış negatifleri(FN) toplamına oranıdır.

$$(TP+TN) / (TP+FP+TN+FN)$$

Özellikle dağılımı eşit olmayan verilerde tek başına yeterli değildir.

**Duyarlılık:** Duyarlılık, modelin pozitif olarak sınıflandırdığı veri noktalarının yüzdesidir. Yani pozitif olarak tahmin etmemiz gerekenleri ne kadar edebildiğimizdir.

$$TP / (TP+FN)$$

Örneđin, bir hastalık teşhis modelinde; duyarlılık, modelin hasta olan kişileri doğru olarak teşhis ettiği yüzdesidir.

**Kesinlik:** Kesinlik, modelin tahminde pozitif olarak elde edilen değerlerin aslında ne kadarının pozitif olduğunu göstermektedir.

$$TP / (TP+FP)$$

Kesinlik değerinin yüksek olması model için önemlidir.

**Özellik:** Özellik, modelin negatif olarak sınıflandırdığı veri noktalarının yüzdesidir. Örneđin; bir hastalık teşhis modelinde, modelin sağlıklı olan kişileri doğru olarak teşhis ettiği yüzdesine özellik denilmektedir.

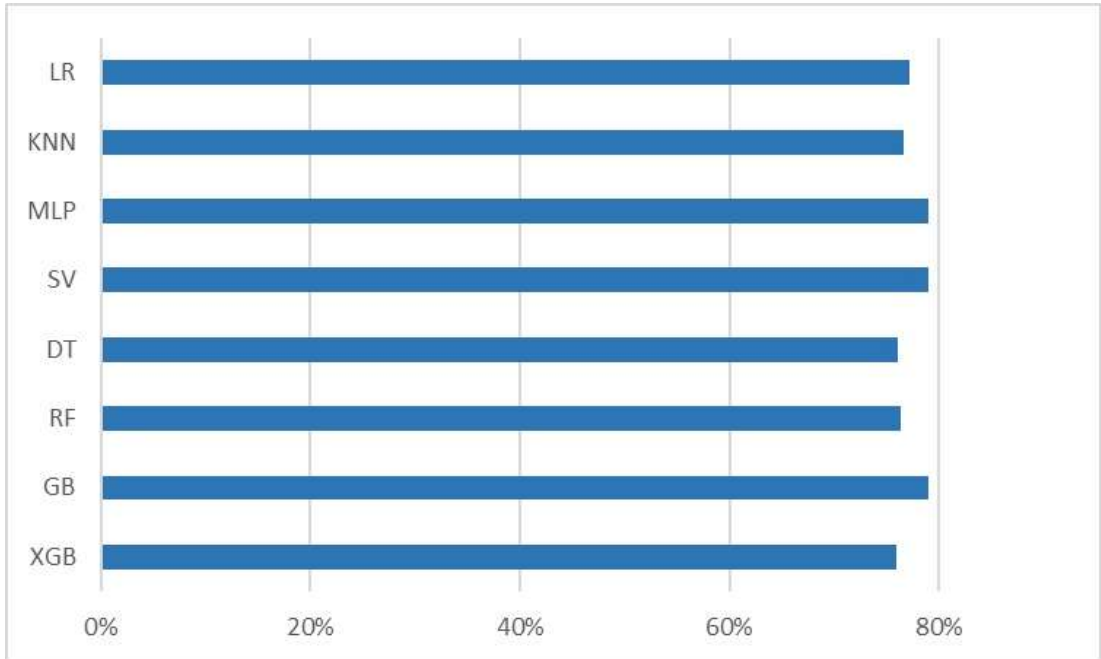
**ROC eğrisi altındaki alan (AUC):** AUC, modelin doğruluğunu ve hassasiyetini bir araya getiren bir ölçüdür. Genel olarak, AUC değeri en yüksek olan model tercih edilmektedir.

**F1 skor:** F1 skor, doğruluk ve duyarlılığı dengeleyen bir ölçüdür. Genel olarak, F1 skor değeri en yüksek olan model en iyisi kabul edilmektedir.

$$2 * (\text{Duyarlılık} * \text{Kesinlik}) / (\text{Duyarlılık} + \text{Kesinlik})$$

Bu faktörleri göz önünde bulundurarak, modellerin performansını karşılaştırabilir ve en iyi sonucu veren seçilebilmektedir. Genellikle modellerin doğruluk oranları ve ROC eğrisi altında kalan alan değerleri kıyaslanarak hangisinin seçileceğine karar verilebilmektedir. Bunun yanında modelin uygulama alanıyla ilgili özel gereksinimlerinizi de göz önünde bulundurmak gerekmektedir. Örneğin, bir güvenlik uygulamasında kullanılan bir model için, yanlış pozitiflerin maliyeti yüksek olabilir. Bu durumda, duyarlılığı daha yüksek olan bir model daha iyi bir seçim olacaktır.

