

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
VERİ YÖNETİMİ VE ANALİZİ PROGRAMI
YÜKSEK LİSANS TEZİ

FARKLI FREKANSLI GOOGLE TRENDS VERİLERİ
İLE İŞSİZLİK ÖNGÖRÜSÜ

Dilan KAYA

Danışman

Doç. Dr. Serkan ARAS

İZMİR - 2024

TEZ ONAY SAYFASI



YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “*Farklı Frekanslı Google Trends Verileri ile İşsizlik Öngörüsü*” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Tarih

..../..../.....

Dilan KAYA

İmza

ÖZET
Yüksek Lisans Tezi
Farklı Frekanslı Google Trends Verileri İle
İşsizlik Öngörüsü
Dilan KAYA

Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü
Ekonometri Anabilim Dalı
Veri Yönetimi ve Analizi Programı

Ekonominin sağlığı açısından kritik bir gösterge olan işsizlik, Türkiye gibi gelişmekte olan ülkelerde sosyal dengeleri büyük ölçüde etkileyebilecek öneme sahiptir. Bu nedenle, iş arayan ancak iş bulamayan bireylerin yüzdesini ifade eden işsizlik oranının öngörülmesi, ekonomik planlamaların etkin bir biçimde gerçekleştirilmesi için değerlidir. Böylece, işsizlik problemini ortadan kaldırmak adına çeşitli politikalar yürütülebilir. Ancak işsizlik oranı öngörüsü gerçekleştirmek için birçok değişken dikkate alınmalıdır. Çünkü işsizlik sorunu enflasyon oranları, istihdam ve nüfus gibi makroekonomik ve demografik toplumsal göstergelerle ilişkilidir. Ayrıca, teknolojinin gelişimiyle işsizlik oranı öngörüsünde çeşitli çevrimiçi kaynakların kullanımı da ön plana çıkmaktadır. Google Trends gibi çevrimiçi kaynaklar, iş arayan bireylerin internet arama trendlerini, toplumdaki ilgi ve eğilimlerini yansıtarak işsizlik oranı öngörüsünde alternatif bir bakış açısı sunabilir. Bu nedenle bu internet tabanlı kaynakların, ilgilenilen değişkene ait öngörü performansını ne derece iyileştirebileceğini tespit etme ilgi konusu olmuştur.

Bu çalışmada, Google Trends üzerinden erişilen farklı frekanstaki veri kaynakları bir araya getirilmiş ve işsizlik oranı tahminine etkileri incelenmiştir. Makroekonomik değişkenler ve Google Trends üzerinden erişilen arama frekans verilerinin çeşitli kombinasyonları dokuz farklı makine öğrenmesi tekniğiyle analiz edilmiştir. Ayrıca belirlenen değişken kombinasyonları üzerinde, değişken seçim tekniği uygulanarak tüm modeller yeniden oluşturulmuştur. Böylece model

performansları daha geniş bir pencerede karşılaştırılmış ve işsizlik oranı tahmininde en iyi sonucu veren değişken grubunun belirlenmesi amaçlanmıştır. Sonuç olarak, en iyi performans 3 farklı frekans türünden elde edilen tüm değişkenlerin modele dahil edildiği durumda elde edilmiştir. LASSO modeli ile gerçekleştirilen değişken seçimi ile model performanslarının arttığı gözlemlenmiştir. Bu doğrultuda Google Trends 'den farklı frekanslarda toplanmış arama terimlerinin oluşturduğu değişken grubunu bir arada kullanmak ve değişken seçimi tekniğinden faydalanmak, işsizlik oranının daha etkin öngörülenmesini ve elde edilen bu tahmin sonuçlarının ekonomik karar süreçlerinde daha iyi rehberlik etmesini sağlayacaktır.

Anahtar Kelimeler: Zaman Serisi, İşsizlik Oranı, Google Trends, Makine Öğrenme Teknikleri, Değişken Seçimi.

ABSTRACT
Master's Thesis
Unemployment Forecast Using Google Trends Data
at Different Frequencies
Dilan KAYA

Dokuz Eylül University
Graduate School of Social Sciences
Department of Econometrics
Data Management and Analysis Program

Unemployment, a critical indicator for the health of an economy, holds significant importance, particularly in developing countries like Turkey, where it can heavily influence social balances. Hence, forecasting the unemployment rate, which represents the percentage of individuals seeking jobs but unable to secure one, holds value for effective economic planning. In this way, various policies can be implemented to eliminate the unemployment problem. However, forecasting the unemployment rate requires considering multiple variables as the problem is associated with several societal indicators. Macroeconomic and demographic factors like inflation rates, employment, and population figures are examples of these indicators. Moreover, with technological advancements, the use of various online sources has gained prominence in predicting unemployment rates. Online resources like Google Trends, reflecting the internet search trends of job-seeking individuals, can offer an alternative perspective in forecasting the unemployment rate. Therefore, it has been a matter of interest to determine to what extent these internet-based resources can improve the prediction performance of the variable of interest.

In this study, different frequency data sources obtained through Google Trends were brought together and their effects on unemployment rate estimation were examined. Different combinations of macroeconomic variables and search frequency data obtained from Google Trends were analyzed using nine different machine learning techniques. Moreover, variable combinations were reassessed

by applying variable selection techniques, resulting in the reconstruction of all models. Thus, it was aimed to determine the variable group that would provide the best prediction for the unemployment rate by comparing the performances of the models on a larger scale. As a result, the best performance in the study was obtained when all variables obtained from 3 different frequency types were included in the model. It has been observed that model performances increase with the variable selection made with the LASSO model. In this regard, using a variable group consisting of search terms collected at different frequencies from Google Trends and benefiting from the variable selection technique will enable the unemployment rate to be predicted more effectively and these prediction results will provide better guidance in economic decision processes.

Keywords: Time Series, Unemployment Rate, Google Trends, Machine Learning Techniques, Variable Selection.

FARKLI FREKANSLI GOOGLE TRENDS VERİLERİ İLE İŞSİZLİK ÖNGÖRÜSÜ

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
KISALTMALAR	xii
TABLolar LİSTESİ	xiv
ŞEKİLLER LİSTESİ	xvi
GİRİŞ	1

BİRİNCİ BÖLÜM

İŞGÜCÜ, İSTİHDAM VE İŞSİZLİK: KAVRAMSAL İNCELEME

1.1. İŞGÜCÜ VE İŞGÜCÜNE KATILMA ORANI	4
1.2. İSTİHDAM	6
1.2.1. Türkiye 'de İstihdamın Çeşitli Boyutları	7
1.2.1.1. Eğitim Seviyeleri ve İstihdam	8
1.2.1.2. Çalışanın İşteki Durumu ve İstihdam	9
1.2.1.3. Sektör Grupları ve İstihdam	10
1.2.1.4. Kamu ve Özel Sektör Açısından İstihdam	11
1.2.1.5. Türkiye 'de Genç Nüfusun İstihdama Katılımı	12
1.3. İŞSİZLİK KAVRAMI	13
1.3.1. İşsizlik Faktörleri	13
1.3.2. İşsizlik Çeşitleri	14
1.3.2.1. Çalışma İsteğine Göre İşsizlik	14
1.3.2.1.1. İstemli/İrادی İşsizlik	15
1.3.2.1.2. İstemsiz/İrade Dışı İşsizlik	15
1.3.2.2. İstihdam Tipine Göre İşsizlik	15

1.3.2.2.1. Gizli İşsizlik	15
1.3.2.2.2. Açık İşsizlik	16
1.3.2.2.2.1. Mevsimsel İşsizlik	16
1.3.2.2.2.2. Teknolojik İşsizlik	16
1.3.2.2.2.3. Konjonktürel İşsizlik	16
1.3.2.2.2.4. Friksiyonel/Arızı/Geçici İşsizlik	17
1.3.3. İşsizlik Sorununun Nedenleri	17
1.3.3.1. İşgücü Arzı Kaynaklı Nedenler	17
1.3.3.2. İşgücü Talebi Kaynaklı Nedenler	18
1.4. DÜNYADA İŞSİZLİK VE EVRİMİ	20
1.5. TÜRKİYE 'DE İŞSİZLİK VE EVRİMİ	22
1.6. İŞSİZLİĞİN TAHMİNLENMESİNİN ÖNEMİ	23

İKİNCİ BÖLÜM

MAKİNE ÖĞRENMESİ ÇEŞİTLERİ VE GOOGLE TRENDS

2.1. MAKİNE ÖĞRENMESİ	24
2.1.1. Makine Öğrenmesinin Diğer Disiplinlerle İlişkisi	24
2.1.2. Makine Öğrenmesinin Önemi	25
2.2. MAKİNE ÖĞRENMESİ ÇEŞİTLERİ	26
2.2.1. Denetimli Öğrenme	26
2.2.2. Denetimsiz Öğrenme	29
2.2.3. Yarı Denetimli Öğrenme	30
2.2.4. Pekiştirmeli Öğrenme	31
2.3. REGRESYON TABANLI DENETİMLİ MAKİNE ÖĞRENMESİ TÜRLERİ	32
2.3.1. Doğrusal Regresyon	32
2.3.2. Ridge, LASSO ve ElasticNet Regresyon	33
2.3.2.1. Ridge Regresyonu	33
2.3.2.2. LASSO Regresyonu	34
2.3.2.3. ElasticNet Regresyon	34
2.3.3. Destek Vektör Makineleri	35
2.3.4. Karar Ağaçları	39
2.3.5. Gradyan Güçlendirmeli Karar Ağaçları	41

2.3.6. Rassal Orman	44
2.3.7. K-En Yakın Komşu Algoritması	46
2.4. REGRESYON MODELLERİNDE PERFORMANS İSTATİSTİKLERİ	46
2.4.1. R-Kare	47
2.4.2. Düzeltilmiş R-Kare	47
2.4.3. Ortalama Mutlak Hata	48
2.4.4. Hata Kareler Ortalaması	48
2.4.5. Ortalama Karesel Karekök Hatası	48
2.4.6. Ortalama Mutlak Yüzdesel Hata	49
2.5. KARMA VERİ ÖRNEKLEMESİ	49
2.6. GOOGLE TRENDS	51
2.6.1. Coğrafi Konum ve Dil	51
2.6.2. Zaman	52
2.6.3. Kategori	52
2.6.4. Google Filtresi	52
2.6.5. İlgili Konular	53
2.6.6. İlgili Sorgular	53

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. ÇALIŞMANIN AMACI VE KAPSAMI	57
3.2. LİTERATÜR TARAMASI	58
3.3. DENEY TASARIMI	63
3.3.1. Veri Toplama	64
3.3.1.1. Google Trends Verileri	64
3.3.1.1.1. Günlük Verilerin Elde Edilmesi	66
3.3.1.1.2. Haftalık Verilerin Elde Edilmesi	67
3.3.1.1.3. Aylık Verilerin Elde Edilmesi	67
3.3.1.1.4. Günlük, Haftalık, Aylık Verilerin Konsepti	67
3.3.1.2. Makroekonomik Veriler	69
3.3.2. Veri Ön işleme	69
3.3.2.1. Veri Temizleme	70

3.3.2.2. Verilerin Hazırlanması	71
3.3.3. Veri Setlerinin İstatistiksel Özellikleri	72
3.3.3.1. Google Trends Veri Seti Tanımlayıcı İstatistikleri	72
3.3.3.2. Makroekonomik Dışsal Veri Setinin Tanımlayıcı İstatistikleri	76
3.3.4. Değişkenlerin Düzenlenmesi	77
3.3.5. Verilerin Durağanlaştırılması	79
3.3.6. Verilerin Standartlaştırılması	80
3.3.7. Eğitim ve Test Veri Setlerinin Belirlenmesi	80
3.3.8. Değişken Seçimi	82
3.3.9. Model Kurulumu	84
3.3.10. Model Performanslarının Değerlendirilmesi	85
3.4. BULGULAR VE BULGULARIN YORUMLANMASI	86
SONUÇ	110
KAYNAKÇA	113

KISALTMALAR

ABD	Amerika Birleşik Devletleri
AdaBoost.	Uyarlanabilir Yükseltme
ANN	Yapay Sinir Ağları
AR	Otoregresif Model
ARIMA	Bütünleşik Otoregresif Hareketli Ortalama
BIC	Bayesyen Bilgi Kriteri
CART	Sınıflandırma ve Regresyon Ağacı
CHAID	Ki-kare Otomatik Etkileşim Tespiti
DiD	Farklılıklar Arasındaki Fark
DT	Karar Ağaçları
DTC	Karar Ağacı Sınıflandırması
DTR	Karar Ağacı Regresyonu
EGE	Ekonomi Güven Endeksi
G20	20 Grubu
GBDT	Gradyan Güçlendirmeli Karar Ağaçları
GDP	Gayrisafi Yurtiçi Hasıla
GJSI	Google İş Arama Endeksi
GSMH	Gayrisafi Milli Hasıla
GT	Google Trends
ID3	İteratif İkiliye Ağırlık Verme Ağacı
ILO	Uluslararası Çalışma Örgütü
IMF	Uluslararası Para Fonu
IP	İnternet Protokol Adresi
İKO	İşgücüne Katılma Oranı
İŞKUR	İş ve İşçi Bulma Kurumu
KNN	K-En Yakın Komşu
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LASSO	En Küçük Mutlak Büyüklük ve Seçim Operatörü
LDA	Gizil Dirichlet Dağılımı
LightGBM	Hafif Gradyan Artırma Makinesi

LSTM	Uzun-Kısa Vadeli Bellek
M5	Model Ağacı
MAE	Ortalama Mutlak Hata
MAPE	Ortalama Mutlak Yüzdesel Hata
MFML	Karma Frekanslı Makine Öğrenmesi
MIDAS	Karma Veri Örneklemesi
MSE	Hata Kareler Ortalaması
OECD	Ekonomik İşbirliği ve Kalkınma Örgütü
RF	Rassal Orman
RMSE	Ortalama Karesel Karekök Hatası
SVC	Destek Vektör Sınıflandırma
SVM	Destek Vektör Makineleri
SVR	Destek Vektör Regresyonu
TCMB	Türkiye Cumhuriyeti Merkez Bankası
TGE	Tüketici Güven Endeksi
TÜFE	Tüketici Fiyat Endeksi
TÜİK	Türkiye İstatistik Kurumu
UMIDAS	Sınırlandırılmamış Karma Veri Örneklemesi
VAR	Vektör Otoregresif Model
XGBoost	Aşırı Gradyan Artırma

TABLÖLAR LİSTESİ

Tablo 1: Mevsim Etkisinden Arındırılmış Temel İşgücü Göstergeleri (2021 Ocak-2023 Temmuz) [15+ Yaş]	s. 6
Tablo 2: Eğitim Durumuna Göre Temel İşgücü Göstergeleri: İstihdam [15+ Yaş] (Bin Kişi)	s. 8
Tablo 3: İstihdam Edilenlerin İşteki Durumu [15+ Yaş] (Bin Kişi)	s. 9
Tablo 4: Sektöre Göre İstihdam Edilenler (%)	s. 11
Tablo 5: Uygulama Modelleri	s. 64
Tablo 6: Google Trends Arama Sorguları	s. 65
Tablo 7: Veri Çekme İşlemi için Oluşturulan Değişken Grupları	s. 66
Tablo 8: Farklı Frekanslı Arama Terimlerini Oluşturmak İçin Veri Çekme Periyotları	s. 68
Tablo 9: Nihai GT Değişkenleri	s. 70
Tablo 10: Google Trends Değişkenlerinin Tanımlayıcı İstatistikleri	s. 73
Tablo 11: Makroekonomik Değişkenlerin Tanımlayıcı İstatistikleri	s. 76
Tablo 12: LASSO Değişken Seçimi ile En Çok Seçilen Değişkenler	s. 83
Tablo 13: Modellerin Hiperparametre Dizini	s. 85
Tablo 14: Otoregresif Modellerin Performans İstatistikleri	s. 86
Tablo 15: Performans Hesaplamaları	s. 87
Tablo 16: Değişken Seçimi Olmadan: Makro + AR(3) Model Performansları	s. 88
Tablo 17: Değişken Seçimi Olmadan: Aylık GT + Makro + AR(3) Model Performansları	s. 89
Tablo 18: Değişken Seçimi Olmadan: Haftalık GT + Makro + AR(3) Model Performansları	s. 90
Tablo 19: Değişken Seçimi Olmadan: Günlük GT + Makro + AR(3) Model Performansları	s. 91
Tablo 20: Değişken Seçimi Olmadan: Tüm Veriler ile Model Performansları	s. 92
Tablo 21: LASSO Değişken Seçimi ile: Makro + AR(3) Model Performansları	s. 93
Tablo 22: LASSO Değişken Seçimi ile: Aylık GT + Makro + AR(3) Model Performansları	s. 94

Tablo 23: LASSO Değişken Seçimi ile: Haftalık GT + Makro + AR(3) Model Performansları	s. 95
Tablo 24: LASSO Değişken Seçimi ile: Günlük GT + Makro + AR(3) Model Performansları	s. 96
Tablo 25: LASSO Değişken Seçimi ile: Tüm Veriler ile Model Performansları	s. 97
Tablo 26: Modellerin En İyi Değişken Kombinasyonları	s. 98



ŞEKİLLER LİSTESİ

Şekil 1: 2023 Ocak – Ağustos Dönemi Yaş Gruplarına Göre İşe Yerleştirmeler (Bin Kişi)	s. 12
Şekil 2: İşsizlik Tipleri	s. 14
Şekil 3: 2010-2022 Dönemi Büyük Gelişmiş Ülkeler ve Türkiye ’deki İşsizlik Oranları (%)	s. 22
Şekil 4: 2005 – 2022 Dönemi Türkiye İşsizlik Oranları (%)	s. 23
Şekil 5: Makine Öğrenmesi Çeşitleri	s. 26
Şekil 6: Sınıflandırma Problemleri İçin Temel Gösterim	s. 27
Şekil 7: Regresyon Problemleri İçin Temel Gösterim	s. 28
Şekil 8: Kümeleme Problemleri İçin Temel Gösterim	s. 29
Şekil 9: Denetimsiz Öğrenme Süreci	s. 30
Şekil 10: Yarı Denetimli Öğrenme Süreci	s. 31
Şekil 11: Pekiştirmeli Öğrenme Süreci	s. 32
Şekil 12: Doğrusal Regresyon Modeli	s. 33
Şekil 13: Tek boyutlu Doğrusal Destek Vektör Regresyonu Modeli	s. 35
Şekil 14: Quadratic Kayıp Fonksiyonu	s. 37
Şekil 15: Laplace Kayıp Fonksiyonu	s. 38
Şekil 16: Huber Kayıp Fonksiyonu	s. 38
Şekil 17: Vapnik Kayıp Fonksiyonu	s. 39
Şekil 18: DT Genel Yapısı	s. 39
Şekil 19: Gradient Boosting Decision Tree Mimari Yapısı	s. 42
Şekil 20: Sınıflandırıcı RF Örnek Gösterimi	s. 45
Şekil 21: Regresyon Tabanlı RF Örnek Gösterimi	s. 45
Şekil 22: Google Trends: “İlgili Konular” Örneği	s. 53
Şekil 23: Google Trends: “İlgili Sorgular” Örneği	s. 54
Şekil 24: Google Trends: “işsizlik” Sorgusunun Zaman İçinde Arama Yoğunluğu (23 Ekim 2023 – 30 Ekim 2023)	s. 54
Şekil 25: Google Trends: “iş” ve “iş ilanları” Sorgularının Zaman İçinde Arama Yoğunluğu (30 Ekim 2022 – 30 Ekim 2023)	s. 55
Şekil 26: Genişleyen Pencere Yöntemi	s. 81

Şekil 27: Değişken Seçimi Olmadan “Makro + AR(3)” Ridge Modeli	s. 101
Şekil 28: Değişken Seçimi Olmadan “Haftalık GT + Makro + AR(3)” DT Modeli	s. 102
Şekil 29: Değişken Seçimi Olmadan “Tümü” SVR Modeli	s. 103
Şekil 30: Değişken Seçimi Olmadan “Tümü” XGBoost Modeli	s. 104
Şekil 31: LASSO Değişken Seçimi ile “Günlük GT + Makro + AR(3)” KNN Modeli	s. 105
Şekil 32: LASSO Değişken Seçimi ile “Tümü” ElasticNet Modeli	s. 106
Şekil 33: LASSO Değişken Seçimi ile “Tümü” LASSO Modeli	s. 107
Şekil 34: LASSO Değişken Seçimi ile “Tümü” RF Modeli	s. 108
Şekil 35: LASSO Değişken Seçimi ile “Tümü” GBDT Modeli	s. 109

GİRİŞ

Gelişmekte olan ülkelerde ekonomik göstergeler, toplum dinamikleri üzerinde önemli etkilere sahiptir. Dolayısıyla işsizlik oranı, bir ülkenin ekonomik açıdan gücünü yansıtan ana unsurlardan biridir. Türkiye gibi ekonomik açıdan hareketli koşullara sahip olan bir ülkede işsizlik kavramı sosyal, politik ve ekonomik boyutlarıyla ele alınmalıdır. İşsizlik oranının öngörülmesi, ekonomik anlamda gerçekleştirilen iyileştirme çalışmaları için önemli bir araç haline gelmiştir. Bu durum, işsizlik sorununu çözmek adına geliştirilecek çözüm stratejileri için önemli bir adımdır. Öte yandan, çok sayıda değişken ile ilişki içerisinde olan işsizlik oranının öngörülmesi süreci oldukça karmaşık bir yapıya sahiptir. Özellikle, klasik istatistiksel ve ekonometrik modellerin performansının yetersiz kaldığı durumlarda, makine öğrenmesi yüksek boyutlu veri kümeleri üzerinde etkin tahminlerde bulunma potansiyeli sunarak son yıllarda öne çıkmıştır.