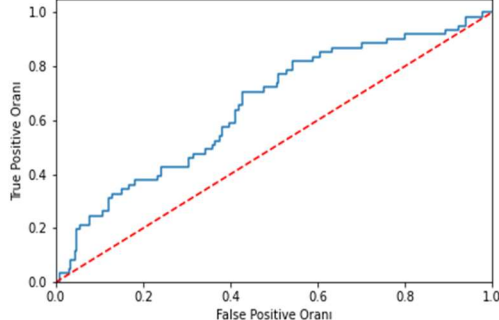
**Şekil 5.2.** Modellere Ait Doğruluk Oranları



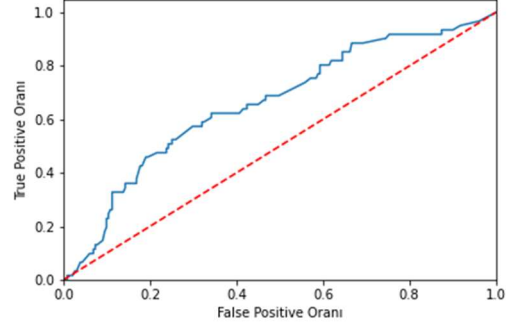
Modellere ait doğruluk oranları Şekil 5.2’de gösterilmektedir. Bu grafiğe göre oranlar birbirine hayli yakın gözükmemektedir. Doğruluk oranları bu kadar yakınken model tercihi yapmak riskli olacaktır. Veri setimiz ve sınıflandırma hipotezimizle en

anlamli ve guclu sekilde calisabilen modeli secibilmek adina ROC egrisi altinda kalan alan degerlerine bakilmasi gerekmektedir.

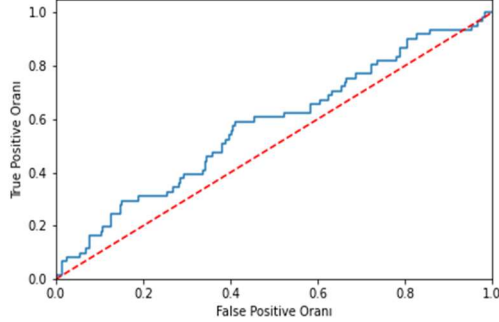
**Ŗekil 5.3.** LR, KNN, MLP, SV, DT, RF, GB ve XGB'nin ROC Egrisi Altında Kalan Alan Grafikleri



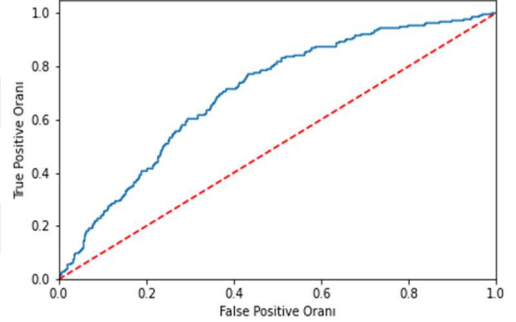
(a) LR-ROC



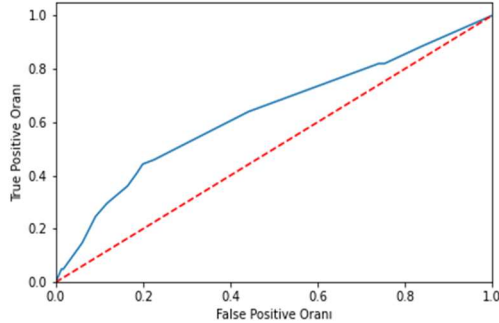
(b) KNN-ROC



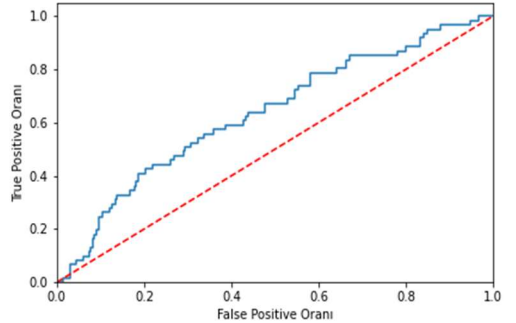
(c) MLP-ROC



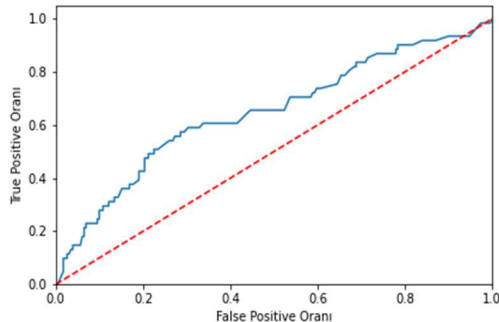
(d) SV-ROC



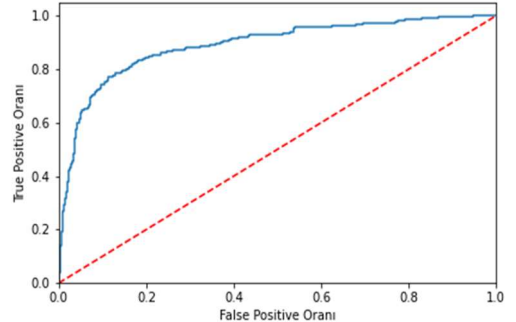
(e) DT-ROC



(f) RF-ROC



(g) GB-ROC



(h) XGB-ROC

ROC grafikleri, modelin doğruluk ve hassasiyet verilerini bir arada gösteren en güzel grafiklerden birisidir. Şekil 5.3'te sekiz farklı modele ait ROC eğrisi altında kalan alan grafikleri yer almaktadır. Sınıflandırma problemleri için çok önemli performans ölçütlerinden biri olan ROC grafiklerine bakıldığında LR, KNN ve MLP'nin AUC değerleri çok düşük gelmektedir. SVC, DT, RF ve GB'nin AUC değerleri dikkate alınacak seviyede olmasa da ortalama bir trend sergilemektedir. AUC değeri en yüksek olan model ise XGB modelidir. Bu sebeple model seçimi için sadece doğruluk kıyaslamalarına bakılarak karar verilemezken ilave olarak AUC değerlerine bakıldığı zaman XGB modelinin en iyi performansa sahip olduğu kolaylıkla gözlemlenmektedir.

XGB modelinin performansının yeterli seviyede olup olmadığını anlamlandırmak için XGB karar matrisi oluşturup F-1 skor değerine bakılması uygun olacaktır. Bu değer incelenmesi, XGB modelinin performansının yeterli seviyede olup olmadığını değerlendirmek için önemlidir. Yüksek bir F-1 skoru, modelin hem doğruluğunu hem de yanlış pozitif ve negatif tahminler arasındaki dengeyi başarılı bir şekilde sağladığını gösterebilir.

Şekil 5.4. XGB Karar Matrisi

	0	1
Gerçek Değerler 0	209	22
1	48	13
Tahminler		

Şekil 5.4'te XGB modeline ait karar matrisi yer almaktadır. Karar matrisi tablosu sırasıyla 0-0; gerçekte azalan tahminde azalan, 0-1; gerçekte azalan tahminde artan, 1-0; gerçekte artan tahminde azalan ve 1-1; gerçekte artan tahminde artan değerleri gösterir. Bu matrise göre azalışları çok iyi tahmin eden model, artışları o kadar iyi tahmin edememiştir. Ancak bu durum veri setimizden kaynaklıdır. Çünkü tabloya bakıldığında azalan örneklem artan örnekleme göre çok daha fazla veri içermektedir. Bu da modelin azalanları iyi öğrendiği ancak verideki dengesizlikten dolayı artanları çok iyi öğrenemediğini göstermektedir.

Çalışmada yapılan ML modelinde azalış ve artış hesaplandığı için doğru ya da yanlış diye ayırım yapmak yanlış olacaktır. Bu sebeple duyarlılık ve kesinliği hesaplarken verinin çok olduğu kısımdan yani azalışlar üzerinden yapmak daha doğru olacaktır. Yani 0'lar doğru ve pozitifleri, 1'ler yanlış ve negatifleri temsil edecektir. Bu şekilde bakıldığında, XGB modeli için F1 skoru şu şekildedir:

$$\text{Kesinlik} = TP / (TP+FP) = 209 / (209+48) = 0,813$$

$$\text{Duyarlılık} = TP / (TP+FN) = (209) / (209+13) = 0,941$$

$$\begin{aligned} \text{F1 Skoru} &= 2 * \text{Duyarlılık} * \text{Kesinlik} / (\text{Duyarlılık} + \text{Doğruluk}) \\ &= 2 * 0,941 * 0,813 / (0,941 + 0,813) = 0,872 \end{aligned}$$

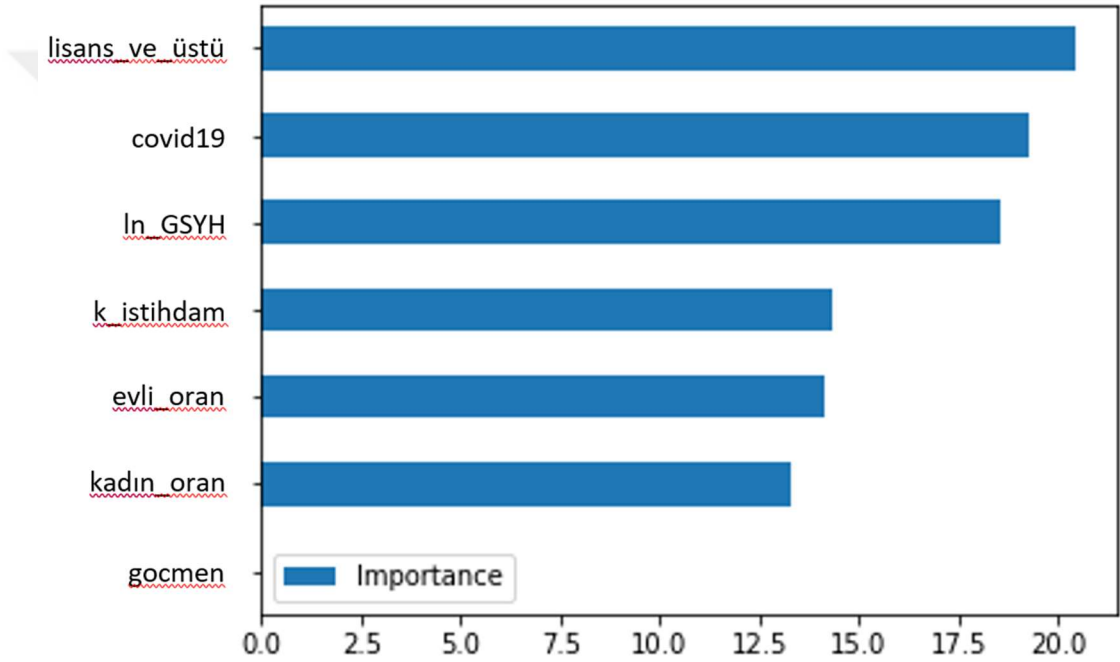
Böylelikle yukarıda F1 skoru hesaplamasına göre azalışları tahmin etme konusunda XGB modeli %87 gibi bir başarıya sahiptir. Tüm bu gözlemler neticesinde en güçlü ve anlamlı performans veren model XGB modeli seçilmektedir.