Bu çalışma, Türkiye 'deki işsizlik oranı öngörüsü yapmak amacıyla günlük, haftalık ve aylık ölçekteki Google Trends verileri ile makroekonomik göstergelerin birleştirilerek analiz edilmesini ele almaktadır. Böylece farklı frekanslarda toplanan Google Trends verilerinin işsizlik öngörüsü üzerinde sağlayabileceği faydaların tespit edilmesi ve karşılaştırmasının yapılması istenmiştir. Değişen frekanslarda kullanılan bu verilerin analize katkısını değerlendirmek amacıyla, makroekonomik dışsal değişkenler ile çeşitli Google Trends arama terimleri, beş farklı modelde bir araya getirilerek modellerin kurulması hedeflenmiştir. Bu aşamada farklı frekanslı değişkenleri aynı modele sokabilmek için UMIDAS model mantığından faydalanılmıştır. Değişken seçim tekniğinin sonuçları iyileştirip iyileştirmeyeceğini tespit etmek amacıyla LASSO modelinden faydalanılmıştır. Bu amaç için, LASSO modeli ile seçilen değişkenlerle tüm makine öğrenme modelleri tekrar oluşturulmuştur. Böylece Türkiye 'de işsizlerin iş ararken internette hangi arama terimlerini sıklıkla tercih ettiklerinin analizi gerçekleştirilmiştir. Çalışmanın sonucunda ise, uygulama modelleri içerisinde en iyi öngörü performansının elde edildiği model belirlenmiş ve işsizlik oranı analizinde uygulanabilecek çözüm önerileri sunulmuştur. Bu entegrasyonun ekonomik analizlerdeki rolü, önemi ve

ekonomik karar süreçlerine katkı sağlamasının yanında, araştırmacılara farklı ve kapsamlı bir bakış açısı sunması hedeflenmiştir.

Çalışmanın birinci bölümünde işsizlik kavramsal olarak incelenmiştir. İşsizlik kavramı incelenirken istihdam ve işgücü gibi işsizlikle doğrudan ilişkili unsurlar da ele alınmıştır. Türkiye 'de istihdamın boyutları detaylandırılmış ve resmi kaynaklardan elde edilen çeşitli tablolar sunularak istihdam kavramı somutlaştırılmıştır. Ayrıca, işsizlik faktörleri ve işsizlik çeşitleri ele alınarak işsizliğin karmaşık yapısını vurgulamak hedeflenmiştir. İşsizliğin işgücü arzı veya talebi kaynaklı nedenleri açıklanarak hem Türkiye hem de küresel düzeyde evrimi incelenmiştir.

Çalışmanın ikinci bölümünde, makine öğrenmesi kavramı ve diğer disiplinlerle olan ilişkisinden bahsedilmiştir. Sonrasında makine öğrenmesi teknikleri açıklanarak dört ana çeşidi detaylandırılmıştır. Aynı zamanda regresyon tabanlı denetimli makine öğrenmesi algoritmalarından ve bu algoritmalarla geliştirilen modellerin performansını ölçmeye yardımcı olan performans istatistiklerinden bahsedilmiştir. Son olarak çalışma kapsamında internet kullanıcılarının arama frekanslarının elde edildiği Google Trends platformunun tanımı yapılmış ve özellikleri görsellerle desteklenerek detaylandırılmıştır.

Çalışmanın üçüncü ve son bölümünde ise, araştırmanın ve uygulamanın amaçlarından ve kapsamından söz edilmiştir. Bunun yanı sıra uygulama için geliştirilen deney tasarımı adım adım açıklanmıştır. Uygulanan bu adımların sözel ve görsel açıklamalarına yer verilmiştir. Ek olarak makroekonomik göstergeler ve Google Trends değişkenlerine istinaden elde edilen ve çalışmada kullanılan verilerin elde edilmesi ile çalışmaya uygun hale getirilmesi süreçleri detaylı bir biçimde aktarılmıştır. Elde edilen bu veri setlerine ilişkin analizi hedeflenen modeller belirtilmiş ve bu modeller değişken seçim tekniğiyle tekrarlanmıştır. Son olarak, analiz sonucunda elde edilen modellere ait performans istatistikleri, işsizlik oranı serisinin gecikmeleri ile oluşturulan kıyaslama modelinin (benchmark) performansı ile karşılaştırılmış ve elde edilen sonuçlar üzerinden tahmin sonuçları yorumlanmıştır. Sonuçlara istinaden çeşitli frekanslara sahip değişkenlerin analiz performansını artırdığı ve Google Trends gibi çevrimiçi kaynaklardan elde edilen kullanıcı

davranışlarının, işsizlik oranını öngörümlemede fayda sağlamış olduğu tespit edilmiştir.



BİRİNCİ BÖLÜM

İŞGÜCÜ, İSTİHDAM VE İŞSİZLİK: KAVRAMSAL İNCELEME

Bu bölümde işgücü, işgücüne katılma oranı, istihdam, işsizlik, işsizlik faktörleri, işsizlik nedenleri, işsizliğin dünyada ve Türkiye 'deki evrimi ele alınarak detaylandırılacaktır.

1.1. İŞGÜCÜ VE İŞGÜCÜNE KATILMA ORANI

Ekonomik üretimde yer alan önemli faktörlerden biri olan işgücü kavramı, genel itibariyle aktif olarak çalışan bireyler ve iş arayanlar (işsizler) toplamı olarak açıklanır (Barro, 1997: 351).

Türkiye 'nin ekonomik, sosyal ve demografik verilerinin anlaşılmasına olanak tanıyan Türkiye İstatistik Kurumu (TÜİK) ise işgücü kavramını belirli bir zaman dilimindeki nüfusu üç temel gruba ayırarak açıklar. Söz konusu olan bu üç temel grup; istihdam edilenler, işsizler ve işgücüne dahil olmayanlar olarak ifade edilmektedir. İstihdam edilenler, belirli bir dönemde düzenli bir işte çalışan kişileri temsil ederken, işsizlikle karşı karşıya olanlar, belirli bir zaman diliminde işe alınmayıp, son dört hafta içinde iş arama kanallarından en az birini kullanmış ve 2 hafta içinde işe başlayabilecek durumda olan ve kurumsal olmayan çalışma çağındaki tüm kişiler olarak ifade edilmektedir. Bu ifadelere göre işsizlikle karşı karşıya olanlar ve istihdam edilenlerin toplamı, bir ülkenin üretkenlik seviyesini değerlendirmeye olanak tanıyan ve ekonomik dinamikleri anlamak açısından önemli rolü olan toplam işgücünü oluşturmaktadır (Bağcı, 2020: 4). Bu kavramlar bir ülkedeki işgücünün büyüklüğünü ölçümlemek için kullanılan işgücüne katılma oranının hesaplanmasına yardımcı olur. Ekonomik literatürde işgücüne katılma oranı, işgücüne dahil olan kişilerin sayısı anlamına gelmektedir.

İşgücüne katılma oranı hesaplanırken istihdam edilenler ve işsizlikle karşı karşıya olanların toplamı, ilgili zaman dilimindeki aktif nüfusa oranlanır. Bir ülkenin işgücü piyasasının sağlığını ve ekonomik faaliyet düzeyini değerlendirmek için önemli bir gösterge olan bu oran sayesinde işgücünün aktif nüfusa göre ne derece büyük olduğunu gözlemleyebiliriz. İşgücüne katılma oranı (İKO) ekonomik kaynaklarda

alıřanlar ve iřsizler toplamının aktif nfusa oranlanarak 100 ile arpılması sonucunda elde edilir (Gndoėan ve diėerleri, 2019: 12).

Tablo 1, 2021 Ocak-2023 Temmuz dnemini kapsayan, TK tarafından 15 yař ve zeri nfus iin hesaplanan ve mevsim etkisinden arındırılmıř olan temel iřgc gstergelerini iermektedir. Bu gstergelerden en nemlisi iřgcne katılma oranıdır. İřgcne katılma oranındaki deėiřimler istihdam, iřsizlik ve aktif nfus arasındaki iliřkileri doėrudan yansıtmaktadır. Trkiye ’de iřgcne katılma oranı 2021 yılı bařında %50,0 iken, 2022 yılına %52,6 ile giriř yapılmıřtır ve ardından 2023 yılının Ocak ayında ise bu oran %53,9 ’a ykselmiřtir. Bu artıř, iřgc piyasasında iř bulma olasılıėının zamanla artabileceėini ve iřsizlik seviyesinin giderek azalabileceėini gsterir. Ancak, bu verilerin anlamını tam olarak kavrayabilmek iin ekonomik faktrlerin, eėitim dzeyinin ve sektrel deėiřikliklerin de analize dahil edilmesi gerekmektedir. Bu faktrler sayesinde iřgc piyasalarındaki deėiřikliklerin kaynaklarını ve bu deėiřikliklerin lkede yarattıėı ekonomik sonuları daha kapsamlı bir aıdan inceleme fırsatı bulunabilir.

Tablo 1: Mevsim Etkisinden Arındırılmış Temel İşgücü Göstergeleri (2021 Ocak-2023 Temmuz) [15+ Yaş]

Yıllar		15+ yaş Nüfus (bin kişi)	İstihdam (bin kişi)	İşsiz (bin kişi)	İşgücüne katılma oranı(%)
2021	<i>Ocak</i>	63.236	27.647	3.955	50
	<i>Şubat</i>	63.323	27.672	4.241	50,4
	<i>Mart</i>	63.401	28.400	4.179	51,4
	<i>Nisan</i>	63.488	28.293	4.345	51,4
	<i>Mayıs</i>	63.572	28.012	4.284	50,8
	<i>Haziran</i>	63.659	28.632	3.561	50,6
	<i>Temmuz</i>	63.744	28.807	3.780	51,1
	<i>Ağustos</i>	63.831	28.988	3.927	51,6
	<i>Eylül</i>	63.918	29.560	3.755	52,1
	<i>Ekim</i>	64.002	29.605	3.666	52
	<i>Kasım</i>	64.089	29.608	3.710	52
	<i>Aralık</i>	64.173	30.036	3.704	52,6
2022	<i>Ocak</i>	64.256	29.987	3.794	52,6
	<i>Şubat</i>	64.334	30.003	3.634	52,3
	<i>Mart</i>	64.405	30.078	3.748	52,5
	<i>Nisan</i>	64.484	30.515	3.699	53,1
	<i>Mayıs</i>	64.560	30.856	3.746	53,6
	<i>Haziran</i>	64.639	30.762	3.568	53,1
	<i>Temmuz</i>	64.714	30.466	3.477	52,5
	<i>Ağustos</i>	64.793	31.036	3.372	53,1
	<i>Eylül</i>	64.871	30.977	3.463	53,1
	<i>Ekim</i>	64.947	31.034	3.499	53,2
	<i>Kasım</i>	65.026	31.498	3.542	53,9
	<i>Aralık</i>	65.102	31.361	3.578	53,7
2023	<i>Ocak</i>	65.166	31.686	3.417	53,9
	<i>Şubat</i>	65.214	31.333	3.526	53,5
	<i>Mart</i>	65.257	31.162	3.465	53,1
	<i>Nisan</i>	65.305	31.633	3.523	53,8
	<i>Mayıs</i>	65.352	31.706	3.328	53,6
	<i>Haziran</i>	65.400	31.353	3.329	53
	<i>Temmuz</i>	65.446	31.671	3.291	53,4

Kaynak: TÜİK, 2023.

1.2. İSTİHDAM

İşgücü piyasasındaki boyutu ve niteliği toplumların ekonomik etkinliği ve direnci açısından son derece önemli olan istihdam kavramı, birçok çalışmaya konu olmuştur. Bu nedenle ekonomik literatürde dar ve geniş açıdan çeşitli tanımlara sahiptir. Dar anlamda, belirli bir zaman diliminde işgücü, toprak ve sermaye gibi üretim faktörlerinin mevcut teknolojik kapasiteye göre ne miktarda kullanıldığını belirtirken geniş anlamda, bireylerin ekonomik faaliyetlerde bulunarak gelir elde etme

süreçlerini tanımlamaktadır (Eğilmez, 2023: 185). TÜİK ise istihdamı iki ana kategori altında değerlendirir: İşbaşında olanlar ve işbaşında olmayanlar. Buna göre bu iki gruba dahil olan kurumsal olmayan çalışma çağındaki tüm bireyler istihdamı oluşturur.

İşbaşında olanlar kategorisi, gelir elde eden dolayısıyla ekonomik faaliyette bulunan bireyleri kapsamaktadır. Yevmiyeli, ücretli, maaşlı, kendi hesabına çalışan, işveren ya da ücretsiz aile işçisi olarak tanımlanan bu bireyler, referans dönemi içerisinde en az bir saat çalışarak ekonomiye katkıda bulunmuşlardır. İşbaşında olmayanlar kategorisi ise, çeşitli sebeplerle çalışma hayatına devam etmeyen ancak işi ile bağlantısı devam eden bireyleri kapsamaktadır. Bu bireyler, kendi hesabına çalışanlar ya da işverenler olarak istihdam edilir. Öte yandan ekonomik faaliyette bulunmadıkları belirli bir dönem vardır.

Bu tanımlamalara göre ekonominin genel sağlığını ve bireylerin ekonomik faaliyetlere katılımını değerlendirmek için kritik öneme sahip olan istihdam oranının önemi ortaya çıkmaktadır. İstihdam oranı literatürde istihdamın aktif nüfusa oranı olarak tanımlanmaktadır (Gündoğan ve diğerleri, 2019: 9).

Tablo 1 'de ifade edilen 2023 Temmuz ayı verilerine göre ilgili tarihte istihdam %48,4 oranındadır. Bu durumda, ilgili tarihte ülkemizde her 100 kişiden 48,4 kişi ekonomik faaliyetlere katılmış ve bu faaliyetlerden gelir elde etmiş, yani istihdam edilmiştir. Bu sayı işsizlik hakkında bir fikir verse de işsizlik oranını etkileyen pek çok unsur olduğu göz önünde bulundurulmalıdır. Bu sebepten yalnızca istihdam oranını dikkate alarak işsizlik oranı hakkında yorum yapmak pek sağlıklı olmayacaktır ancak istihdam ve işsizliğin birbirini tamamlayan kavramlar olduğu göz ardı edilmemelidir. Bu iki kavram her zaman birbiriyle ilişkisi değerlendirilerek incelenmelidir (Ay, 2012: 322).

1.2.1. Türkiye 'de İstihdamın Çeşitli Boyutları

Türkiye 'nin çeşitlilik gösteren sosyoekonomik koşullara sahip olması istihdamı farklı açılardan incelemenin gerekliliğini ortaya koymaktadır. Bu nedenle, istihdamı değerlendirirken eğitim seviyesi, sektörel dağılım, coğrafi dağılım gibi farklı perspektifleri de ele almak büyük önem taşır.

1.2.1.1. Eğitim Seviyeleri ve İstihdam

Eğitim seviyesi, bir toplumun gelişmişlik seviyesi hakkında bilgi verme yönüyle ekonomik dinamiklere katkıda bulunur. Dolayısıyla bu yönüyle bireylerin işgücü piyasasında elde edebileceği pozisyonları ve gelir seviyelerini doğrudan etkiler. Genel bir kanı olarak yüksek eğitim seviyesine sahip bireylerin, uzmanlık gerektiren veya daha yüksek maaşlı işlere erişim olasılığı daha yüksektir. Bu nedenle eğitim seviyesi arttıkça istihdam edilme olasılığı da artmaktadır. Tablo 2 'de 2021 yılının 1. çeyreğinden 2023 yılının 2. çeyreğine kadar olan dönemde eğitim seviyesine göre temel işgücü göstergeleri ifade edilmektedir.

Tablo 2: Eğitim Durumuna Göre Temel İşgücü Göstergeleri: İstihdam [15+ Yaş] (Bin Kişi)

Yıllar	Okur-yazar olmayanlar	Lise altı eğitilmişler	Lise	Mesleki veya teknik lise	Yüksek-öğretim
2021 I. Çeyrek	639	12.889	3.025	3.055	7.783
II. Çeyrek	676	13.565	3.168	3.114	7.925
III. Çeyrek	825	13.901	3.541	3.317	8.067
IV. Çeyrek	798	13.848	3.587	3.396	8.186
2022 I. Çeyrek	754	13.357	3.677	3.417	8.214
II. Çeyrek	801	14.059	3.930	3.491	8.484
III. Çeyrek	832	14.024	4.239	3.536	8.529
IV. Çeyrek	837	14.096	4.170	3.605	8.847
2023 I. Çeyrek	689	13.509	4.125	3.643	8.890
II. Çeyrek	769	13.759	4.302	3.619	9.058

Kaynak: TÜİK, 2023.