### 5.2.b. XGB Modelinde Değişkenlerin Önem Düzeyi

Değişken önem düzeyi tablosu, bir modelin iç mekanizmasını anlamak için önemli bir araçtır. Hangi değişkenlerin model tahminlerini ne ölçüde etkilediğini göstererek,

modelin yorumlanabilirliğini artırmaktadır. Ayrıca modelin önemli değişkenlerinin belirlenmesi, gereksiz değişkenlerinin tespit edilmesi ve modelin genel güvenilirliğinin anlaşılması hususlarında destek sağlamaktadır. Bu sayede, modelin kararlarını ve tahminlerini daha iyi kavramamıza yardımcı olmaktadır. En başarılı model olarak seçilen XGB modelinin değişken önem düzeyi uygulama sonuçlarını incelemek ve buna göre yorum yapmak gerekmektedir.

Şekil 5.5. XGB Modeli Değişken Önem Düzeyi Uygulama Sonuçları



Şekil 5.5'te XGB modeline ait değişken önem düzeyi uygulama sonuçları yer almaktadır. Göçmen kukla değişkeni için önem düzeyi sıfır gelmiştir yani model için bu değişkenin bir anlam ifade etmediği söylenebilmektedir. Kadın istihdam oranı, evli oranı ve kadın nüfusunun oranınının 13-14 seviyelerinde benzer önem düzeylerine sahip oldukları görülmektedir. Modeli açıklamak için en önemli değişkenler olmasalar da bu model için yadsınamaz seviyede öneme sahip oldukları söylenebilmektedir. Modeldeki en önemli değişken lisans ve üzeri eğitim oranı olurken GSYH seviyesi ve COVID-19 kukla değişkeni ardından gelen diğer önemli iki değişken olarak

gözükmektedir. COVID-19 kukla değişkeninin hesaplanması itibariyle 2020 ve 2021 yıllarını içermesi ve bu iki yılın da doğum oranı değişimi ile ilgili benzer sonuçlar getirmesi, bu şekilde yüksek önem düzeyinde gelmesinin başlıca nedenidir. Bu sebeple bu modeli açıklamada en önemli iki değişkenin lisans ve üzeri eğitim oranı ve kişi başına düşen GSYH seviyesi olduğunun ifade edilmesi yanlış olmayacaktır.

Özetle; Türkiye’de bir sonraki yılın doğum oranı değişimini artış ve azalış etiketiyle sınıflandırıp mevcut veri setini kullanarak ML sınıflandırma modelleriyle incelendiğinde XGB modeli en iyi performans veren modeldir. XGB modeli için göçmen kukla değişkeni hariç bütün değişkenler anlamlıdır. Modelin en kritik değişkeni ise lisans ve üzeri eğitim oranıdır.



## BÖLÜM VI

### SONUÇ

Doğum oranı, demografik analizler ve nüfus çalışmalarında önemli bir ölçüt olarak hizmet vermektedir. Bir ülkenin veya bölgenin doğurganlık seviyesini anlamak, nüfusun değişimini ve büyümesini değerlendirmek açısından kritik bir faktördür. Doğum oranındaki değişiklikler, ekonomik, sosyal ve kültürel faktörlerin yanı sıra nüfus politikalarının etkilerini incelemek için kullanılmaktadır.

Bu çalışmanın amacı, geçmişte yüksek doğum oranlarına sahip olan Türkiye'nin doğum oranındaki düşüş başta olmak üzere değişimlerle ilgili etkili olan faktörler hakkında yorum yapabilmektir. Bu nedenle doğum oranını etkileyen faktörleri belirlerken literatürde yapılan çalışmalar ve veri setinin anlamlandırılabilir olma durumu göz önünde bulundurulmuştur. Bunun neticesinde, il ve yıl bazında doğum oranının; lisans ve üzeri eğitim oranı, kadın istihdam oranı, nüfustaki kadın oranı, gayri safi yurt içi hasıla (GSYH) seviyesi, evli nüfus oranı, covid19 ve göçmen kukla değişkenleri ile ilişkisi, panel veri analizi ve ML modelleriyle incelenmiştir.

Panel veri analizi incelemelerini yapmak için 3 farklı model (Havuzlanmış OLS, rastgele etkiler ve sabit etkiler) kurulmuştur. Bu modellerin sonuçlarına bakıldığında  $R^2$  değeri, sabit etkiler ve rastgele etkiler modellerinde havuzlanmış OLS'ye göre daha yüksek çıkmıştır. Ayrıca veri detayında il ve yıl bilgileri mevcutken sabit ve rastgele etkiler modellerini kurmak için panel veri seti imkânlarından faydalanmak daha doğru olacaktır. Sabit ve rastgele etkiler modellerine bakıldığında iki modelde de göçmen kukla değişkeni hariç bütün değişkenler istatistiki olarak anlamlı çıkmış ve bu değişkenlerin etki yönlerinin de aynı olduğu görülmüştür.