Bu verilere göre ilgili tarihler arasında istihdam edilen bireylerin çoğunluğunun lise altı eğitim seviyesine sahip olduğu görülmektedir. Bu veri, Türkiye 'nin gelişmekte olan bir ülke olarak eğitim seviyesinin hala yetersiz olduğuna işaret etmektedir. Ancak, “lise altı eğitilmişler” sınıfını “lise”, “mesleki veya teknik lise”,

“yüksek öğretim” gibi yükselen eğitim seviyesine sahip sınıfların toplamı ile kıyasladığımızda bir miktar daha düşük istihdam sayısına sahip olduğu söylenebilir. “Okur-yazar olmayanlar” sınıfı ise en az istihdam edilenler grubunda yer almaktadır. Bu bağlamda, ülkemizde eğitim seviyesi yükseldikçe istihdam edilen sayısının artması, daha yüksek eğitim seviyesine sahip bireylerin iş gücüne katılımının ekonomik büyümeyi destekleyebileceği ve daha fazla beceri gerektiren işlerin oluşturulmasına katkı sağlayabileceği fikrini verir. Yüksek eğitim seviyesine sahip bireyler, genellikle daha üretken ve yaratıcı olabilirler. Bu da ülkeye ekonomik kalkınma açısından olumlu bir etki sağlayabilir.

1.2.1.2. Çalışanın İşteki Durumu ve İstihdam

İstihdamın değerlendirilmesinde işgücündeki bireylerin işteki statüleri de büyük bir önem taşımaktadır. Bu, bir ülkenin ekonomik gelişmişlik düzeyini yansıtan önemli göstergelerden biridir. Tablo 3, 2021 yılının 1. çeyreğinden 2023 yılının 2. çeyreğine kadar olan dönemde istihdam edilen bireylerin işteki durumu yani statülerine göre dağılımını göstermektedir.

Tablo 3: İstihdam Edilenlerin İşteki Durumu [15+ Yaş] (Bin Kişi)

<i>Yıllar</i>		<i>Toplam</i>	<i>Ücretli /Yevmiyeli</i>	<i>İşveren</i>	<i>Kendi hesabına</i>	<i>Ücretsiz aile işçisi</i>
2021	<i>I. Çeyrek</i>	27.391	19.140	1.265	4.569	2.416
	<i>II. Çeyrek</i>	28.448	19.918	1.328	4.483	2.719
	<i>III. Çeyrek</i>	29.652	20.532	1.344	4.729	3.047
	<i>IV. Çeyrek</i>	29.815	21.019	1.329	4.884	2.583
2022	<i>I. Çeyrek</i>	29.418	20.844	1.335	4.921	2.317
	<i>II. Çeyrek</i>	30.765	21.583	1.379	5.072	2.732
	<i>III. Çeyrek</i>	31.160	21.855	1.374	4.995	2.937
	<i>IV. Çeyrek</i>	31.556	22.604	1.417	5.068	2.467
2023	<i>I. Çeyrek</i>	30.856	22.194	1.426	5.017	2.219
	<i>II. Çeyrek</i>	31.507	22.415	1.472	5.127	2.493

Kaynak: TÜİK, 2023.

Bu tabloya göre istihdam edilen bireylerin işteki durumları; yevmiyeli, ücretli, maaşlı, kendi hesabına çalışan, işveren ya da ücretsiz aile işçisi olarak ayrıştırılmıştır. İlgili veriler, ülkemizde “ücretli veya yevmiyeli” sınıfındaki bireylerin en yüksek istihdam edilen grup, “işveren” sınıfındaki bireylerin en az istihdam edilen grup olduğunu gösterir. Gelişmiş toplumlarda, genellikle en yüksek istihdam oranı ücretli sınıfında bulunur (Şahin, 2007: 559). En yüksek istihdam sayısının ücretli sınıfında olması, ekonominin sağlam bir temele dayandığını ve işçi haklarının daha iyi korunduğuna işaret edebilir. Ayrıca, ücretli işlerin genellikle daha istikrarlı olduğu ve çalışanlara daha fazla sosyal güvence sağladığı söylenebilir. Öte yandan, işveren olmak, işte daha yüksek bir statüyü temsil edebilir. Ancak tablodan da görülebileceği üzere bu sınıftaki bireylerin sayısı genellikle daha düşüktür çünkü bu pozisyon riskli olarak bilinir ve daha fazla sorumluluk getirecektir. Sonuç olarak, istihdamın farklı sınıfları, ülkemizin ekonomik yapısı ve iş dünyasının dinamiklerini yansıtan önemli bir göstergedir. İşgücündeki bu çeşitlilik, ekonomik büyüme, gelir dağılımı ve sosyal refah gibi makroekonomik faktörler üzerinde doğrudan etkili olabilir.

1.2.1.3. Sektör Grupları ve İstihdam

Ekonomik faaliyetlerin gerçekleştiği sektörlerle göre istihdamın dağılımı, ekonominin hangi alanlarda daha aktif olduğu ve hangi sektörlerin büyüme potansiyeline sahip olduğu hakkında genel anlamda bilgi verebilir. Bu açıdan sektörel dağılım bir ülkenin ekonomik yapısının önemli bir yansımasıdır ve ülkenin ekonomik ve sosyal gelişmişlik düzeyini değerlendiren temel göstergelerden birini oluşturur (Şahin, 2007: 557). Tarım, sanayi ve hizmet sektörleri, ülkemizde de istihdamın ana sektörleridir ve her biri farklı beceri veya eğitim seviyesi gerektirir. Tablo 4, 2021 'in 1. çeyreğinden 2023 'ün 2. çeyreğine kadar olan dönemde tarım, sanayi, inşaat ve hizmet sektörlerinde istihdam edilen bireylerin sayısını göstermektedir.

Tablo 4: Sektöre Göre İstihdam Edilenler (%)

Yıllar		Toplam	Tarım	Sanayi	İnşaat	Hizmetler
2021	I. Çeyrek	100	16,7	21,7	5,9	55,6
	II. Çeyrek	100	17,8	21,4	6,2	54,6
	III. Çeyrek	100	18,3	20,9	6,3	54,5
	IV. Çeyrek	100	15,9	21,7	6,1	56,3
2022	I. Çeyrek	100	14,8	22,2	5,5	57,4
	II. Çeyrek	100	16,2	22	6	55,9
	III. Çeyrek	100	16,9	21,3	6,1	55,7
	IV. Çeyrek	100	14,7	21,3	6,4	57,5
2023	I. Çeyrek	100	13,5	22,1	5,9	58,5
	II. Çeyrek	100	14,9	21,6	6,1	57,3

Kaynak: TÜİK, 2023.

Bu zaman dilimi sektörel bazda incelendiğinde, istihdam oranlarında belirgin bir artış veya azalış yaşanmamıştır. Ancak genel olarak değerlendirildiğinde, ülkemizde diğer sektörler göre en az istihdamın sağlandığı sektörün inşaat, en fazla istihdamın sağlandığı sektörün ise hizmet sektörü olduğu görülmektedir. Tarım ve sanayi sektörleri de istihdam açısından önemli bir rol oynamaktadır.

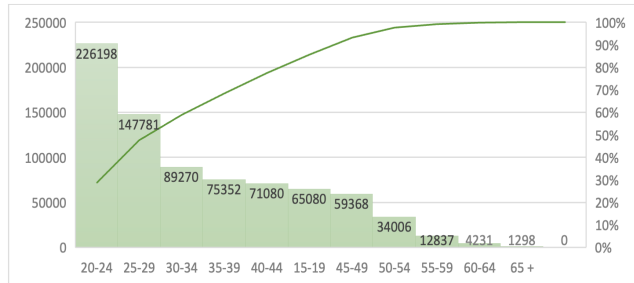
1.2.1.4. Kamu ve Özel Sektör Açısından İstihdam

Toplumdaki bireylere genellikle daha iyi çalışma koşulları sunan kamu sektörü, daha istikrarlı iş imkanlarıyla, daha çok kariyer fırsatı sunan özel sektör ise daha dinamik ve rekabetçi yapısıyla bilinir. İş ve İşçi Bulma Kurumu (İŞKUR) 'un yayınladığı 2023 Ocak-Ağustos dönemi toplam işe yerleştirme verilerine göre kamu sektöründe istihdam 10.534 iken özel sektörde istihdam 775.967 'dir. Toplam istihdam ise 786.501 'dir. Bu değerlere göre ilgili dönemin istihdam sayısında kamu sektörünün yeri %1 'lik bir dilimi kapsamaktadır (İŞKUR, 2023). Dolayısıyla güncel verilere göre işe yerleştirmelerde özel sektörün oldukça büyük bir paya sahip olduğu söylenebilir.

1.2.1.5. Türkiye 'de Genç Nüfusun İstihdama Katılımı

Tüm dünyada genç nüfus, işgücü piyasasına yeni dahil olan ve bu anlamda ekonomiye bulunduğu katkıyla ekonomik büyümeyi ve toplumsal refahı destekleyen en dinamik grup olma özelliğine sahiptir. Dolayısıyla genç nüfusun istihdam edilenler arasındaki payının incelenmesi toplumların ekonomisi hakkında birçok fikir sunar. Öte yandan, genç nüfusun işe alınması ve işgücü piyasasına katılması ile ilgili olan bu süreç pek çok faktöre bağlıdır. Bu faktörlere bireylerin eğitim seviyeleri, becerileri ve toplumun ekonomik koşulları örnek verilebilir. Türkiye açısından bu durum incelendiğinde, hızla artan genç nüfus yapısı dikkat çekmektedir. Türkiye 'de toplam nüfusun %34,8 'i 15-24 yaş aralığındadır. Buna ek olarak, 2023 yılı Ağustos ayı verilerine göre Türkiye 'de 15-24 yaş aralığındaki genç nüfusun istihdam oranı %33,8 olarak kayda geçmiştir. Bu oranın bir önceki yıla göre 1,2 puan arttığı söylenebilir (TÜİK, 2023). Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) ülkelerinde 2023 yılı ortalama genç nüfus istihdam oranı %54,7 'dir (OECD, 2023). Türkiye 'nin bu oranla OECD ülkeleri ortalamasının altında olduğu açıktır. Genç nüfus istihdam oranının artması için eğitim seviyesinde iyileştirmeler, işgücü piyasası politikalarındaki sağlamlaştırmalar ve cinsiyet ayrımıyla mücadele çalışmaları yapılabilir. Şekil 1 'de 2023 yılı Ocak-Ağustos dönemi İŞKUR verilerine göre, Türkiye 'de yaş gruplarına göre işe yerleştirme verileri sunulmuştur.

Şekil 1: 2023 Ocak – Ağustos Dönemi Yaş Gruplarına Göre İşe Yerleştirmeler (Bin Kişi)



Kaynak: İŞKUR, 2023.

Bu verilere göre 20-24 ve 25-29 yaş gruplarında işe yerleştirmelerinin diğer yaş gruplarına kıyasla oldukça yüksek olduğu görülmektedir. Genç nüfusun güncel verilerde istihdam edilenler arasındaki payının diğer yaş gruplarına göre yüksek olması ekonomiye bulunduğu katkıdan dolayı ülkemiz için oldukça kıymetlidir. Genç nüfusun istihdamı, genç nüfusun işsizlik oranı ve istihdamın zaman içindeki seyri gibi faktörlerin bir arada değerlendirilmesi daha sağlıklı bir analiz yapmamıza olanak tanıyabilir.

1.3. İŞSİZLİK KAVRAMI

Gelişmekte olan ülkelerin en büyük makroekonomik problemlerinden biri olarak kabul edilen ve birçok tanımı bulunan işsizlik, ekonomik literatürde genellikle emeğin arz ve talebi arasındaki uyumsuzluk sonucu meydana gelen bir durum olarak nitelendirilir (Zengin, 2000: 4).

İşsizlik oranı ise işsiz nüfusun toplam iş gücü içindeki oranı olarak nitelendirilir ve matematiksel gösterimi aşağıdaki gibidir (Gündoğan ve diğerleri, 2019: 10).

$$\text{İşsizlik Oranı} = \frac{\text{İşsizler}}{\text{İşgücü}} \times 100 \quad (\text{İşsizler} = \text{İşgücü} - \text{İstihdam edilenler}) \quad (1.1)$$

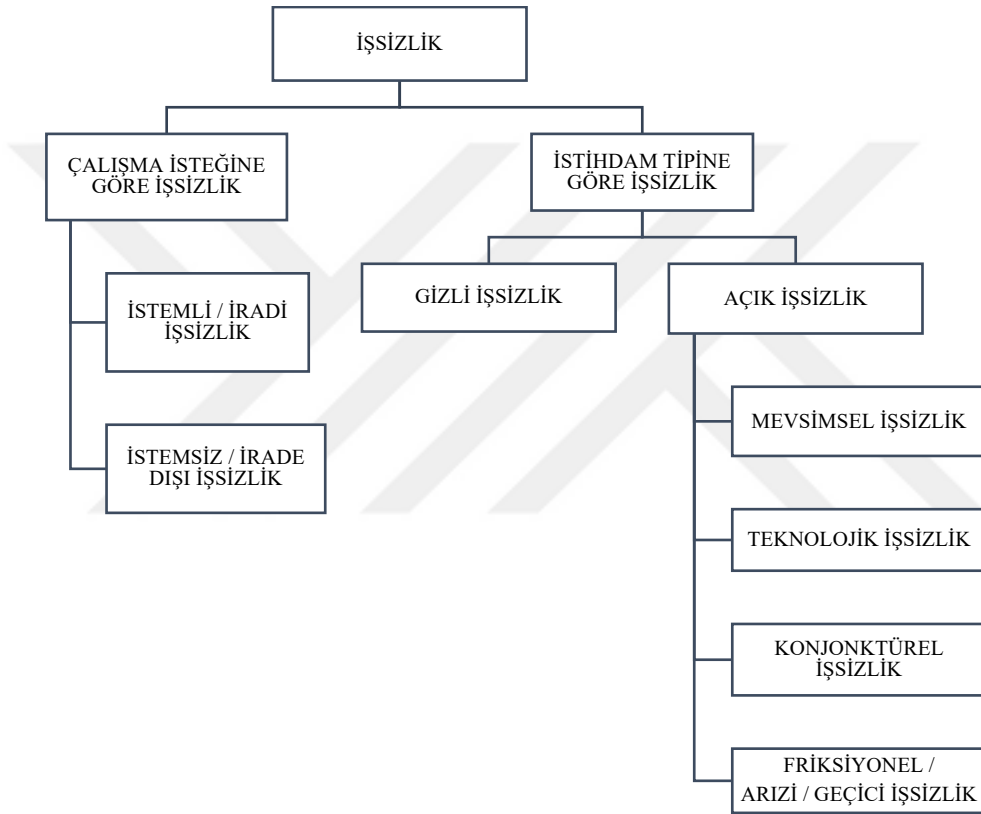
1.3.1. İşsizlik Faktörleri

İşsizliği iş arayan kişi sayısına bağlı iş bulma olasılığının düşmesi olarak düşündüğümüzde işsizlik oranlarına doğrudan etkisi olan faktörlerden birisi ülkede var olan iş fırsatı miktarı ve bu iş fırsatlarının niteliğidir. Dolayısıyla bir ülkede iş fırsatlarının yani istihdam olanaklarının kısıtlı olması söz konusuysa işsizlik oranının bu durumdan etkilenmemesi kaçınılmazdır. Ayrıca mevsimsel değişiklikler ve ekonomik dalgalanmalar da işsizlik oranlarını etkileyen önemli faktörler arasında yer almaktadır. (Gündoğan, 2022).

1.3.2. İşsizlik Çeşitleri

İşsizlik çeşitleri değişen toplum dinamiklerine bağlı çoğalmaktadır. Ancak en bilinen ve ekonomik kaynaklar tarafından sıkça değinilen işsizlik tipleri Şekil 2 'deki gibi kişilerin çalışma isteğine göre veya istihdam tipine göre sınıflandırılabilir.

Şekil 2: İşsizlik Tipleri



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

1.3.2.1. Çalışma İsteğine Göre İşsizlik

İşsizliği kişilerin çalışma isteğine bağlı sınıflandırmak mümkündür. Bu nedenle kişilerin çalışmayı isteyip istemediği durumlara bağlı gelişen işsizlik tipleri aşağıda detaylandırılmıştır.

1.3.2.1.1. İstemli/İradi İşsizlik

Belirli bir çalışma kapasitesine sahip olmasına karşın bilinçli bir şekilde işsiz olma haline istemli veya iradi işsizlik denir. Bu tip işsizlikte bireyler mevcut açık işlerin koşullarını veya işverenlerin sunduğu ücretleri hoş karşılamazlar. Sosyal güvence kaygıları ve daha iyi iş fırsatlarını arama gibi etkenler de istemli işsizliği şekillendiren faktörler arasında yer almaktadır (Kanca, 2012: 3).

1.3.2.1.2. İstemsiz/İrade Dışı İşsizlik

İstemsiz veya irade dışı işsizlik, çalışma yeteneklerine sahip bireylerin, uygun ücret ve çalışma şartlarına sahip işler arayıp bulamadıkları bir durumu ifade eder (Kanca, 2012: 3).

1.3.2.2. İstihdam Tipine Göre İşsizlik

İşsizlik türlerini istihdam tipine göre sınıflandırmak mümkündür. Açıkça kendini gösteren işsizlik tipine açık işsizlik, hali hazırda çalışma durumu içinde olan bireylerin verimlilik seviyesindeki azalışla ortaya çıkan işsizlik tipine ise gizli işsizlik denmektedir (Zengin, 2000: 5).

1.3.2.2.1. Gizli İşsizlik

Gizli işsizlik, dışarıdan bakıldığında bir işte çalışıyor gibi görünen ancak iş verimliliği belirli bir standardın altında olan bireylerin durumunu ifade etmektedir. Bu işsizlik tipinin en belirgin özellikleri bireylerin istihdam edilmesi, bu bireylerden bir kısmının işlerini yapmaması ve düşük seviyede verim göstermesidir. Fakat bu durumun yanı sıra üretim seviyesinde azalma veya artış gözlenmez. Gizli işsizliğin en çok kamu sektöründe görülmesi genellikle iktidarın aldığı kararlarla ortaya çıkan politikaların bir sonucudur. Öte yandan tarım sektöründe de sıklıkla karşılaşılabilmektedir (Zengin, 2000: 6).

1.3.2.2.2. Açık İşsizlik

Açık işsizlik, iş arayan, iş bulamayan ve iş aramaya devam eden kişilerden oluşan bir işsizlik kategorisidir. Açık işsizlik sınıfındaki kişiler, çalışmaya istekli ve işgücü piyasasına katılmış kişilerden oluşmaktadır. Ancak bu kişilerin genellikle mevcut ücret düzeylerinde uygun bir iş bulamamış gruba dahil olduğu bilinmektedir (Zaim, 1997: 170).

1.3.2.2.2.1. Mevsimsel İşsizlik

Belirli dönemlerde, doğal şartlar veya sosyal faktörler gibi etkenler, ekonomik aktivitelerin yavaşlamasına ve üretimin azalmasına yol açabilir. Bu tür durumlar, işsizliği artırabilir ve kişiler geçici olarak işsiz kalabilirler. Bu tür işsizlik, mevsimsel işsizlik olarak nitelendirilir (Unay, 1996: 350). Başka bir deyişle, bazı işler belirli mevsimlerde veya dönemlerde daha az talep görebilir ve bu nedenle o işlerde çalışan kişiler işsizlikle karşılaşabilirler.