Çalışma öncesi hipotezde şehirlere göre etkilerin farklılaşacağına dair sahip olunan düşünceyi test etmek için Hausman test sonucundan faydalanılmıştır. Hausman test sonucuna göre sabit etkiler modelini tercih etmenin daha doğru olduğu görülmektedir. Sabit etkiler modeliyle yapılan panel veri analizi sonucuna göre hem model hem de bütün değişkenler %5 önem düzeyinde istatistiksel olarak anlamlı çıkmıştır.

Değişkenlerin ifade ettikleri incelendiğinde, lisans ve üzeri eğitimin nüfusun doğum oranını azalttığı görülmektedir. Eğitim oranının lisans düzeyi ve üzerine çıkmasıyla birlikte kadınların evlilik yaşının ilerlemesi, kişisel kariyere daha çok önem verilmesi, doğum kontrol konusunda daha bilgili olunması vb. birçok sebepten dolayı böyle bir sonuca ulaşmak beklenen ve olağan bir durumdur. Kadın istihdam oranının artması da doğum oranını düşürmektedir. Çalışma hayatında aktif rol oynayan bir kadın için çocuk sayısının artması sorumluluk, çocuk bakıcı maliyeti, kreş maliyeti, ulaşım imkânlarının sağlanması, işinden olma tehlikesi vb. yeni düşüncelere sebep olmaktadır. Tüm bunları birlikte değerlendiren anne-baba adayları geçmişe nazaran daha az çocuk sahibi olmak istemeleri beklenen negatif ilişkinin açıklayıcısı olabilmektedir.

Kadın ve evli nüfusun artması doğum oranını artırmaktadır sonucu hayatın doğal akışında beklediğimiz bir netice iken doğum oranının kişi başına düşen GSYH seviyesi ile pozitif ilişki içerisinde çıkması genel literatürden farklı gözükmektedir. Ancak bu değişkeni Türkiye örneği için gelişmiş ülkelere ait sonuçlarla kıyaslamak doğru değildir. Bu durum hayat pahalılığı vb. sebeplerden insan davranışlarının Türkiye nezdinde değişmesi ile açıklanabilmektedir. Çocuk bakım masraflarının her geçen gün yükselmesi ve GSYH'nin yıllara göre ortalamasının giderek azalması çocuk sahibi olmanın maddi yönlerine de odaklanmayı gerektirdiği sonucunun çıkarılabilmesini sağlamaktadır.

Kukla deęişkenlerden COVID-19 kukla deęişkeninde alıřmanın bařlangıcında beklenildięi gibi doęum oranını azaltıcı bir etkisinin olduęu gözlemlenmektedir. Özellikle dünya geneli bir salgın olmasının ciddi etkisiyle insanların saęlık ve gelecek kaygılarının en üst seviyede olduęu bu süreçte çocuk düşünmedikleri gözlemlenmektedir. Göçmen kukla deęişkeninde ise beklenin aksine doęum oranını azaltıcı bir etki göstermektedir. Rastgele etkiler modelinde anlamsız, havuzlanmış OLS modelinde ise pozitif korelasyona sahip gelen göçmen kukla deęişkeninin sabit etkiler modelinde ıkan sonucunun tam olarak anlamlı olduęu söylenememektedir. Sabit etkiler modelinde doęum oranı ile negatif iliřkili ıkmasının temel sebebi; göçmenlerin aęırlıklı olarak yařadıkları illerde, hâlihazırdaki doęum oranlarının zaten yüksek olmasından kaynaklanmaktadır. Göçmenlerin yüzde 5'in üzerinde olduęu illerde yařayan nüfustan daha yüksek doęum oranına ulaşamıyor oldukları ıkarımı yapılabilmektedir. Örneęin doęum oranı en düşük illerden biri olan anakkale'de göçmen oranı hiçbir yılda yüzde 5'in üzerine ıkamamıřtır. Doęum oranının en yüksek olduęu řehirlerden biri olan řanlıurfa'da ise her yıl yüzde 5'in üzerine ıkmıřtır.

ML incelemelerinde doęum oranındaki yıllara göre deęişim, artıř için 1 azalıř için 0 olacak řekilde etiketli sınıflandırma problemine dönüřtürülmüřtür. Literatürde sıkça geen sınıflandırma modellerinin hepsi uygulanmıřtır. alıřmadaki veri seti ile daha doęru sonuçlar üretebilen, performansı en yüksek ML sınıflandırma modeli XGB modeli ıkmaktadır. XGB modelinde doęum oranındaki deęişim için hangi deęişkenlerin belirleyici olduęuna bakılmıřtır. Panel veri analizinde de pek anlamlı ıkmayan göçmen kukla deęişkeninin herhangi bir etkisi gözükmemektedir. Dięer tüm deęişkenlerin önem düzeyi mevcuttur. Lisans ve üzeri mezuniyet, doęum oranındaki deęişimi açıklamak için en önemli gösterge gözükmemektedir. Kadın istihdam oranı, evli oranı ve kadın nüfusu oranının ok yüksek olmamakla beraber benzer önem

düzeylelerine sahip oldukları görülmektedir. Modeldeki en önemli deęişken lisans ve üzeri eğitim oranı olurken GSYH seviyesi ve COVID-19 kukla deęişkeni ardından gelen dięer önemli iki deęişken olarak gözükmektedir. Eğitim durumunun ve kişi başına düşen GSYH seviyesinin toplumların doğum oranındaki deęişimlerde ciddi etkisi olduęu çıkarımı yapılabilmektedir.