1.3.2.2.2.2. Teknolojik İşsizlik

Üretim süreçlerinde işgücü bir diğer deyişle emek yerine otomasyon ve makinelerin kullanılması sonucu ortaya çıkan işsizlik türüne “teknolojik işsizlik” adı verilmektedir (Pekin, 1996: 108). Bu işsizlik türü, iş dünyasındaki teknolojik ilerlemelerin insan gücüne olan etkisini yansıtmaktadır ve iş piyasasındaki değişimlerin bir sonucu olarak meydana gelir.

1.3.2.2.2.3. Konjonktürel İşsizlik

Konjonktürel dalgalanma, ekonomide belirli bir dönemde ekonominin büyüme hızının veya genel ekonomik aktivitelerin dalgalanması anlamı taşımaktadır. Bu dalgalanmaların nedenleri çeşitli olabilir veya çeşitli faktörlerin etkileşimi sonucunda ortaya çıkabilir. Konjonktürel dalgalanmaların ortaya çıktığı durumlarda, işsizlik oranları, enflasyon, üretim düzeyleri gibi ekonomik göstergeler etkilenebilir. Bu

bağlamda konjonktürel işsizlik meydana gelir. Bu işsizlik tipi Keynesçi işsizlik olarak da bilinir. Bu işsizliğin en önemli nedeni efektif talep yetersizliği olarak nitelendirilir (Eyüboğlu, 2003: 15).

1.3.2.2.4. Friksiyonel/Arızı/Geçici İşsizlik

Literatürde arızı, friksiyonel veya geçici gibi isimlerle nitelendirilen bu işsizlik tipi kişilerin hür iradesine veya zorunlu nedenlere bağlı gelişebilir. Ancak çoğunlukla geçici nedenlerle ortaya çıkan bu tip işsizliğin ekonominin genelini etkileme gibi bir durumu söz konusu değildir. İşgücü piyasasında etkili olmayan düzenlemeler, bilgi seviyesinin düşüklüğü, işgücünün yer değiştirebilme yeteneğinin zayıflığı, üretim kaynaklarına rahatlıkla erişmedeki sorunlar, daha iyi koşullarda iş bulabilmek adına hali hazırda çalışılan işi bırakma gibi nedenler örnek gösterilebilir (Kanca, 2012: 3). Emekli olan iş gücü kategorisinde yer alan bireylerin yerine yeni işgücünün geçişinin oldukça zaman alışı, eğitim programlarına katılan işgücü, işgücünün artan vasıf ve yetenekleri doğrultusunda iş arayışındaki artış, özelleştirmeler, işsizlik ödeneklerindeki pasif önlemlerin miktarı ve bu önlemlerin süresindeki artışlar bu işsizlik tipinin artış nedenleri olabilir (Aydın, 2012: 123).

1.3.3. İşsizlik Sorununun Nedenleri

İşsizliğin kökenleri, temel olarak işgücü arzı ve işgücü talebi perspektiflerinde değerlendirilebilir. Bu nedenle işgücü arzı ve talebi kaynaklı nedenler bu başlıkta detaylandırılmıştır.

1.3.3.1. İşgücü Arzı Kaynaklı Nedenler

İşsizlik, geçmişten günümüze birçok çalışmaya konu olmuş ve toplumların gelişmişlik düzeyinde büyük rol oynayan karmaşık bir problemdir. İşsizliğin nedenleri arasında işgücü arzı kaynaklı etkenler de bulunmaktadır. Toplumların ekonomik faaliyetlere katılıma hazır bireylerinin toplamı olarak ifade edilen işgücü arzı seviyesinin, işgücü talebi seviyesi ile uyum sağlayarak artması, ekonomik büyümeyi

sağlar ve istihdama olanak tanırken, çeşitli sebeplerin bir sonucu olarak işgücü talebi seviyesinden daha altta kalan işgücü arzı, işsizlik oranlarının artmasına yol açabilir (Tansel, 2012: 50). Bu nedenle, işgücü piyasasının işgücü talebi ve işgücü arzını dengeye getirmesi işsizliği önlemek için büyük önem taşımaktadır. İşgücü arzını etkileyen en önemli etkenler; ülkenin nüfusu, bu nüfusun yaş bakımından yapısı, eğitim seviyesi, işgücü, işgücüne katılma oranı ve göç hareketleridir. Bu etkenler aşağıda detaylandırılmıştır.

Nüfusun Büyüklüğü ve Yapısı: Tüm toplumlarda genel nüfusun büyüklük seviyesi işgücü arzını etkilemektedir. Genç nüfusu genel nüfusa oranla daha yüksek olan ülkelerin işgücü arzı daha yüksek olabilir. Ancak bu genç ve dinamik nüfusun işgücü piyasasına dahil olabilmesi için genç nüfusun yapısına bağlı uygun iş fırsatlarının olması gerekmektedir.

Eğitim Seviyesi: Eğitim seviyesi işgücü piyasasında doğrudan bireylerin rekabet edebilme yeteneklerini etkilemektedir. Eğitim seviyesi yüksek bireylerin daha nitelikli ve kapsamlı işlerde çalışma potansiyeline sahip olduğu görülmüştür. Ancak, eğitim seviyesinin işgücü talebiyle uyumlu olmaması durumu nitelikli işsizlik sorununun meydana çıkmasına sebebiyet verebilir.

İşgücüne Katılma Oranı: Ekonomik faaliyetlere katılmaya hazır ve istekli bireylerin toplam nüfusa oranı olarak ifade edilen işgücüne katılma oranının yüksek olması, işgücü arzının da yüksek olduğunu göstermektedir.

Göç: Ekonomik nedenlerle gerçekleşen göçler ani değişiklikler yaratarak işgücü arzını doğrudan etkiler.

Sonuç olarak, işgücü arzına dayalı etkenler, işsizlik oranlarını etkileyen temel unsurlardandır. Ekonomik çalışmaların işgücü arzı göz önünde bulundurularak gerçekleştirilmesi, sürdürülebilir ekonomik büyüme ve toplumsal refah için son derece kritik bir öneme sahiptir.

1.3.3.2. İşgücü Talebi Kaynaklı Nedenler

İşsizlikte iş gücü talebi kaynaklı nedenler, işgücü arzı kaynaklı nedenler gibi ekonomik literatürde geniş bir yere sahiptir. İşgücü talebinin yetersiz olması, işsizlik oranlarının artışına neden olabilir. İşgücü talebini etkileyen en önemli faktörler; bir

ülkenin mevcut ekonomik yapısı, küresel konumu, gelişmişlik düzeyi, teknolojik seviyesi, işveren açısından işgücü maliyetleri, verimlilik seviyesi ve istihdamla sonuçlanmayan ekonomik büyüme oranı gibi maddelerden oluşur (Akkaya, 2017: 13). Artan işsizlik oranının anlaşılabilmesi ve çözümlenebilmesi için söz konusu bu maddeler aşağıda detaylandırılmıştır.

Ülke ekonomisinin mevcut yapısı: İşgücü piyasasını etkileyen faktörlerden en önemlisi ülke ekonomisinin yapısıdır. Örneğin sektörler arası dağılımlar (tarım, sanayi ve hizmet) işgücü talebini oldukça etkilemektedir. Gelişmekte olan toplumlarda tarım ve hizmet sektöründe yüksek oranda işgücü talebi görülebilir.

Küresel konum: Küresel konum ülkelerin uluslararası ilişkileri bakımından önem taşımaktadır. Bu ilişkilerin bir sonucu olarak işgücü talebinde artış veya azalış gibi durumlar gözlemlenebilir.

İstihdam olmadan ekonomik büyüme: Ekonomik büyümenin söz konusu olduğu toplumlarda işsizlik oranının azalabileceği söylenebilir fakat bu büyüme sayesinde ülkenin istihdam oranı artmıyorsa buna istihdam olmadan gerçekleşen ekonomik büyüme denir. İstihdam oranında artış olarak sonuçlanmayan bu büyüme işgücü talebini etkilemektedir.

Gelişmişlik düzeyi: İşgücü talebini etkileyen kaçınılmaz unsurlardan biri ülkelerin gelişmişlik düzeyidir. Gelişmişlik düzeyi arttıkça vasıflı ve nitelikli işgücüne olan ihtiyaç artmaktadır.

Ülkenin teknolojiye geldiği nokta: Teknolojik seviyenin artmasıyla beraber nitelikli işgücüne olan ihtiyaç ve talep artar. Dolayısıyla bu durumla orantılı olarak teknolojik seviyesi düştükçe daha az nitelikli ve daha az vasıflı işgücüne ihtiyaç artabilir.

İşveren açısından işgücü maliyetleri: İşveren kategorisindeki bireyler için en önemli unsur işgücü maliyetidir. Bu durum işgücü talebini oldukça etkilemektedir. İşveren için yüksek maliyet işgücü talebini azaltabilirken düşük maliyet işgücü talebinin artmasına neden olabilir.

Verimlilik seviyesi: Yüksek verim elde eden kuruluşlar daha fazla üretim elde eder ve böylelikle daha fazla çalışana ihtiyaç duyar. Bu durumda işgücü talebinin artmasından söz edilebilir.

1.4. DÜNYADA İŞSİZLİK VE EVRİMİ

İşsizlik, dünya genelinde araştırılan ve çözümü bulunmaya çalışılan bir sorundur. Artan nüfus, değişen teknolojik şartlar, ekonomik krizler ve savaşlar gibi faktörlerle birlikte büyüyen işsizlik problemi, sanayi devrimine kadar uzanan bir sürecin sonucudur. Geçmişten günümüze küresel çerçevede işsizliği etkileyen önemli hareketler aşağıda detaylandırılmıştır.

Sanayi Devrimi: Dünya tarihinde büyük etkiler yaratan sanayi devrimleri, işgücü piyasalarını da büyük oranda etkilemiştir. Her bir devrim, yapısı ve etkisi bakımından diğer devrimlere göre değişkenlik gösterse de bu devrimlerin en büyük ve ortak özelliği giderek kol gücüne olan ihtiyacın azalmasına neden olmalarıdır.

Birinci Sanayi Devrimi 18. yüzyılın sonlarında buhar gücünün fabrikalarda kullanılmasıyla ortaya çıktı. Bu devrim makineleşmenin artışına neden oldu. Bu artışla beraber büyük fabrikalarda üretim artarken küçük atölyeler kapanmak durumunda kaldı. Ardından 19. yüzyılın ortalarına doğru elektrik üretim faaliyetlerinde kullanılmaya başlandı ve böylece ikinci Sanayi Devrimi başlamış oldu. Elektriğin kullanılmasıyla birlikte seri üretim çeşidi yaygınlaştı ve şehirler hızlı bir ivmeyle büyümeye başladı. 20. yüzyılın ortalarında elektrikli hesap makineleri icat edildi bu yenilikle birlikte üçüncü Sanayi Devrimi başladı. Bu devrimin sonucunda ise kol gücü giderek yerini makinelere, otomasyona bıraktı. Bu dönemde dünya kaynakları hızla tükenmeye başladı. Son olarak günümüzde yani 21. yüzyılın ortalarında dünya teknolojik anlamda seviye atladı. Bu dönemde internet kullanımı zirveye ulaştı. Dijitalleşme ve akıllı makineler dünyaya hâkim oldu. Bu döneme Endüstri 4.0 da denilmektedir. Kol gücüne ihtiyaç yok denecek kadar azaldı. Bu devrimle birlikte yeni iş kolları ortaya çıktı. Ancak tıpkı diğer sanayi devrimlerindeki gibi nitelikli işgücüne olan ihtiyaç arttı ve birçok kişi de işsizlikle yüzleşti (Taş, 2018: 1820).

Savaşlar ve Büyük Buhran: Dünya tarihindeki birçok savaş hem ulusal hem de uluslararası düzeyde önemli sonuçlar doğurmuştur. Ancak, bu savaşlar içinde en büyük yankı uyandıran ve dünya tarihine büyük etkiler bırakanlardan biri Birinci Dünya Savaşı'dır. Bu savaş, başlangıçta Avrupa'da başlamış, ancak hızla küresel boyutta etkisini göstermiştir. Milyonlarca insan hayatını kaybetmiş ve birçok kişi işsiz kalmıştır. Üretim ve ticaret durgunluğa girmiş, dünya ekonomisi ciddi bir sarsıntı

geçirmiştir. Bu dönemin ardından, 1929 'da tarihteki en büyük ekonomik çöküş olan Büyük Buhran meydana gelmiştir. Para birimlerinin değer kaybı ve üretimdeki gerileme sonucunda her dört kişiden biri işsizlikle mücadele etmiştir. İkinci Dünya Savaşı da buhranın hemen ardından gelmiş ve dünya çapında işsizlik oranlarını daha da artırmıştır.

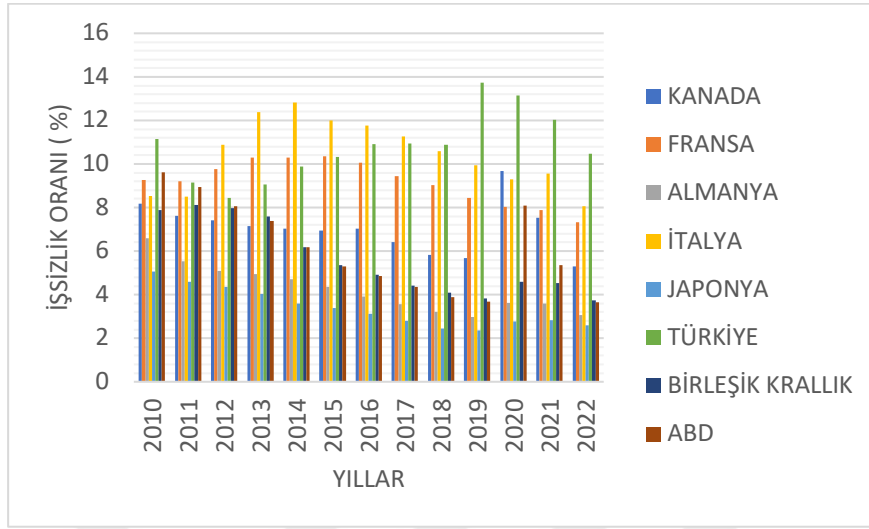
2008 Küresel Finans Krizi: Küresel finans krizi, 2007 'den itibaren tüm dünyayı ekonomik açıdan olumsuz anlamda etkileyen bir krizdi. 2008 yılında Lehman Brothers 'ın iflas etmesi ekonomik anlamda büyük çapta bir krize neden oldu ve finansal piyasaları sarstı. Hükümetler ve merkez bankaları ekonomik büyümeyi desteklemek için büyük mali yardımlar sağlasa bile mali dengelerin bozulması kaçınılmazdı. Ekonomik büyüme düştü, güvensizlik ve belirsizlik had safhaya ulaştı. Bu süreçte işsizlik oranları tüm dünyada tarihi seviyelerdeydi ve bu durumun ekonomilerde yarattığı olumsuzlukların iyileşmeye başlaması epey zaman aldı (Otker-Robe ve Podpiera, 2013: 2).

COVID-19: Küresel bir salgın olarak başlayan COVID-19 Salgını, 2020 'de başlayarak 2022 'ye kadar devam etti. Bu süre zarfında pandemi dünya genelinde iş piyasalarında büyük olumsuzluklar yarattı. Bu etki ile para, döviz kuru, maliye ve finans politikaları gibi unsurlar büyük değişikliklere uğradı. İşsizlik seviyesi ve ekonomik belirsizlik artış gösterdi. Birçok hükümet, işsizliği azaltma amacıyla teşvikler ve para politikaları gibi önemli araçlarını devreye soktu.

Kökeni, etkileri ve yayılma yolları açısından diğer küresel krizlerle karşılaştırıldığında tamamen farklı özelliklere sahip olan COVID-19 salgınına yönelik tedbirlerle birlikte kısıtlamalar kademeli olarak kaldırılmaya başlandı ve ekonomiler toparlanma sürecine girdi. Ancak Uluslararası Çalışma Örgütü (ILO), COVID-19 salgınının yarattığı işsizlik boyutunu şu şekilde açıklıyor; "2022 'nin üçüncü çeyreğinde, küresel çalışma saatleri hala 2019 öncesi döneme göre %1,5 daha düşüktü, bu da yaklaşık 40 milyon tam zamanlı iş kaybına işaret ediyor." (ILO, 2023: 3).

Uluslararası Para Fonu (IMF) tarafından yayınlanan Dünya Ekonomik Görünümü raporuna göre (IMF, 2022: 104) büyük gelişmiş ekonomiler; Kanada, Japonya, Fransa, Almanya, İtalya, Birleşik Krallık ve Amerika Birleşik Devletleri 'dir. Bu ülkeler ve Türkiye için OECD 'nin yayınladığı en güncel işsizlik oranları Şekil 3 'deki gibidir.

Şekil 3: 2010-2022 Dönemi Büyük Gelişmiş Ülkeler ve Türkiye ’deki İşsizlik Oranları (%)



Kaynak: OECD, 2023.

Şekil 3 'e göre 2010 yılından 2022 yılına kadar ülkelerin işsizlik oranlarında büyük değişimler olmuştur. Ancak zaman içinde Türkiye 'nin işsizlik oranları gelişmiş ekonomilerin işsizlik oranlarına göre giderek yükselmiştir.

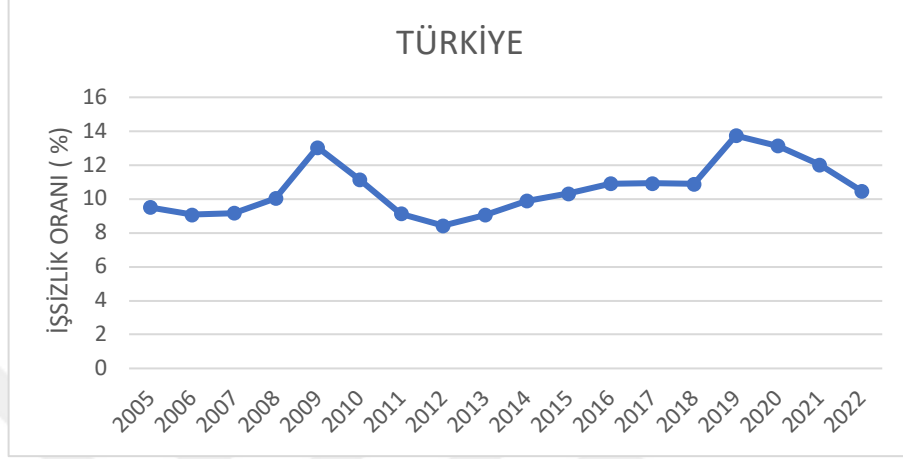
1.5. TÜRKİYE 'DE İŞSİZLİK VE EVRİMİ

Türkiye tarihinde 1960 'lı yıllardan itibaren başlayan ve 1980 'lerden sonra gelişen teknoloji ve küreselleşme etkisiyle işsizlikte önemli artışlar yaşanmıştır. 1990 'lı yıllardan itibaren uluslararası işsizlik oranı standartlarını aşarak sürekli büyümüştür. Özellikle 2001 'deki ulusal ekonomik kriz, işgücü piyasalarını olumsuz yönde etkileyerek toparlanma sürecinin yavaş ilerlemesine sebep olmuştur. İşgücü arzı ve talebiyle ilgili faktörler, işsizlik sorununu giderek karmaşık hale getirmiştir, bu da Türkiye için çözümünün daha da zorlaştığı bir problemi ortaya koymaktadır (Ay, 2012: 322).

OECD verileri ile oluşturulan Şekil 4 'deki grafiğe göre 2005 – 2022 yılları arasında Türkiye 'de işsizlik oranının en düşük olduğu dönem %8,43 ile 2012 'dir. En yüksek olduğu dönem ise %13,74 ile 2019 'dur. 2018 'de daha düşük seyrinde ilerleyen

işsizlik oranının 2019 'da sıçrama yapması COVID-19 ile ilişkili olabilirken 2020 ile bu oran düşüşe geçmiştir.

Şekil 4: 2005 – 2022 Dönemi Türkiye İşsizlik Oranları (%)



Kaynak: OECD, 2023.

1.6. İŞSİZLİĞİN TAHMİNLENMESİNİN ÖNEMİ

Hızla değişen ekonomik koşulların yarattığı belirsizlik, doğal afetler, küresel krizler ve pandemiler işsizlik oranını etkileyen, dolayısıyla tahminini zorlaştıran unsurlardır. Öte yandan işsizliğin birçok faktörle meydana gelmesi, güvenilir veya güncel veriye erişimin kısıtlılığı ve yanıltıcı istatistikler de öngörümlemeyi büyük oranda zorlaştırmaktadır. Ancak bu sorunlar ortadan kaldırıldığında işsizlik oranının tahminlenmesi hükümetlerin ekonomik planlamalarını doğru bir biçimde yapabilmesi açısından oldukça önem taşımaktadır. Gerçekleştirilen tahminleme sonucu işsizlik oranları yüksek ise ekonomik karar süreçlerinde çeşitli önlemler alınabilir veya yeni politikalar geliştirilebilir. Ayrıca işsizlik oranlarının tahmini sayesinde işgücü piyasalarındaki dinamikler yakından takip edilebilir ve gelecekteki eğilimler öngörülebilir. Bu nedenle işsizlik oranlarının tahmini karmaşık bir sürece dayalı olsa da büyük önem taşımaktadır.

İKİNCİ BÖLÜM

MAKİNE ÖĞRENMESİ ÇEŞİTLERİ VE GOOGLE TRENDS

Bu bölümde makine öğrenmesinin kavramsal tanımı, diğer disiplinlerle olan ilişkisi ve önemi ayrıntılı olarak açıklanmıştır. Makine öğrenmesinin ana kategorileri ele alınarak en sık tercih edilen regresyon tabanlı yöntemler detaylı bir biçimde aktarılmıştır. Ek olarak çalışma kapsamında yer alan Google Trends arayüzünün sunduğu imkanlara ve kullanım alanlarına değinilmiştir.

2.1. MAKİNE ÖĞRENMESİ

Teknolojideki ilerleme ile büyük veriler saklanıp işlenebilir bir seviyeye gelmiştir. Ancak saklı tutulan bu veriler yalnızca detaylı bir biçimde analiz edildiğinde değerli bilgiler sunabilir. Bazı uygulamalarda, araştırmacılar erişebildikleri bu kapsamlı verilere rağmen doğru algoritmayı bulmakta zorluk çekebilmektedir. Çünkü bu veriler işlendikleri modellerin yüksek bir tahmin doğruluğuna sahip olmasıyla anlam kazanmaktadır. Yalnızca veri madenciliğiyle ilişkili olmayan ve yapay zekanın çok değerli bir parçası olarak nitelendirilen makine öğrenmesi, bir sistemi programlamadan öğrenme yeteneği ve çeşitli algoritmaları ile bilgiye ulaşmayı kolaylaştırmaktadır. Makine öğrenmesi; bilgisayarları, saklı tutulan verileri veya geçmişteki deneyimleri kullanarak performans ölçütlerini optimize etmek için, öğrenerek programlayan teknikler bütünüdür. Makine öğrenmesi uygulamalarında üretilen modeller, veri kümesi ve analiz amacına bağlı olarak tahminleme gerçekleştirmek veya bilgi üretmek için kullanılmakta ya da aynı anda bu iki faydayı sağlamayı hedeflemektedir (Alpaydın, 2010: 2-3).

2.1.1. Makine Öğrenmesinin Diğer Disiplinlerle İlişkisi

Makine öğrenmesi, bir veri seti üzerinden mevcut dinamikleri öğrenerek gelecekte karşılaşılabilecek muhtemel veri setini ve yapıyı tahminleyebilme yeteneği sayesinde oldukça geniş uygulama alanlarına sahiptir. Bu çeşitliliğin getirdiği bir durum olarak yolu birçok disiplinle kesişmektedir. Matematiksel kavramlarını ve

olasılık temelini istatistik biliminden alan makine öğrenmesi, bu temeli bilgisayar bilimi ile algoritmaya dönüştürüp çalıştırmaktadır. Mühendislik bilimi ile otomasyon çalışmalarına katkıda bulunmakta; tıp ve biyoloji bilimlerinde hastalıklarla ilgili olabilecek çeşitli çalışmalarda güçlü tahminlerde bulunup canlı vücudu hakkında perspektifler sunmaktadır. Ekonomi ve finans bilimlerinde toplumların işgücü piyasalarındaki problemlerini netleştirerek araştırmacıların çözüm üretmesini kolaylaştırmaktadır. Elbette bu örneklerle kısıtlı olmayan makine öğrenmesinin uygulama alanlarının kapsamı, devamlı gelişen teknolojik koşullar ve veri kullanımının yaygınlaşmasıyla daha fazla çeşitliliği meydana getirecektir.

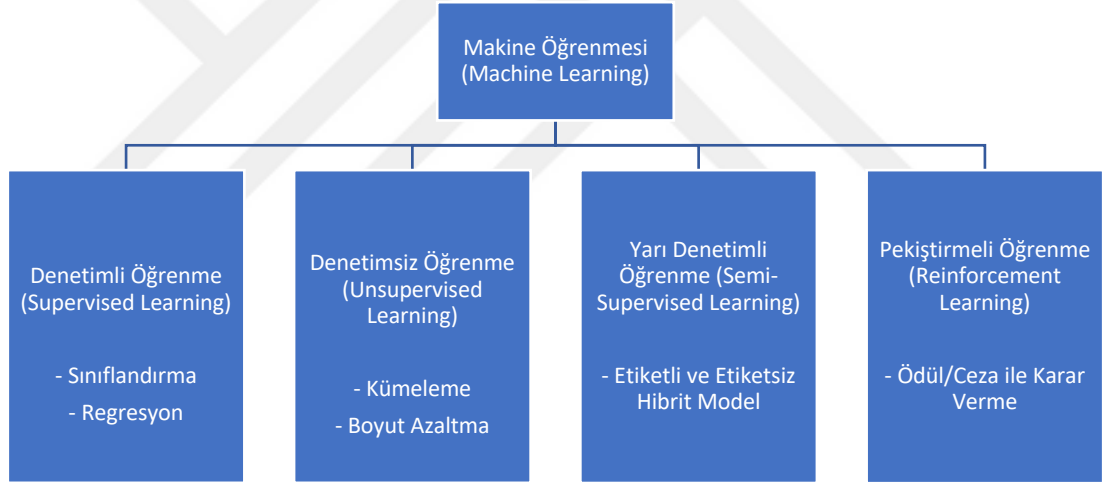
2.1.2. Makine Öğrenmesinin Önemi

1990 'ların başlarında bilimsel çalışmalarda kullanımı artan makine öğrenmesi, geçmişten günümüze gelişen teknolojik koşullarla birlikte, donanım verimliliğinin artışı ve büyük verilerin ulaşılabilir olması ile yapay zekâ alanında önemli bir dal haline gelmiştir. Makine öğrenmesinin önemini anlamak için geleneksel yöntemlerden farklılaşan yönlerini değerlendirmek gerekmektedir. Makine öğrenmesinin klasik istatistikten ayrıştığı en önemli nokta ise karmaşık milyonlarca veri ile çalışabilme yeteneğidir (Chollet, 2021: 4). Bu durum, büyük verileri analiz edebilme ve bu verilerden anlamlı örüntüler çıkarabilme özelliğini ortaya koymaktadır. Geleneksel istatistiki yöntemler bu tür büyüklükte verilerle çalışma noktasında zorlanabilirken, makine öğrenmesinde genellikle bu sorun yaşanmaz. Öte yandan makine öğrenmesi klasik yöntemlere göre değişken ortamlara daha çok uyum sağlayarak yüksek performans sergileyebilme kapasitesine sahiptir (Geron, 2019: 5). Sonuç olarak, bu özelliklere sahip olması sayesinde makine öğrenme teknolojisi büyük veri kümelerinden değerli bilgileri çıkarmak, mevcut veya gelecekte ortaya çıkma olasılığı olan karmaşık problemleri çözüme kavuşturmak için oldukça değerlidir.

2.2. MAKİNE ÖĞRENMESİ ÇEŞİTLERİ

Makine öğrenmesi çeşitleri genellikle Şekil 5 'deki gibi denetimli, denetimsiz, yarı denetimli ve pekiştirmeli öğrenme olarak dört ana kategoriye ayrılmaktadır. Ancak bu ana kategorilerin haricinde takviyeli öğrenme ve diğer özel makine öğrenmesi türleri de mevcuttur. Bu alt kategoriler, makine öğrenmesinin özelleştirilmiş farklı uygulamalarına odaklanan tekniklerden meydana gelmektedir. Bu bölümde makine öğrenmesi çalışmalarında en sık tercih edilen ana kategoriler detaylandırılarak açıklanacaktır.

Şekil 5: Makine Öğrenmesi Çeşitleri



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

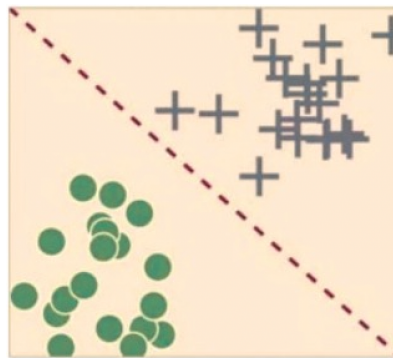
2.2.1. Denetimli Öğrenme

Makine öğrenmesinin ana kategorilerinin başında gelen denetimli öğrenmeyi anlayabilmek için öncelikle tahminleyici modelin detaylandırılması gerekir. Tahminleyici model, genel tanımıyla bir değeri tahminleyebilmek için ilgili veri kümesinin içindeki diğer verilerden yararlanarak oluşturulan modellerdir. Bu

modellere öğrenmeleri gereken işlem ve nasıl öğrenmesi gerektiği konusunda birtakım komutlar verilir. Modelin bu eğitim süreci ise denetimli öğrenme olarak nitelendirilir. Etiketli veri kümeleri ile çalışan denetimli öğrenme algoritmaları, girdi olarak verilen veri topluluklarını kullanarak hedef çıktıya neden olabilecek özellikleri bulmayı amaçlar ve oluşturulan modeli optimize etmeye yani modeli daha etkin bir yapıya getirmeye çalışırlar (Lantz, 2019: 21). Denetimli öğrenme algoritmaları Şekil 6 ve Şekil 7 'deki gibi çözüm ürettikleri problemler açısından genellikle hedef değişkenin tipine bağlı olarak sınıflandırma ve regresyon olarak ikiye ayrılmaktadır. Tercih edilen algoritmaların bazıları yalnızca regresyon problemlerine ya da yalnızca sınıflandırma problemlerine çözüm üretebilir, ancak bazı algoritmalar her iki tür problemin çözümünde de kullanılabilir.

Sınıflandırma tabanlı denetimli öğrenmede hedef değişkeni kategorik yapıdadır. Bu açıdan ayrık değerler ile bir örneğin içinde bulunduğu kategoriyi tahminlemek için kullanılmaktadırlar. Sınıflandırma problemlerine en sık verilen “e-posta kullanımında gelen mesajın spam olup olmadığının tespiti” örneğindeki gibi tahminlenmesi hedeflenen bir özellik vardır ve sınıf olarak nitelendirilebilen, kendi içinde seviyelere ayrılabilen kategorik özellikler söz konusudur (Lantz, 2019: 21). Dolandırıcılık tespiti ya da müşteri bağlılığını ölçen uygulamalar da sınıflandırma problemlerine örnek olarak verilebilir (Khanzode, 2015: 14).

Şekil 6: Sınıflandırma Problemleri İçin Temel Gösterim



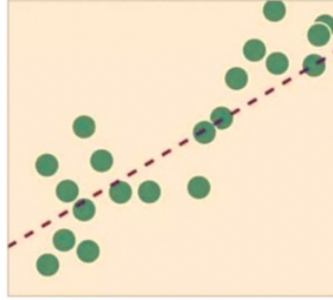
Kaynak: Khanzode, 2015: 14.

Denetimli öğrenme kapsamında sınıflandırma problemlerinde en sık kullanılan algoritmalar aşağıda listelenmiştir.

- Lojistik Regresyon
- K-En Yakın Komşu
- Destek Vektör Makineleri
- Naive Bayes
- Karar Ağacı Sınıflandırması (DTC)
- Rastgele Orman (RF)
- Gradyan Artırma (Gradient Boosting)
- Yapay Sinir Ağları (ANN)

Regresyon problemlerinde ise hedef değişken sürekli ve sayısal bir yapıdadır. Bağımlı ve bağımsız değişkenlerle bu değişkenler arasındaki neden sonuç ilişkilerine dayalı bir yapı söz konusudur. Bu nedenle bu yapıdaki uygulamalarda regresyon tabanlı denetimli öğrenme teknikleri ile sayısal veriler tahminlenir. Bu tahminler sayesinde değişkenler arasındaki ilişkinin büyüklüğü ve belirsizliği gözlemlenebilir (Lantz, 2019: 21). Regresyon tabanlı yöntemlerin örnek uygulama alanları, süreç optimizasyonu veya daha önce tespit edilmemiş yeni bilgiler elde etme konularını içerebilir (Khanzode, 2015: 14).

Şekil 7: Regresyon Problemleri İçin Temel Gösterim



Kaynak: Khanzode, 2015: 14.

Denetimli öğrenme kapsamında regresyon problemlerinde en sık kullanılan algoritmalar aşağıda listelenmiştir.

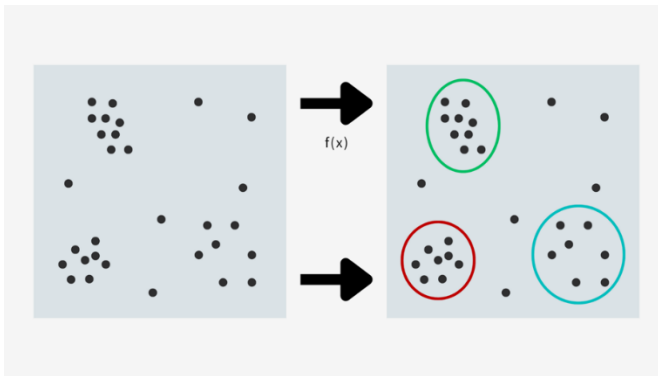
- Doğrusal Regresyon
- Lojistik Regresyon

- Destek Vektör Regresyonu
- Karar Ağacı Regresyonu (DTR)
- Gradyan Artırma Karar Ağacı (GBDT)
- Rastgele Orman (RF)
- ANN
- Ridge Regresyon
- En Küçük Mutlak Büyüklük ve Seçim Operatörü (LASSO)
- ElasticNet
- K – En Yakın Komşu Algoritması (KNN)

2.2.2. Denetimsiz Öğrenme

Makine öğrenmesinin en temel öğrenme tiplerinden bir diğeri denetimsiz öğrenmedir. Bu öğrenme tipinde, denetimli öğrenmenin aksine etiketli eğitim örnekleri söz konusu değildir. Yani hiçbir verinin bir etiketi veya sınıfı yoktur. Etiketsiz veri kümeleri ile çalışması bu öğrenme tipinin karmaşık olarak nitelendirilmesinin en büyük nedenlerinden biridir. Başlangıçta belirsiz özelliklere sahip olan veri kümesinin, sınıflanması veya kümelenmesi gerekmektedir. Bu öğrenme çeşidine verilebilecek en iyi örneklerden biri e-ticaret sitelerinde müşterilerin segmentlerinin alışveriş davranışlarına göre kümelenmesi olabilir. Kümeleme problemleri için temel gösterim Şekil 8 'de ifade edildiği gibidir.

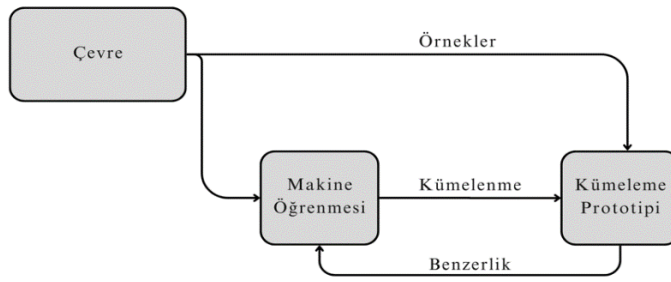
Şekil 8: Kümeleme Problemleri İçin Temel Gösterim



Kaynak: Arya, 2023.

Kümeleme problemlerinin ilk adımında veri noktası rassal bir biçimde başlatılmaktadır. Bu veri noktalarının benzerlikleri tespit edildikten sonra gruplara ayrılarak verilerin içsel özellikleri araştırılmaktadır. Bu verilerin benzerlik gösterdiği veya ayrıştığı noktaların tespit edilmesi amaçlanmakta ve son aşamada ise sınıflar ortaya çıkarılarak veriler bu sınıflara göre kümelenmektedir. Şekil 9 'da küme örnekleri ile veri ögeleri arasındaki benzerliğin maksimum noktaya getirilmesini hedefleyen denetimsiz öğrenme süreci aktarılmıştır. Şekle göre süreç içerisinde küme sayısının belirlenip küme örneklerinin rassal olarak başlatıldığı ve her bir veri noktasının, en benzer olduğu kümelere göre düzenlendiği bu döngü açıklanmıştır.

Şekil 9: Denetimsiz Öğrenme Süreci



Kaynak: Jo, 2021: 13.

En sık tercih edilen denetimsiz öğrenme algoritmalarından birkaçı aşağıda listelenmiştir.