Birebir aynı durumu modellemeseler de panel veri analizi sonuçlarıyla ML sonuçlarını kıyasladığımızda, her ikisi içinde göçmen kukla deęişkeni harici deęişkenler anlamlıdır. ML ile yapılan incelemelerde doğum oranı deęişimlerinin tahmininde en etkili deęişkenlerinin lisans ve üzeri eğitim, COVID-19 ve kişi başına düşen GSYH seviyesi olduęu görülmekte iken; panel veri analizi tarafında negatif yönlü en yüksek etkiye sahip deęişkenler COVID-19 ve lisans ve üzeri eğitim, pozitif yönlü en yüksek etkiye sahip deęişkenler ise kadın nüfusu ve kişi başına düşen GSYH seviyesi gözükmektedir.

Sonuç olarak kullanılan modellerin ikisi de doğum oranını belirleyen faktörler için etkin bir şekilde çalışabildiğini göstermektedir. Türkiye’de doğum oranı uzunca süredir düşüş eğilimindedir. Bu durum göz önünde bulundurarak doğum oranındaki deęişim üzerine yapılacak politikalarda lisans ve üzeri eğitim, kadın istihdam oranı, kişi başına düşen GSYH seviyesi çok ciddi önem arz etmektedir. Doğum oranlarındaki düşüş eğiliminin deęiştirilmesi hedeflendięi takdirde; özellikle lisans ve üzeri eğitimin ve kadın istihdamının insanlar için çocuk istememe sebebi olmasını azaltacak politikalar geliştirilmelidir. Ayrıca COVID-19 gibi salgınların doğum oranlarına etkisi göz ardı edilmemelidir.

ML modelleri, demografik verilerdeki mikro detaylar arttıkça ve daha dengeli veri setleri elde edildiğinde başarısını artıracaktır. Bu tarz veri setleri ile doğrudan ilgili

tahminde bulunma görevleri yüklenilmesi ve bu şekilde Türkiye’de gelecek yıllarda doğum oranlarının ne olacağına yönelik tahminler yapılması ilerideki çalışmalarda hedeflenmektedir. Veri setinde GSYH seviyesinin yıllara göre oransal değişimini kullanarak analizler yapılması ve ML modelleri için şehirlerin de kukla değişken olarak önem düzeylerinin tespit edilmesi yine ilerideki çalışmalarda hedeflenmektedir.





## KAYNAKÇA

Akadlı-Ergöçmen, Banu. 1997. "An overview of women's status and fertility in Turkey." *Women and Families: Evolution of the Status of Women as a Factor and Consequence of Changes in Family Dynamics*, Maria Eugenia COSIO-ZAVALA (ed), Paris, CICRED: 79-105.

Akca, Mehmet Fatih. 2020. "Nedir Bu Destek Vektör Makineleri? (Makine Öğrenmesi Serisi-2)" Erişim Tarihi: 6 Aralık 2023. <https://medium.com/deep-learning-turkiye/nedir-bu-destek-vekt%C3%B6r-makineleri-makine-%C3%B6%C4%9Frenmesi-serisi-2-94e576e4223e>.

Akca, Mehmet Fatih. 2021. "Lojistik Regresyon Nedir? Nasıl Çalışır" Medium. Erişim tarihi: 6 Aralık 2023. <https://mfakca.medium.com/lojistik-regresyon-nedir-nas%C4%B1l-%C3%A7al%C4%B1%C5%9F%C4%B1r-4e1d2951c5c1>.

Akça, Haşim ve Mehmet Ela. 2012. "Türkiye’de Eğitim, Doğurganlık ve İşsizlik İlişkisinin Analizi." *Maliye Dergisi* 163: 223-242.

Akin, Mustafa Seref. 2005. "Education and Fertility: A Panel Data Analysis for Middle Eastern Countries." *The Journal of Developing Areas* 39(1): 55-69.

Ali, Fatma Romeh M. ve Shiferaw Gurmu. 2018. "The Impact of Female Education on Fertility: A Natural Experiment From Egypt." *Review of Economics of the Household* 16: 681-712.

Al-Jarrah, Omar Y., Paul D. Yoo, Sami Muhaidat, George K. Karagiannidis ve Kamal Taha. 2015. "Efficient Machine Learning for Big Data: A Review." *Big Data Research* 2(3): 87-93.