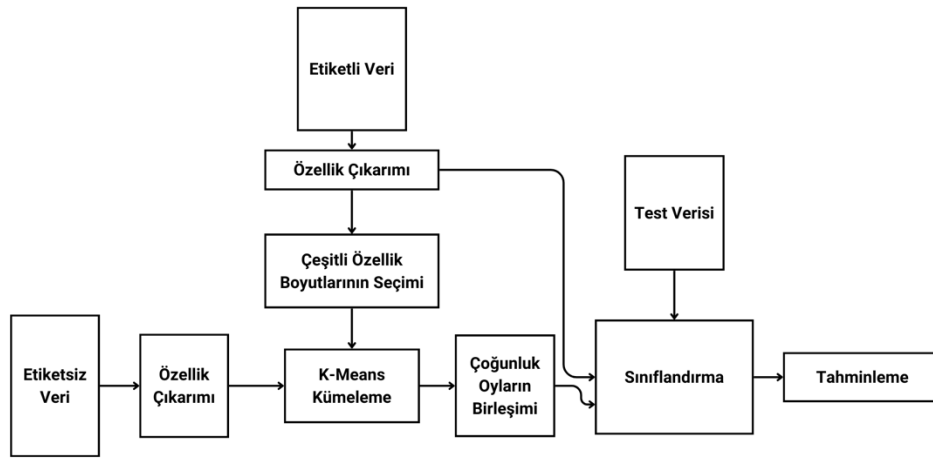
- K-Means
- Apriori Algoritması
- Gizil Dirichlet Dağılımı (LDA)

2.2.3. Yarı Denetimli Öğrenme

Denetimli ve denetimsiz öğrenmenin bir karışımı olan yarı denetimli öğrenme ne tamamen etiketli ne de tamamen etiketlenmemiş verilere sahiptir. Yani yarı denetimli öğrenme çeşidinde veriler hem etiketlenmiş hem de etiketlenmemiştir.

Denetimli öğrenme yöntemleri ile aynı amaçlara sahip olsa da etiketlenmemiş verilerle daha iyi bir modele ulaşmayı hedeflemektedir. Çok sayıda veri ile çalışılan ve maliyeti yüksek olan bu öğrenme türünde genellikle etiketli olmayan veri sayısı daha fazladır. Bu nedenle belirsiz sonuçlara neden olacağı düşünülmektedir. Ancak daha büyük örnek çoğunlukla daha fazla bilgi anlamını taşımaktadır (Burkov, 2019: 2). Yarı denetimli öğrenme süreci Şekil 10 'da özetlenmiştir.

Şekil 10: Yarı Denetimli Öğrenme Süreci

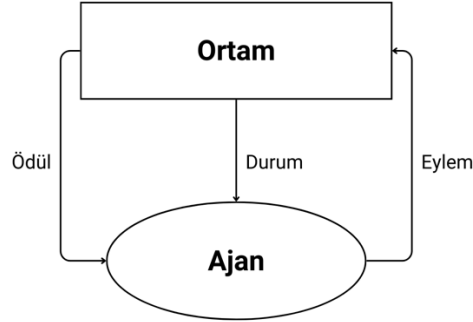


Kaynak: Gupta, 2023.

2.2.4. Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme, ödül/ceza sistemine dayanır. Bu sistemde bir ajan vardır ve bu ajan belirli bir ortamda çeşitli eylemler gerçekleştirerek, bu eylemlerin bir sonucu olarak ödüle veya cezaya ulaşır. Tüm bu ödül/ceza sistemi ile en sık ödülü elde ettiği en iyi eylem adımlarını tespit etmek hedeflenmektedir. Yapısı itibariyle, problemi çözüme kavuşturmadan ziyade gerçekleştirilen eylem adımlarının etki seviyesini değerlendirmek önemlidir. Gelişmiş sonuçlara ulaşmak amacıyla eylemler dizilerine sahip olan bu ortam, genellikle Markov karar süreçleri kullanılarak modellenmektedir (Alpaydın, 2010: 447-449). Pekiştirmeli öğrenme çeşidinin temel işleyişi Şekil 11 'de ifade edilmiştir.

Şekil 11: Pekiştirmeli Öğrenme Süreci



Kaynak: Alpaydın, 2010: 448.

2.3. REGRESYON TABANLI DENETİMLİ MAKİNE ÖĞRENMESİ TÜRLERİ

Denetimli makine öğrenmesinde analiz edilecek veri kümelerine ve problemlere bağlı olarak birçok algorithmadan yararlanılabilir fakat problem çözümünde en temel adımlardan birisi tercih edilecek algoritmanın amaca yönelik olmasıdır. Bu nedenle bu başlıkta çalışma kapsamına uygun bir biçimde en sık tercih edilen regresyon tabanlı denetimli makine öğrenmesi tekniklerine yer verilmiştir.

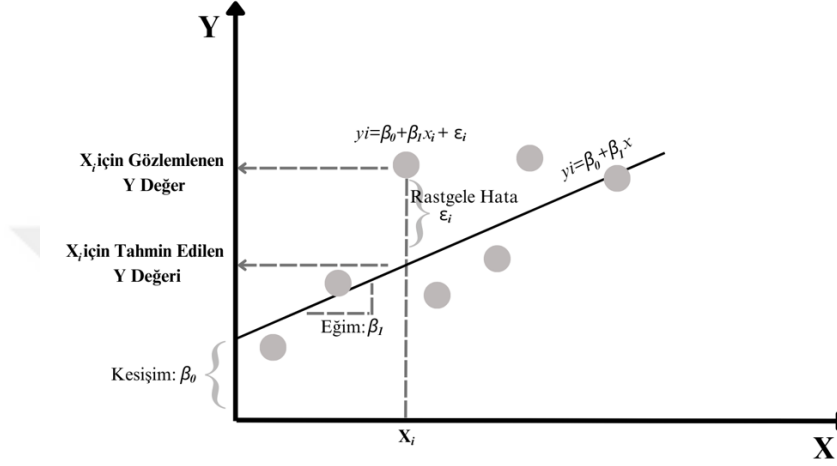
2.3.1. Doğrusal Regresyon

Doğrusal regresyon bir veya daha fazla bağımsız değişkenin bağımlı değişkenle olan ilişkisini açıklamak amacıyla çok sık tercih edilen bir algoritmadır. Bu algorithmada ilişkinin doğrusal olduğu varsayılmaktadır ve en temel amaçlarından biri veri kümesine en uygun çizgiyi tespit etmektir. Ancak veriler arası ilişkiler çok karmaşık olabilir. Bağımlı değişken ve bağımsız değişken arasındaki ilişkiler tahminlenirken hata karelerinin toplamını minimize etmek hedeflenmektedir. Böylelikle tahminlenen çizgi mümkün olduğu kadar veri noktalarına yaklaşmaktadır. Bu noktada model, verileri en iyi açıklayacağı seviyeye gelecektir.

Temelde bir bağımsız değişkenle ilişkilendirilen basit doğrusal regresyon modelinin matematiksel gösterimi ve şekil üzerinde gösterimi aşağıda detaylandırılmıştır.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \varepsilon \quad (2.1)$$

Şekil 12: Doğrusal Regresyon Modeli



Kaynak: Mali, 2021.

Birden fazla bağımsız değişkenin modele eklenmesini gerektiren karmaşık problemlerde bu temel fonksiyon çoklu doğrusal regresyon modeli haline gelir ve matematiksel gösterimi aşağıdaki gibidir:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i + \varepsilon \quad (2.2)$$

2.3.2. Ridge, LASSO ve ElasticNet Regresyon

2.3.2.1. Ridge Regresyonu

Doğrusal regresyonun düzenlenleştirilmiş hali olarak bilinen Ridge regresyon (veya Tikhonov düzenlenleştirilmesi), aşırı uyumla (overfitting) mücadele etmek ve hata kareler toplamını minimize etmek için genellikle çoklu korelasyona sahip özelliklerin bulunduğu durumlarda tercih edilen, katsayıları sıfıra yakın olacak şekilde düzenleyen bir regresyon çeşididir. Bu regresyon modelindeki en önemli unsur maliyet

fonksiyonuna dahil edilen ceza terimidir. Bu ceza terimi sayesinde model ağırlıkları olabildiğince küçük tutulur ve modelin düzenlileştirme seviyesini kontrol etmek için bir hiperparametre (λ) kullanılmaktadır (Geron, 2019:135). L_2 düzenlileştirilmesi olarak da ifade edilen bu yapının matematiksel gösterimi aşağıdaki gibidir:

$$HKT_{L_2} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (2.3)$$

2.3.2.2. LASSO Regresyonu

Doğrusal regresyonun düzenlileştirilmiş olarak ifade edilen farklı çeşitlerinden bir diğeri LASSO regresyonudur. Tıpkı Ridge regresyonu gibi bu regresyon modelinde de hata kareleri toplamını minimize etmek amacı söz konusudur. Bu nedenle maliyet fonksiyonuna $\lambda \sum_{j=1}^p |\beta_j|$ ceza terimi eklenir. Ridge regresyonunda katsayıların 0 'a yaklaştırılması söz konusuyken LASSO regresyonunda katsayıların sıfır olarak belirlenmesi bu modelleri birbirinden ayıran tek özelliştir. L_1 düzenlileştirme olarak da bilinen bu modelin matematiksel gösterimi aşağıdaki gibidir:

$$HKT_{L_1} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (2.4)$$

2.3.2.3. ElasticNet Regresyon

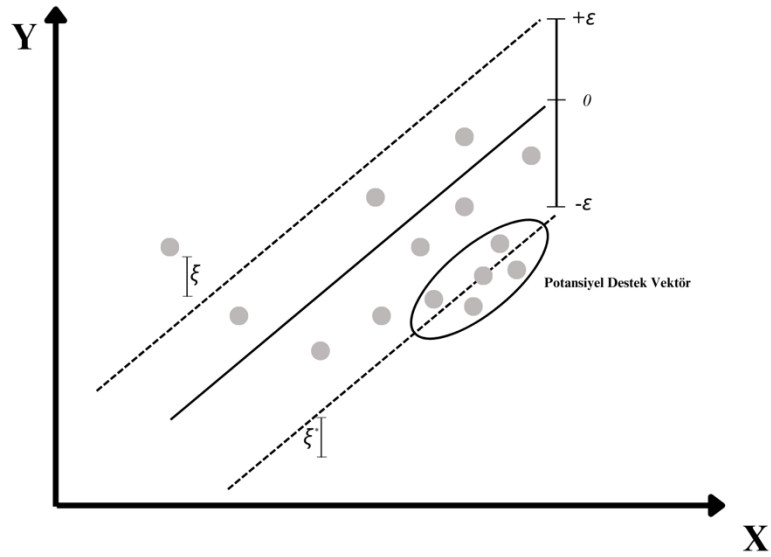
Ridge ve LASSO regresyonlarında olduğu gibi hata kareleri toplamını minimize etmeyi amaçlayan bir diğer düzenlileştirme normu ElasticNet regresyonudur. Ridge ve LASSO regresyonun bir karışımı olarak bilinmektedir. Bunun nedeni L_1 ve L_2 düzenlileştirmelerini birleştirerek işlemleri gerçekleştirmesidir. Bu nedenle eşitlikte (2.5) de görüldüğü gibi iki adet hiperparametresi ve iki adet ceza terimi ile tahminlerde bulunur. $\lambda = 0$ iken Ridge regresyonuna, $\lambda = 1$ iken LASSO regresyonuna denktir (Nacar ve Erdebili, 2021: 313-314).

$$HKT_{\text{Enet}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^p \beta_j^2 + \lambda_2 \sum_{j=1}^p |\beta_j| \quad (2.5)$$

2.3.3. Destek Vektör Makineleri

Makine öğrenmesi teknikleri içerisinde en güçlü algoritmalarından biri olan destek vektör makineleri (SVM), farklı çıktıları hedefleyen regresyon ve sınıflandırma problemleri olmak üzere iki kategoride de tercih edilebilir. Bu nedenle destek vektör makinelerini, destek vektör sınıflandırma (SVC) ve destek vektör regresyon (SVR) olarak iki ayrı çalışma prensibine ayırmak mümkündür. Sınıflandırma problemlerinde kesin sınıflandırmalar söz konusuken regresyon problemlerinde tahminlerde değişkenlik söz konusudur. Örneğin işsizliğin gelecekteki oranı tahmin edilmek istendiğinde sınıflandırma tekniklerinden yararlanılamaz. Bu noktada sonuçlar farklılık gösterebileceği için SVR tercih edilmelidir.

Şekil 13: Tek boyutlu Doğrusal Destek Vektör Regresyonu Modeli



Kaynak: Rasifaghihi, 2022.

Şekil 13 'ü incelediğimizde SVR 'de, basit regresyon modelinden farklı değerler olduğu görülmektedir. Belirli seviyede hataya izin vermek ya da aşırı uyumu (overfitting) önlemek adına “ ξ ”ve “ ξ^* ” sembolleri ile gösterilen bir alan mevcuttur. Bu alan genellikle e-tube ya da duyarsız bölge olarak nitelendirilmektedir. Duyarsız

bölgede yer alan gözlemler aykırı gözlemlerdir ve bu bölgenin genişliğini kontrol edebilmek için ε mesafesi dışarıda tutulur. Bu sayede eğitim verilerine maksimum seviyede uyum gösterebilmek için tahmin hataları kontrol altında tutulurken modelin karmaşık yapısı dengelenmeye çalışılır. Eğer ε büyükse, model daha fazla hataya izin verir ve daha az destek vektörü kullanır böylece tahminlerin daha basit olmasına yardımcı olur (Awad ve Khanna, 2015: 67-68). Bu algorithmada benzersiz bir sonuca ulaştıracak olan “konveks optimizasyon” olarak nitelendirilen bir matematiksel problem çözümlenmelidir. Buna göre hiper düzlem, ε -tube adı verilen bu alanın sınırları dışında bulunan eğitim örnekleri ile ilişkilendirilmektedir. Bu eğitim örneklerine ise “destek vektörler” denmektedir. Destek vektörleri sayesinde bu alanın şekli netleşmektedir. Çözümlemesi gereken minimizasyon problemi ve kısıtlar aşağıdaki gibi ifade edilir:

Minimizasyon Problemi:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (2.6)$$

Kısıtlar:

$$\begin{aligned} y_i - (w \cdot x_i) - b &\leq \varepsilon + \xi_i \\ (w \cdot x_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0, \quad i = 1, \dots, m \end{aligned} \quad (2.7)$$

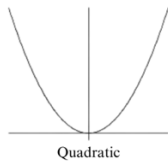
Ceza ya da kontrol parametresi olarak tanımlanan ve tahminlerin kabul edilebilir hata sınırlarını belirleyen C, modelin karmaşıklığı ile tahmin hataları arasındaki dengeyi belirlemektedir. Yani, C parametresi ile aykırı veya uç değerler (ξ ve ξ^*) üzerinde bir kontrol mekanizması çalıştırılır. Dolayısıyla C ne kadar büyükse, model karmaşıklığı o kadar fazla olur, ancak tahmin hataları daha düşük olur. C çok büyükse, model eğitim verilerini mükemmel bir şekilde uydurabilir, ancak bu aşırı karmaşıklık sorununa yol açabilir. C küçüldükçe, tahmindeki esneklik payı artar ve hatalara daha fazla izin verir.

Kısıtlarda yer alan x_i destek vektörleri ifade ederken, b ve w parametreleri hiper düzlemi tanımlamaktadır. ε , tahmin hatasının ne kadar büyük olabileceğini belirtir. Kısıtlar sayesinde gerçek değerler ile tahmin edilen değerler arasındaki farkların, regresyon eğrisinin iki yönünden de belirli bir ε ve ξ değerinden daha uzakta

olmaması sağlanır. Duyarsız bölgenin boyutu modelin dinamikleri doğrultusunda maksimum seviyeye getirilmeye çalışılır. Çünkü olabildiğince verinin bu alanın içerisinde kalması hedeflenmektedir. Gerçek ile tahmini değerler arasındaki farkların bu mesafeden daha uzakta kalmaması önemlidir. Artıklar belirli bir ε oranı dahilinde kaldığı sürece önemsiz sayılmaktadır. Bu durum doğrunun eğimini belirlemede ε 'dan küçük olan değerlerin herhangi bir etkiye sahip olmadığı anlamına gelmektedir. Özetle, SVR için doğru parametreleri ayarlamak, iyi tahminler elde etmek için oldukça önemlidir. Doğru C, ε ve γ değerlerini ayarlamak, SVR modelinin performansını optimize etmeye olanak tanımaktadır. Bu parametreler doğru bir biçimde ayarlandığında, tahminler daha güvenilir olacaktır (Yolasığmaz, 2015: 29-33).

Öte yandan, SVR 'de tahminlerin gerçek değerlerden uzaklığını ölçmek ve bu farkı minimize etmek amacıyla Smola tarafından tanıtılan kayıp fonksiyonlar (loss function) da oldukça önemlidir. En sık kullanılan kayıp fonksiyonlar aşağıda detaylandırılmıştır. Buna göre veri yığınının dinamiklerine bağlı hangi kayıp fonksiyonun seçilebileceğine karar verilir. Bu da destek vektör regresyonunun destek vektör sınıflandırmaya göre daha esnek yapıda olduğunun göstergelerinden biridir (Kor, 2015: 71-72). Şekil 14 'de ifade edilen quadratic yani kare kayıp fonksiyonu, en küçük kareler hesaplamasına dayanır. Bu kayıp fonksiyonla birlikte model tahminleri arası fark karelerinin toplamını minimize etmek hedeflenmektedir.

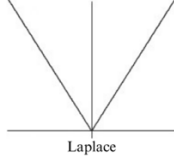
Şekil 14: Quadratic Kayıp Fonksiyonu



Kaynak: Gunn, 1998:29.

Şekil 15 'de ifade edilen laplace kayıp fonksiyonunun sapmalara karşı hassasiyeti quadratic fonksiyona göre daha azdır. Bu nedenle veri sapmaları daha çok gerçekleşir yani daha esnek bir yapıdadır.

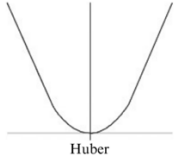
Şekil 15: Laplace Kayıp Fonksiyonu



Kaynak: Gunn, 1998:29.

Şekil 16 'da ifade edilen ve Huber tarafından 1964 'de sunulan Huber kayıp fonksiyonu genellikle veri dağılımının bilinmediği problemlerde iyi sonuç sağlamaktadır. Ancak bu fonksiyonda bazı noktalardaki yoğunlaşmadan kaynaklı modellerde dengesizlik olabilir. Bu durum da bu kayıp fonksiyonu diğerlerine göre dezavantajlı yapmaktadır.

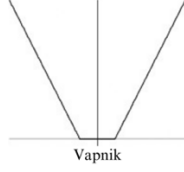
Şekil 16: Huber Kayıp Fonksiyonu



Kaynak: Gunn, 1998:29.

Şekil 17 'de ifade edilen Vapnik tarafından tanıtılan kayıp fonksiyon, Huber 'ın yaklaşımına benzerdir, ancak destek vektörlerinin daha dengeli bir şekilde dağılmasını sağlar (Vapnik, 1982: 430). Bu da modelin daha dengeli ve güvenilir tahminler yapmasına yardımcı olmaktadır.

Şekil 17: Vapnik Kayıp Fonksiyonu



Kaynak: Gunn, 1998:29.

2.3.4. Karar Ağaçları

Karar Ağaçları (DT) hem regresyon hem de sınıflandırma problemlerini çözmek için yaygın olarak kullanılan bir makine öğrenmesi türüdür. DT modelinin amacı sıralı kararları belirli bir yapı dışına çıkmadan ifade etmek ve bir hedefe ulaşma ihtimalini maksimum seviyeye taşımaktır. Sınıflandırma problemlerinde doğru kategoriye atamak için kullanılan DT, regresyon problemlerinde değişkenlerin yarattığı etkileri görmek için tercih edilmektedir (Sharma ve Kumar, 2016: 2094). DT, Şekil 18 'deki gibi öğrenme sürecindeyken edindiği bilgiyi ağaç üzerinde modellemektedir.