- Ashraf, Quamrul H., David N. Weil ve Joshua Wilde. 2013. "The Effect of Fertility Reduction on Economic Growth." *Population and Development Review* 39(1): 97-130.
- Aydın, Olgu, Pınar Aslantaş Bostan ve Ertuğrul Murat Özgür. 2018. "Mekânsal veri analizi teknikleriyle Türkiye’de toplam doğurganlık hızının dağılımı ve modellenmesi." *Coğrafya Dergisi* (37): 27-45.
- Basu, Alaka Malwade. 2002. "Why does education lead to lower fertility? A critical review of some of the possibilities." *World Development* 30(10): 1779-1790.
- Becker, Sascha O., Francesco Cinnirella ve Ludger Woessmann. 2010. "The Trade-Off Between Fertility and Education: Evidence From Before The Demographic Transition." *Journal of Economic Growth* 15(3): 177-204.
- Behrman, Julia ve Pilar Gonalons-Pons. 2020. "Women’s employment and fertility in a global perspective (1960–2015)." *Demographic Research* 43: 707.
- Bernhardt, Eva M. 1993. "Fertility and Employment." *European Sociological Review* 9(1): 25-42.
- Borenstein, Michael, Larry V Hedges, Julian PT Higgins ve Hannah R Rothstein. 2010. "A basic introduction to fixed-effect and random-effects models for meta-analysis." *Research synthesis methods* 1(2): 97-111.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45: 5-32.
- Brownlee, Jason. 2016. *Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End*. Machine Learning Mastery.
- Chen, Tianqi ve Carlos Guestrin. 2016. "Xgboost: A scalable tree boosting system." KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining San Francisco, CA, 13-17 Ağustos.

Cornett, A. 2020. "Analyzing the Relationship Between Female Education and Fertility Rate." *Journal of Drake Undergraduate Social Science*.

D'Addio, Anna Cristina ve Marco Mira d'Ercole. 2005. "Policies, institutions and fertility rates: a panel data analysis for OECD countries." *OECD Economic Studies* 2005(2): 7-45.

Da Rocha, José María ve Luisa Fuster. 2006. "Why Are Fertility Rates and Female Employment Ratios Positively Correlated across O.E.C.D. Countries?" *International Economic Review* 47(4): 1187-1222.

Dağlı, Hüseyin ve Hasan Ayaydın. 2012. "Gelişen piyasalarda hisse senedi getirisini etkileyen makroekonomik değişkenler üzerine bir inceleme: Panel veri analizi." *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi* 26(3-4): 45-65.

Dielman, Terry E. 1983. "Pooled cross-sectional and time series data: A survey of current statistical methodology." *The American Statistician* 37(2): 111-122.

Dünya Bankası. 2023. "Fertility rate, total (births per woman)" Erişim tarihi: 5 Aralık 2023.

[https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?end=2021&name\\_desc=false&start=1960&view=chart](https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?end=2021&name_desc=false&start=1960&view=chart).

El Bouchefry, Khadija ve Rafael S de Souza. 2020. "Learning in big data: Introduction to machine learning." *Knowledge discovery in big data from astronomy and earth observation içinde*, 225-249. Elsevier.

Emara, Noha. 2016. "Fertility and Female Employment: A Panel Study on Developing Countries." *Applied Economics and Finance* 3.

Eryurt, Mehmet Ali ve Banu Akadli-Ergöçmen. 2008. "Ebeveyn Eğitiminin Doğurganlık Üzerindeki Etkisi." *Nüfusbilim Dergisi* 30(1): 13-28.

Gallagher, James. "Türkiye'de yaşanan son gelişmeler." BBC. 15 Temmuz 2020.

<https://www.bbc.com/turkce/haberler-dunya-53417604?xtor=AL-73-%5Bpartner%5D-%5Bt24.com.tr%5D-%5Blink%5D-%5Bturkish%5D-%5Bbizdev%5D-%5Bisapi%5D>.

Galor, Oded ve Hyoungsoo Zang. 1997. "Fertility, income distribution, and economic growth: Theory and cross-country evidence." *Japan and the World Economy* 9(2): 197-229.

Herzer, Dierk, Holger Strulik ve Sebastian Vollmer. 2012. "The long-run determinants of fertility: one century of demographic change 1900–1999." *Journal of Economic Growth* 17(4): 357-385.

Hsiao, Cheng. 2005. "Why panel data?" *The Singapore Economic Review* 50(02): 143-154.

Iwasaki, Ichiro ve Kazuhiro Kumo. 2020. "Determinants of regional fertility in Russia: a dynamic panel data analysis." *Post-Communist Economies* 32(2): 176-214.

Jafari, Hasan, Ebrahim Jaafari-pooyan, Abouali Vedadhir, Abbas Foroushani, Bahman Ahadinejad ve Abolghasem Pourreza. 2016. "Socio-Economic Factors Influencing on Total Fertility Rate in Iran: A Panel Data Analysis for the Period of 2002-2012." *Electronic physician* 8: 2551-2556.

Karra, Mahesh, David Canning, and Joshua Wilde. 2017. "The Effect of Fertility Decline on Economic Growth in Africa: A Macrosimulation Model." *Population and Development Review* 43 (S1): 237–63.

Klasen, Stephan ve Andrey Launov. 2006. "Analysis of the determinants of fertility decline in the Czech Republic." *Journal of Population Economics* 19: 25-54.

Koppes, Derek Richard. 2022. "Influence of GDP per Capita on Birth Rates." Proceedings of the Jepson Undergraduate Conference on International Economics.

- Kögel, Tomas. 2004. "Did the Association between Fertility and Female Employment within OECD Countries Really Change Its Sign?" *Journal of Population Economics* 17(1): 45-65.
- L'Heureux, A., K. Grolinger, H. F. Elyamany ve M. A. M. Capretz. 2017. "Machine Learning With Big Data: Challenges and Approaches." *IEEE Access* 5: 7776-7797.
- Lal, Sumeet, Rup Singh, Keshmeer Makun, Nilesh Chand ve Mohsin Khan. 2021. "Socio-economic and demographic determinants of fertility in six selected Pacific Island Countries: An empirical study." *PLOS ONE* 16(9): e0257570.
- Meisenberg, Gerhard. 2008. "How Universal is the Negative Correlation between Education and Fertility?" *The Journal of social, political, and economic studies* 33: 205-227.
- Michie, Donald, David J Spiegelhalter ve Charles C Taylor. 1994. *Machine learning, neural and statistical classification*. Overseas Press.
- Mitchell, Tom M. 1997. *Machine Learning*. McGraw Hill.
- Nargeleçekenler, Mehmet. 2009. "Makroekonomik ve finansal serilerin ekonometrik analizi: Panel veri yaklaşımı." Doktora Tezi, Bursa Uludağ Üniversitesi.
- Natekin, Alexey ve Alois Knoll. 2013. "Gradient boosting machines, a tutorial." *Frontiers in neurorobotics* 7: 21.
- Nyoni, Smartson P. ve Thabani Nyoni. 2021. "Forecasting ART coverage in Malawi using the Multilayer perceptron neural network." *International Research Journal of Innovations in Engineering and Technology* 5(3): 222.
- Osili, Una Okonkwo ve Bridget Terry Long. 2008. "Does female schooling reduce fertility? Evidence from Nigeria." *Journal of Development Economics* 87(1): 57-75.
- Pesaran, M Hashem. 2015. *Time series and panel data econometrics*. Oxford University Press.

Phillips, Peter CB ve Zhijie Xiao. 1998. "A primer on unit root testing." *Journal of Economic Surveys* 12(5): 423-470.

Piyush. 2020. "Different Types of Machine Learning." Erişim tarihi: 6 Aralık 2023. <https://machinelearning8.blogspot.com/2020/05/different-types-of-machine-learning.html?zx=f14ca6c88a666769>.

Rahman Abdur, Akhtarul Islam, and Samia Yeasmin, vd. (2020). "Influencing Factors of Fertility in Developing Countries: Evidence from 16 DHS Data." *Journal of International Women's Studies* 21(6): 416-426.

Rayhan, Istihak, K Akter ve MS Islam. 2018. "Determinants of fertility rate decline in the south asian countries: A panel data approach." *International Journal of Development Research* 8(07): 21583-21589.

Rindfuss, Ronald R., S. Philip Morgan ve Kate Offutt. 1996. "Education and the Changing Age Pattern of American Fertility: 1963-1989." *Demography* 33(3): 277-290.

Rokach, Lior ve Oded Maimon. 2005. "Clustering methods." *Data mining and knowledge discovery handbook*: 321-352.

Rovny, Allison E. 2011. "Welfare state policy determinants of fertility level: A comparative analysis." *Journal of European Social Policy* 21(4): 335-347.

Shirahase, Sawako. 2000. "Women's increased higher education and the declining fertility rate in Japan." *Review of population and social policy* 9(2000): 47-63.

Teguh, Dartanto. 2009. "The determinants of fertility in Southeast and South Asian countries: An analysis of panel data." *LPEM FEUI Staff Paper* 6(2009): 1-18.

Tilburg Science Hub. 2023. "Regressions with Panel Data - Tilburg Science Hub." Erişim tarihi: 5 Aralık 2023. <https://tilburgsciencehub.com/building-blocks/analyze-data/regressions-paneldata/>.

Torres-Reyna, Oscar. 2007. "Panel data analysis fixed and random effects using Stata (v. 4.2)." *Data & Statistical Services, Princeton University* 112: 49

TÜİK. 2023. "Doğum İstatistikleri, 2022." Erişim tarihi: 5 Aralık 2023. <https://data.tuik.gov.tr/Bulten/Index?p=Dogum-Istatistikleri-2022-49673>.

Uyar, Gökhan. 2015. "Panel verilerle regresyon modelleri." Yüksek Lisans Tezi, Anadolu Üniversitesi.

Van den Broeck, Goedele ve Miet Maertens. 2015. "Female employment reduces fertility in rural Senegal." *PloS one* 10(3): e0122086.

Vollset, Stein Emil, Emily Goren, Can Yuan, Jackie Cao, Amanda E. Smith, Thomas Hsiao, Catherine Bisignano, vd. 2020. "Fertility, Mortality, Migration, and Population Scenarios for 195 Countries and Territories from 2017 to 2100: A Forecasting Analysis for the Global Burden of Disease Study." *The Lancet* 396 (10258): 1285–1306.

Yaffee, Robert. 2003. "A Primer for Panel Data Analysis A Primer for Panel Data Analysis." *Connect: 2-11*.

Yang, Hsiu-ling. 2000. "Education, married women's participation rate, fertility and economic growth." *Journal of Economic Development* 25(2): 101-118.

Yerdelen, Tatoğlu. 2016. *Panel Veri Ekonometrisi*. İstanbul: Beta Basım Yayım.

Yeşilkaya, Fatma. 2022. "Kadın İstihdamı ile Doğum Oranı Arasındaki İlişki Üzerine Eşbütünleşme Analizi: İsveç ve ABD Örneği." *Sosyoekonomi* 30(52): 333-348.

Yurtseven, Çağlar. 2015. "The Socioeconomic Determinants of Fertility Rates in Muslim Countries: A Dynamic Panel Data Analysis." *Economics & Sociology* 8: 165-178.