Şekil 18: DT Genel Yapısı



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Kök düğüm (root node), DT'nin başlangıcını temsil etmektedir. En başta, tüm veri kümesi bu kök düğümüdür ve buradan çeşitli özelliklere veya koşullara göre ayrılmaya başlar. Yani, karar ağacının en başında yer alan bu düğüm, verinin başlangıç noktasına işaret eder ve bu noktadan itibaren veriyi belirlenmiş birtakım kriterlere bölerek gruplamaya başlar. Karar düğümleri (decision node), söz konusu bu kök düğümlerinin bölünerek ayrışmasıyla oluşan düğümlerden meydana gelir. Bu düğümler de veri kümesinin farklı özelliklere göre nasıl ayrılacağına karar verir. Bu yanıyl DT için ana unsurlarından biri olma özelliğini taşımaktadır. Yaprak düğümleri (terminal düğümleri), daha fazla bölünemeyecek olan düğümleri temsil etmektedir ve nihai sonucu yansıtır. Yani karar ağacının sonuçlarını taşımaktadırlar. Alt-ağaç (sub-tree), karar ağacının belirli bir kısmını ifade etmek için kullanılan bir terimdir. İfade ettiği bu bölümde çeşitli kararlar bulunur. Budama, bir karar ağacındaki belirli düğümlerin kaldırılmasına denir. Buradaki amaç aşırı uyum (overfitting) sorunlarını önlemek ve modeli daha anlaşılır bir boyuta taşımaktır. Yani budama işlemi yapılarak karar ağacında olabilecek gereksiz karmaşıklık önlemek amaçlanmaktadır.

DT modelini uygun dallara ayırabilmek için çeşitli yöntemler kullanılmaktadır. Bu yöntemler kullanılan değişkenlere göre çeşitlilik gösterebilir. Kapsamdaki değişkenler kategorikse entropi endeksi, gini endeksi veya sınıflandırma hatası endeksi tercih edilebilecek yöntemlerden birkaçına örnek olarak gösterilebilir. Ancak değişkenler sürekli ise genellikle en küçük kareler sapması yöntemi tercih edilmektedir. DT modellerinde tercih edilen bazı algoritmalar aşağıdaki gibidir:

- Sınıflandırma ve Regresyon Ağacı (CART)
- İteratif İkिलilikçi Ağacı (ID3)
- C4.5
- C5.0
- Ki-kare Otomatik Etkileşim Tespiti (CHAID)
- Karar Kökü
- Model Ağacı (M5)
- Koşullu DT

DT, kolayca yorumlanabilmesi sayesinde büyük avantajlar sağlasa bile karmaşıklığa olan yatkınlığı göz önünde bulundurulmalıdır. Çünkü bazı durumlarda

ikili, kategorik ve sayısal tahminciler aynı anda yer alabilir. Bu da süreci daha kompleks olmaya itebilir. Öte yandan verilerdeki küçük oynamalar bile karar ağacının tüm yapısını değiştirebilir. Bu nedenle, DT modellenirken dikkatli olunmalıdır. Karmaşıklığı önlemek veya aşırı uyum problemini kontrol altında tutmak için en uygun yöntem tercih edilmelidir.

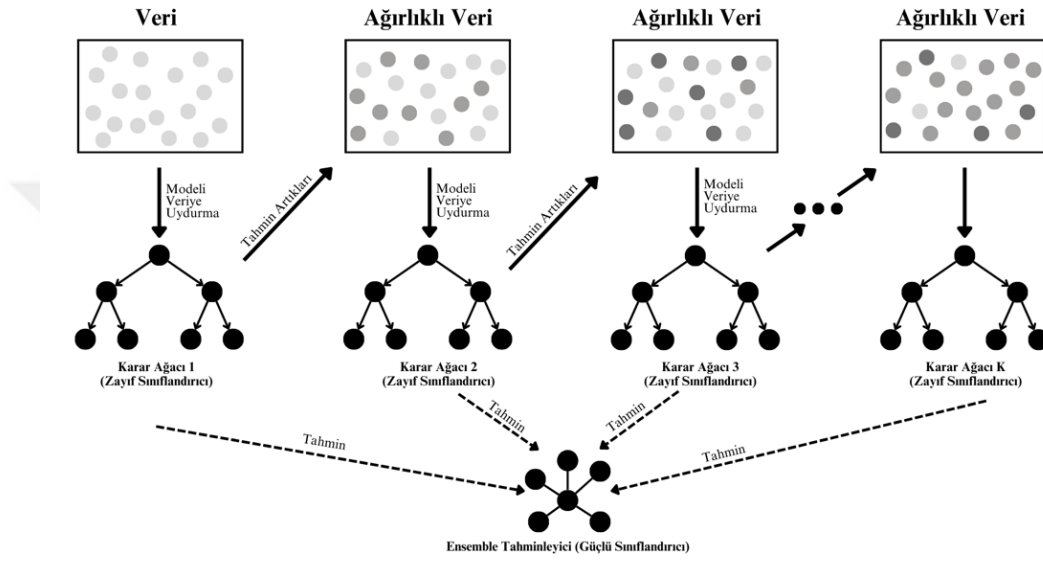
2.3.5. Gradyan Güçlendirmeli Karar Ağaçları

Makine öğrenmesi algoritmalarının göstermiş olduğu performansı geliştirmek ve farklı model tahminlerini bir araya getirerek üzerinde çalışılan problemleri çözmek adına kullanılan Ensemble yöntemleri, genelde üç ana kategoriye ayrılır. Bu kategoriler: torbalama (bagging), artırma (boosting) ve oylama (voting) olarak adlandırılır (Brownlee, 2016: 91). Ensemble yöntemleri içerisinde yer alan boosting kategorisinde en sık tercih edilen yöntemler; AdaBoost. (uyarlanabilir yükseltme) ve gradyan güçlendirmeli karar ağaçlarıdır (GBDT).

GBDT, Leo Breiman'ın 1997 yılında yayımlanan “*Arcing the Edge*” adlı çalışmasında ilk kez dile getirilmiş ve Jerome H. Friedman'ın 1999 yılında yayımlanan “*Greedy Function Approximation: A Gradient Boosting Machine*” başlıklı çalışmasında daha ileri bir seviyeye taşınarak geliştirilmiştir (Géron, 2019: 203). Hem regresyon hem de sınıflandırma problemlerinde yaygın bir biçimde tercih edilen GBDT algoritmaları genel yapı itibarıyla zayıf öğrencileri birleştirerek güçlü tahmin modelleri oluşturmak için kullanan ve gerçek değerler ile tahmin edilen değerlerin farkından meydana gelen artıkların optimize edilmesine dayalı bir ağaç yöntemidir (Julian, 2016: 179). Burada bahsi geçen kötü tahminlerde bulunan zayıf öğrencileri iyi tahminde bulunan güçlü öğrenciler haline getirmek hedeflenmektedir. Bu hedef doğrultusunda ardışık modeller dizisi oluşturulur. Oluşturulan bu modeller ise kendilerinden önceki modellerin hatalarını düzeltmeye çalışmaktadır. Öte yandan, bir tahminin iyi veya kötü olduğuna karar vermek için kullanılan değerlendirme ölçütü ise hata kareleridir. Hata kareleri ne kadar büyürse tahminin o derece kötü olduğu düşünülmektedir. Modeller tahminlerde bulunur ve tahminde bulunan bu modellerin doğruluk dereceleri tespit edilir. Doğruluk derecelerine bağlı ise ağırlıklandırılabilirler. Özetle hata karelerinin toplamını minimum seviyede tutacak

katsayılar bulmayı hedefleyen tüm bu süreçte, elde edilen sonuçlar birleştirilerek nihai bir tahmine ulaşılmaktadır (Brownlee, 2016: 94). Ulaşılan bu nihai tahmindeki son model, söz konusu ardışık zayıf modellerin bir kombinasyonu olarak nitelendirilmektedir (Burkov, 2019: 83). Bu süreç Şekil 19 'da detaylandırılmıştır.

Şekil 19: Gradient Boosting Decision Tree Mimari Yapısı



Kaynak: Zhang ve diğerleri, 2021: 4-5.

GBDT algoritması için başlangıçta, bir eğitim modeli (f_0) oluşturulur ve tahminler gerçekleştirilir. Gerçekleştirilen bu tahminlerle gerçek değerler arasındaki farklar hesaplanır ve her bir örneğin tahminleri düzeltilmiş etiketlerle değiştirilir. Sonrasında eğitim setindeki her örnek $i = 1, \dots, N$ eklenir. Tahminlerin yanlıtıcılık seviyesi arttıkça farklar da o kadar büyük olur.

$$f = f_0(x) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N y_i \quad (2.8)$$

$$\hat{y}_i \leftarrow y_i - f(x_i) \quad (2.9)$$

Sonra, bu düzeltilmiş etiketlerle yeni bir model (f_1) oluşturulur. Bu iki model birleştirilerek bir artırma modeli (f_d) elde edilir ve bu noktada öğrenme hızı (α)

denilen bir hiper parametre devreye girer. Bu parametre, her iki modelin katkısını dengelemek amacıyla eklenir. Bu aşamada matematiksel gösterim aşağıdaki gibi olur:

$$f \stackrel{\text{def}}{=} f_0 + \alpha f_1 \quad (2.10)$$

Daha sonra, bu yeni modelin tahminlerini kullanarak tekrar artıklar hesaplanır ve eğitim verilerindeki etiketleri güncelleyerek bir sonraki model (f_2) oluşturulur. Ardından bu yeni model öncekilerle birleştirilerek artırma modeli güncellenir. Bu işlem, belirlenen maksimum ağaç sayısına (m) ulaşana kadar devam etmektedir.

$$f \stackrel{\text{def}}{=} f_0 + \alpha f_1 + \alpha f_2 \quad (2.11)$$

Yani temelde, her yeni modelin öncekilerin hatalarını düzeltmeye çalıştığı bir dizi model oluşturulmaktadır. Bu şekilde tahminleri daha iyi hale getirmek hedeflenmektedir. Gradyan artırma üzerinde ayarlama yapılması gereken üç temel hiperparametre; ağaç sayısı, öğrenme hızı ve ağaçların derinliğidir. Çünkü bu üç parametre de modelin doğruluğunu etkilemektedir. Ağaçların derinliği aynı zamanda eğitim ve tahmin hızını da etkilemektedir. Derinlik az ise tahminin daha hızlı olduğu söylenebilir (Burkov, 2019: 85).

Son olarak, GBDT algoritmalarının birçok farklı çeşidi vardır. GBDT yöntemlerinin hızını artıran ve performansını iyileştiren ölçeklenebilir versiyonu aşırı gradyan artırma (XGBoost) yöntemidir. Hafif gradyan artırma makinesi (LightGBM) ise XGBoost yönteminin eğitim süresi performansını arttırmaya yönelik geliştirilen GBDT türüdür. Kategori artırma (CatBoost) yöntemi ise kategorik değişkenler ile otomatik bir biçimde mücadele edebilen, hızlı olmasıyla bilinen GBDT örneğidir (Koç, 2022). Kendi içerisinde birçok uyarlanabilir yöntemle sahip olan GBDT, aykırı verilere dirençli ve karmaşık verilere karşı dayanıklı bir algoritmadır. Ancak büyük veri kümeleriyle çalışırken ölçeklendirme zorluğu dezavantaj olarak ortaya çıkabilmektedir. Bu nedenle problem çözümlerinde bu avantaj ve dezavantajların dikkate alınması gerekmektedir (Julian, 2016: 179).

2.3.6. Rassal Orman

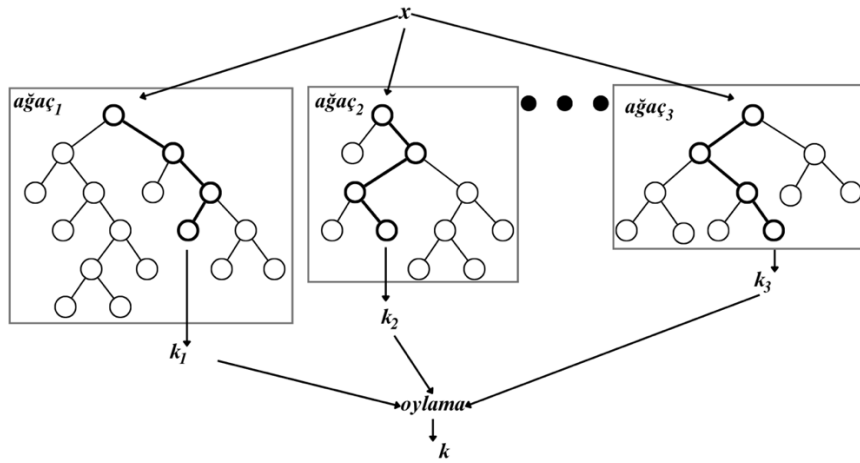
Leo Breiman 'ın öne sürdüğü hem sınıflandırma hem de regresyon problemlerinde kullanılan rassal orman (RF) karar ağaçlarının gelişmiş bir versiyonu olarak bilinmekte ve birden çok ağacın veya algoritmanın tahminlerini birleştirilerek sonuç elde etmesi nedeniyle ensemble öğrenme teknikleri arasında yer almaktadır (Lantz, 2019: 367). Ensemble tekniklerinde yer alan torbalama (bagging veya bootstrap aggregation) kategorisinde en yaygın kullanılan algoritmalar çanta karar ağaçları (bagged decision trees), RF ve ekstra ağaçlardır (extra trees) (Brownlee, 2016: 92).

RF, eğitim verilerini torbalama (bagging veya bootstrap aggregation) yöntemi ile rastgele örneklemeler oluşturarak seçer ve değişkenleri rastgele alt uzay yöntemi (random subspace) metodu ile rastgele seçerek birden çok karar ağacının oluşmasına olanak tanır. Bu yapısı sayesinde ise varyansa karşı dirençli olur ve aşırı uyum (overfitting) sorununu azaltmayı hedefler. Ayrıca, hata kareler ortalamasını düşürerek diğer yöntemlere kıyasla daha net tahminler elde edilmesine katkıda bulunmaktadır (Brownlee, 2016: 93). Bu karar ağacı topluluğuna RF denmesinin ana nedeni, bu ağaç topluluğunun her bir alt kümesinin eğitim veri setinden rastgele örneklemeler içermesinden kaynaklanmaktadır (Jo, 2021: 160-162). RF, karar ağaçlarını oluştururken daha fazla rastgelelik eklemektedir. Her bir ağacı oluştururken, en iyi özelliği seçmez. Bunun yerine rastgele birkaç özellikten birini oy kullanarak (voting) seçer. Bu yaklaşımdan dolayı farklı ağaçlar oluşturulmuş olur. Yani oluşturulmuş her bir ağaç benzersiz rastgele özelliklerini kullanmaktadır. Bu da model çeşitliliğini artıran bir unsurdur (Géron, 2019: 197). Bu sayede aşırı uyum (overfitting) problemi azalır. Böylece bu ağaçların birleştirilmesiyle RF daha kuvvetli ve genelleştirilebilir tahminlerde bulunabilir (Brownlee, 2016: 93).

RF yönteminin öğrenme süreci ve sınıflandırma süreci Şekil 20 ve Şekil 21 'de sunulmuştur. Öncelikle x eğitim seti alt kümelere ayrıştırılır. Ayrıştırılan her bir alt küme kullanılarak n adet karar ağacı oluşturulur. Yeni bir örnek, rastgele orman içerisinde yer alan her karar ağacı tarafından sınıflandırılır. Yeni ögenin nihai çıktısı, probleme bağlı olarak karar ağaçlarının çıktılarının (k_n) oylama ya da ortalaması kullanılarak belirlenir. Şekillerde belirtilen oy kullanma (voting) ise birden fazla karar

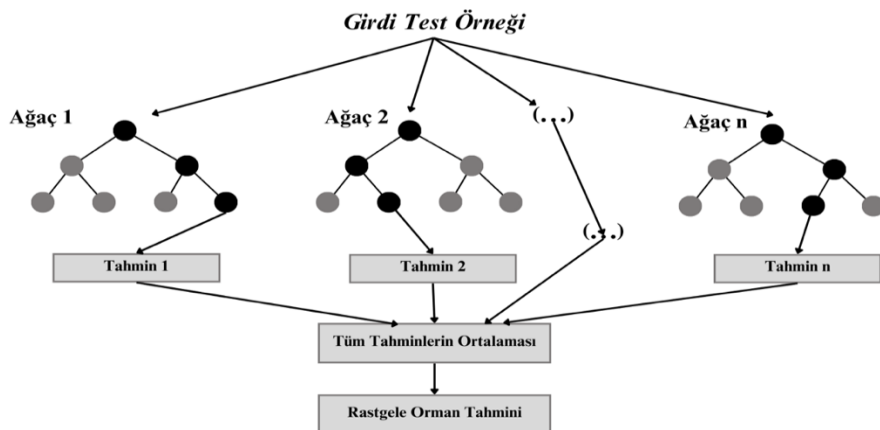
ağacının ya da bir dizi farklı tahmin modelinin bir araya getirilmesi sonucu model tahminlerinin toplu bir biçimde değerlendirilmesini ifade etmektedir. Voting ya da oy kullanma çoğunlukla sınıflandırma problemlerinde kullanılsa da birtakım regresyon problemlerinde de kullanılabilir (Jo, 2021: 160-162).

Şekil 20: Sınıflandırıcı RF Örnek Gösterimi



Kaynak: Dangeti, 2017: 198-200.

Şekil 21: Regresyon Tabanlı RF Örnek Gösterimi



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

2.3.7. K-En Yakın Komşu Algoritması

Parametrik bir yapıya sahip olmayan ve tahmin edilen değerin çeşidine bağlı olarak hem sınıflandırma hem de regresyon problemlerinde sıkça tercih edilen KNN algoritması temel olarak eğitim verilerinde yer alan ve tahmini yapılan veriye en yakın gözlemleri tespit ederek etiketleme işlemi gerçekleştirmektedir.

Genellikle model oluşturulduktan sonra eğitim verilerinin atıldığı bir düzene sahip olan birçok algoritmanın aksine KNN algoritmasında eğitim verileri atılmamaktadır. KNN, yeni eklenen örnek için, öznitelik vektörleri uzayında bu örneğe en yakın k eğitim örneğini bularak sınıflandırma problemlerinde en sık tekrar eden etiketi, regresyon problemlerinde ise ortalama etiketi tespit etmektedir. Burada söz konusu olan k parametresi, sınıf tahmini veya regresyon uygulamalarında performans seviyesini artırmayı sağlamak için ayarlama yapılabilen bir değerdir. Bu parametrenin tam sayı olma koşulu vardır ve komşu sayısını ifade etmektedir. K parametresi küçüldükçe verilere uyum artabileceğinden aşırı uyum (overfitting) sorununun meydana gelebileceği göz önünde bulundurulmalıdır. Öte yandan KNN, genellikle bir gözlemin komşularını uygulamaya bağlı olarak tercih edilebilecek çeşitli mesafe ölçüleri ile belirler. Öklid mesafesi en sık tercih edilen mesafe ölçüsüdür ve matematiksel gösterimi aşağıdaki gibidir:

$$d(p, q) = \sqrt{\sum_i (p_i - q_i)^2} \quad (2.12)$$

Öklid mesafesine ek olarak Manhattan mesafesi ve kosinüs benzerliği de en sık tercih edilen metriklere örnek olarak gösterilebilir.

2.4. REGRESYON MODELLERİNDE PERFORMANS İSTATİSTİKLERİ

Model performansını ve sonuçlarını değerlendirmek üzere tercih edilen performans istatistikleri analiz tipine, veri tipine veya oluşturulacak modelin hedefine bağlı olarak değişkenlik göstermektedir. Ancak regresyon problemlerinde sıkça tercih edilen belli başlı performans metrikleri bulunmaktadır. Bu başlıkta bu metrikler detaylandırılmıştır.

2.4.1. R-Kare

Belirleme katsayısı olarak da bilinen R^2 , bir regresyon modelinin gerçek değerlere olan uyumunu ölçmektedir. Doğrusal regresyon modellerinin başarı oranı hakkında fikir veren bu değer sıfır ile bir arasındadır ve bire ne derece yakınsa modelin verileri o kadar iyi açıkladığı yani modelin uyum sağlama yeteneğinin yüksek olduğu ve modelde kullanılan bağımsız değişkenlerin bağımlı değişkendeki varyansı daha iyi açıkladığı bilgisini vermektedir. Sıfıra yakınlığı ise modelin verilere uyum sağlamadığı hakkında bilgi vermektedir (Brownlee, 2016: 68). Bu değerın matematiksel gösterimi aşağıdaki gibidir:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.13)$$

2.4.2. Düzeltilmiş R-Kare

Düzeltilmiş R^2 regresyon modeline eklenen bağımsız değişkenlerin R^2 'yi artırarak modeli çok daha karmaşık bir hale getirmesi sonucu ortaya çıkabilecek problemleri (örneğin overfitting) minimumda tutmak için kullanılır. R^2 'nin bir varyasyonu olan bu metrik eklenen her bir bağımsız değişkenin modelin uygunluğunu artırıp artırmadığını değerlendirmektedir. Bu değerlendirme sonucu modelin gerçek performansını yansıtmayı amaçlar. Genellikle çoklu regresyon problemlerinde tercih edilen bu ölçüt tıpkı R^2 gibi regresyon modelindeki bağımsız değişkenlerin bağımlı değişkendeki toplam varyansı ne derece açıklayabildiği bilgisini sıfır ile bir arasında bir değer ile verir. Bire olan yakınlığı arttıkça regresyon modelinin verileri açıklama kabiliyeti artar. Düzeltilmiş R^2 'nin matematiksel gösterimi aşağıdaki gibidir:

$$Düzeltilmiş R^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \quad (2.14)$$

2.4.3. Ortalama Mutlak Hata

Ortalama mutlak hata (MAE), tahmin değeri ile gerçek değer arasındaki mutlak farkların ortalamasını ifade etmektedir. Bu ölçüt sayesinde tahminin ne derece hataya sahip olduğu bilgisine ulaşılır. Hataların yönünden ziyade boyutu hakkında fikir veren bu istatistiksel değer ne kadar düşükse modelin tahminlerinin gerçeğe yakınlığı o kadar yüksektir. Bu değerın matematiksel gösterimi aşağıdaki gibidir:

$$MAE = \frac{1}{N} \sum_{j=1}^N |y_j - \hat{y}_j| \quad (2.15)$$

2.4.4. Hata Kareler Ortalaması

Hata kareler ortalaması (MSE), gözlemlenen ve tahminlenen değerler arasındaki ortalama kare farkını değerlendiren bir istatistiktir. Bu değer düştükçe regresyon modelinin doğruluk derecesi artmaktadır ve modelde hata mevcut değilse ortalama kare hata değeri sıfıra eşittir. Hataların karelerinin alınması sebebiyle büyük hataların daha fazla etkisinde kalan MSE, MAE ölçütüne göre aykırı değerlere daha az dayanıklılık gösterir. Bu değerın matematiksel gösterimi aşağıdaki gibidir:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.16)$$

2.4.5. Ortalama Karesel Karekök Hatası

Ortalama karesel karekök hatası (RMSE), hataların gerçek veriler üzerindeki boyutunu açıklar ve bu yönüyle regresyon modellerinin performansını değerlendirmek için sıkça tercih edilir. RMSE değeri ne kadar düşükse modelin de o kadar iyi olduğu kanısına varılır. Mükemmel olarak nitelendirilebilecek modellerde RMSE değeri sıfıra eşit olmalıdır. Bu değerın matematiksel gösterimi aşağıdaki gibidir:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.17)$$

2.4.6. Ortalama Mutlak Yüzdesel Hata

Ortalama mutlak yüzdesel hata (MAPE), hataların mutlak değerlerinin ortalaması alınarak, bu ortalamaların gerçek değerlerin yüzdesi olarak göstermek amaçlanmaktadır. Bu değer % 10 'un altında olduğunda genellikle tahmin modellerinin yüksek oranda doğruluğa sahip olduğu varsayılmaktadır. %10 ile %20 arasında değerlere sahip olan modeller ise doğru modeller olarak nitelendirilmektedir. Eğer bu değer %50 üstünde ise tahmin modellerinin hatalı olduğu veya genel olarak yanlış olduğu kabul edilir. Bu ölçütün matematiksel gösterimi aşağıdaki gibidir:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (2.18)$$

2.5. KARMA VERİ ÖRNEKLEMESİ

Klasik zaman serisi analiz yöntemlerinde bağımlı ve bağımsız değişkenler aynı frekanslara sahiptir. Örneğin bağımlı değişken aylık ise bağımsız değişkenlerin de aylık olması beklenmektedir. Ancak regresyon modelinde tercih edilecek veriler farklı frekanslara sahipse, Ghysels ve diğerleri (2004) tarafından öne sürülen ve farklı frekans değerlerine sahip verilerin modelde kullanılabilmesi imkanını sunan “karma veri örnekleme” tekniğinden (MIDAS) faydalanılabilir.

Aylık verilere sahip ekonomik göstergelerle birlikte günlük, haftalık ya da aylık değişkenlerin de analize dahil edilebildiği MIDAS yöntemine göre bağımlı değişken frekansı bağımsız değişken frekansından her zaman düşük tutulmaktadır. Bu sayede MIDAS, yüksek frekansa sahip seri içerisindeki her bir gözlemde yer alan maksimum bilginin kullanılabilmesine olanak tanır. Söz konusu düşük frekans kavramı, tercih edilen değişkenlerin gözlemlenme sıklığı anlamına gelmektedir. Yani günlük, haftalık ve aylık gözlemlere sahip veri setlerimiz var ise, günlük gözlemler yüksek frekanslı iken aylık gözlemler düşük frekanslıdır. Çünkü günlük veriler için yılda 365 gözlem varken, aylık veriler için 12 gözlem mevcuttur (Armesto ve Engemann, 2010: 523).

Orijinal frekansların analiz sürecine dahil olmasına olanak tanıyan MIDAS regresyonu ile eksik veriler tamamlanmakta veya gelecek dönem verilerini öngörmeyi

hedefleyerek düşük frekanslı değişkenlerin geçmiş değerlerinden yararlanılmaktadır. Ardından yüksek frekanslı açıklayıcı değişkenlere ağırlıklandırma işlemi uygulanarak seri dönüştürülmekte yani sınırlandırılarak modele dahil edilmektedir (Ghysels ve diğerleri, 2004: 4). MIDAS ile oluşturulan modelin matematiksel gösterimi aşağıdaki gibidir:

$$Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \sum_{k=1}^m \gamma_k L_{HF}^k X_t + \varepsilon_t \quad (2.19)$$

Matematiksel gösterime göre Y_t bağımlı değişken iken, $L^i Y_t$ bu bağımlı değişkenin önceki zamanlardaki değerlerini yani gecikmeli değerlerini ifade etmektedir. Dolayısıyla düşük frekansa sahip değişkendir. β_i parametreyi, i model içerisinde yer alacak gecikmelerin başlangıç noktasını ve p maksimum gecikme değerini ifade etmektedir. Ayrıca, $L_{HF}^k X_t$ ifadesi yüksek frekansa sahip değişkenin gecikmeli değerini temsil etmektedir. Gecikme operatörü L , sabit terim α , parametre γ_k , gecikme başlangıç değeri k , maksimum gecikme sayısını gösteren m ve hata terimini ifade eden ε_t ise diğer model parametreleridir (Armesto ve Engemann, 2010: 521-536).

Öte yandan, MIDAS yönteminin çeşitli türevleri de mevcuttur. Örneğin, temel MIDAS regresyonu, farklı zaman dilimlerindeki verileri birleştirme işleminde ağırlıkları kullanan bir matematiksel formül içermekte ve düşük frekansa sahip verilere daha az, yüksek frekansa sahip verilere daha çok ağırlık verebilmektedir. Ancak belirsizlik ve oynaklığın da dikkate alındığı varyasyonu “sınırlandırılmamış karma veri örnekleme” (UMIDAS), zaman serilerini tahminlerken belirsizlikleri dikkate almakta ve bunları modellemeyi amaçlamaktadır. Buradaki temel hedef tahminleri sağlamlaştırmaktır. Dolayısıyla UMIDAS yönteminde modeller kurulurken belirsizlik ve oynaklık faktörlerinin eklendiği özel modeller kullanılmaktadır. Matematiksel gösterimleri genellikle tercih edilen modelin karmaşıklık seviyesine ve çalışma alanına bağlı olarak değişkenlik gösterebilmektedir.

2.6. GOOGLE TRENDS

Google Trends (GT), 2006 yılında Google tarafından kullanıcıların arama sorguları içerisinde en popüler olan kelime, kelime grupları ya da konuları sunan bir hizmet olarak sunuldu. 2018 yılına kadar çeşitli güncellemeleri ekleyerek gelişim sürecine devam eden GT, 2018 'de bugün hala kullanılan “Google Trends Keşfet” bölümü ile özelleştirme seçeneklerini artırarak son halini almıştır. Bugünkü GT, kullanıcılara 2006 'da sunulan orijinal sürümü ile kıyaslandığında, daha kapsamlı ve detaylı veri görüntüleyebilme imkanına sahip olması ile ön plana çıkmaktadır. Ek olarak sunduğu bu verilerin analiz için çeşitli sınırlamalara tabi tutularak özelleştirilebilmesi ve grafiklerle desteklenmesi de önemli özelliklerindendir.

Kelime ya da kelime grupları ile gerçekleştirilen arama sorguları bu kısımdan itibaren “sorgular” olarak nitelendirilecektir. Sorgular GT platformunun “keşfet” bölümünde aratıldığında uygulanabilecek özelleştirilmiş sınırlamalar aşağıda detaylandırılmıştır.

2.6.1. Coğrafi Konum ve Dil

Sorgular dünya genelinde incelenebilirken, tercih edilen bir ülke, bir il ya da bir ilin herhangi bir bölge ya da bölgeleri için sınırlandırılabilir. Bu sayede GT, kullanıcının tercih ettiği coğrafi konum çerçevesinde kalarak sorgu bilgilerini sunmaktadır. Ancak GT çeşitli mantıksal ya da teknik nedenlerle bazı bölgeler için bu işlemi gerçekleştiremeyebilir. Örneğin “işsizlik” sorgusu için “Afganistan” bölgesi tercih edildiğinde, yeterli veri olmaması nedeniyle GT herhangi bir sonuç üretmeyecektir. Burada herhangi bir sonuç üretilmemesinin nedeni Afganistan 'da pek tercih edilmeyen Türkçe dili ile bir arama yapılmasıdır. Genellikle tercih edilen bölgeye has bir dil ya da tüm dünyada yaygın olan diller ile aramalar gerçekleştirilip, sağlıklı sonuçlar alınabilir.

2.6.2. Zaman

Sorgu sonucunu sınırlandırmak amacıyla belirli bir zaman aralığı tercih edilebilir. “2004 – bugün”, “Son 5 Yıl”, “Son 90” gün gibi seçeneklerle ilerlenebilir veya kullanıcının tercih ettiği belirli bir zaman dilimi manuel olarak girilerek sonuçlar görüntülenebilir. Google ’ın 2004 yılından itibaren dünya çapında kullanılmaya başlanması ve bu tarihten itibaren büyüyerek veri kaydetme kapasitesini artırması nedeniyle GT, en eski tarih olarak 2004 yılını baz alır ve bu yıldan itibaren veri sunabilmektedir. Ek olarak, verileri tercih edilen zaman aralığına bağlı belirli ölçeklerde gruplayarak sunmaktadır. Örneğin son bir haftalık “işsizlik” sorgu sonuçlarına erişim istendiğinde saatlik ölçeklerde verileri sunarken, son bir yıllık “işsizlik” sorgu sonuçları için bir haftalık ölçekler halinde verileri sunacaktır.

2.6.3. Kategori

“Tüm Kategoriler” ya da “Alışveriş”, “Bilim”, “Haberler”, “Seyahat” ve “Spor” gibi ana kategoriler tercih edilerek sorgu frekansları kategori bazlı görüntülenebilmekte ya da bu ana kategorilerin alt kategorileri seçilerek detaylandırılabilir. Örneğin “Bilim” ana kategorisinin “Kimya”, “Matematik”, “Fizik” gibi alt kategorileri seçilerek ulaşılmak istenen verinin kapsamı netleştirilebilir.

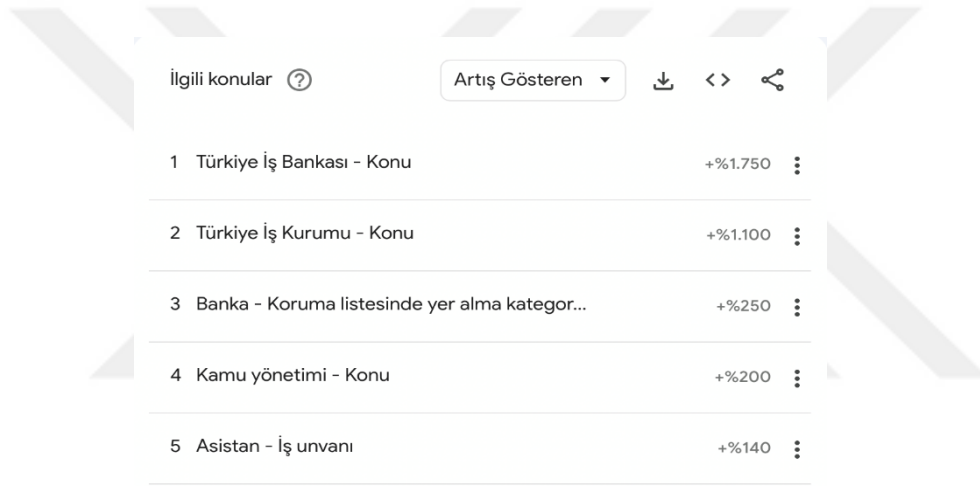
2.6.4. Google Filtresi

Google ’ın sahip olduğu arama filtreleri içerisinde “Görsel Arama”, “Google Haberler Arama”, “Google Alışveriş” veya “YouTube Arama” seçeneklerinden birisi baz alınarak arama yapılan sorgu için sınırlandırma yapılabilmektedir. Örneğin herhangi bir alanda yayınlanan videolar ile ilgili bir çalışma yapılmak istendiğinde “YouTube Arama” seçeneği ile verileri görüntülemek daha sağlıklı sonuçlar verebilir.

2.6.5. İlgili Konular

GT, “İlgili konular” hizmeti aracılığıyla, kullanıcının girdiği arama sorgusuyla en ilgili olan yirmi bir konuyu veya artış gösteren sekiz konuyu sunmaktadır. Bu, kullanıcının belirli bir arama sorgusuyla ilişkilendirilen diğer konuları incelemesine ve sorguyla ilişkili olarak gündemde olan konuları anlayabilmesine olanak tanımaktadır. Şekil 22 ’de “iş” sorgusu aratıldığında GT arayüzünün sunduğu ilgili artış gösteren konulardan birkaçı ifade edilmiştir.

Şekil 22: Google Trends: “İlgili Konular” Örneği



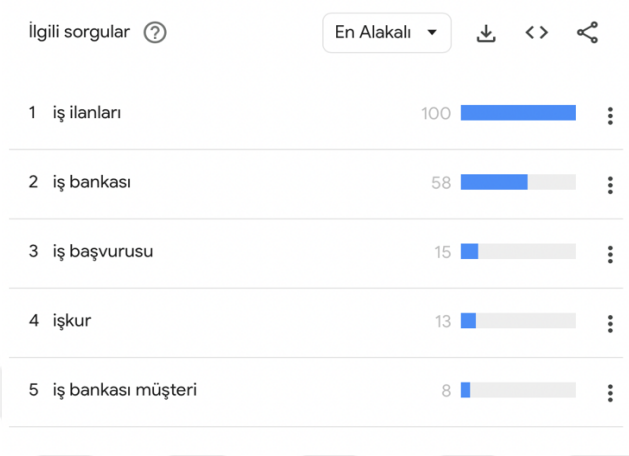
Kaynak: GT, 2023.

2.6.6. İlgili Sorgular

GT, “İlgili sorgular” hizmeti aracılığıyla, kullanıcıların yaptığı belirli bir arama sorgusu ile en yakından ilişkilendirilen ilk yirmi beş sorguyu veya artış gösteren ilk yirmi sorguyu sunmaktadır. Bu hizmet, kullanıcının girdiği arama sorgusu ile ilgili olan başka kelime veya kelime gruplarını belirlemek için kullanılmaktadır. Özellikle, “İlgili sorgular” hizmeti, belirli bir arama sorgusunun bağlantılı olduğu diğer sorguları görmek için kullanılırken, “İlgili konular” hizmeti, belirli bir arama sorgusunun ilgili olduğu konuları görmek için kullanılmaktadır. Bu nedenle bu iki hizmet aynı gibi

görünse de aslında farklı içeriklere sahiptir. Şekil 23 'de "iş" sorgusu aratıldığında GT 'in sunduğu sorgu ile en alakalı diğer sorgular gösterilmektedir.

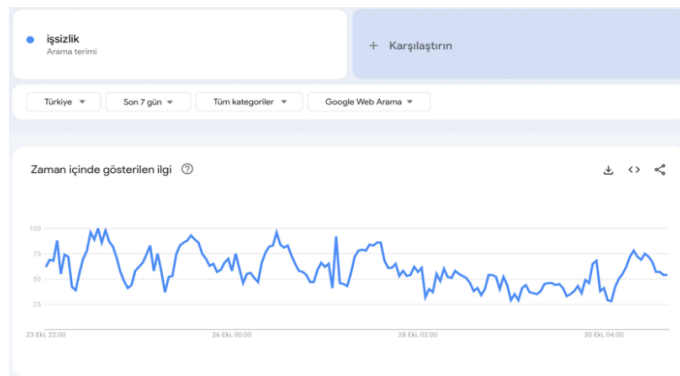
Şekil 23: Google Trends: "İlgili Sorgular" Örneği



Kaynak: GT, 2023.

GT, bir sorgunun zaman içindeki arama yoğunluğunu grafikleştirir ve aynı dönemde birden fazla sorgunun karşılaştırılmasına imkân sağlar. Şekil 24, Türkiye 'de Google Web Arama filtresi altında tüm kategorilerde sorgulama yapılan gün baz alınarak "son 7 gün" içinde gerçekleştirilen "işsizlik" sorgusunun arama yoğunluğu grafiğini sunmaktadır.

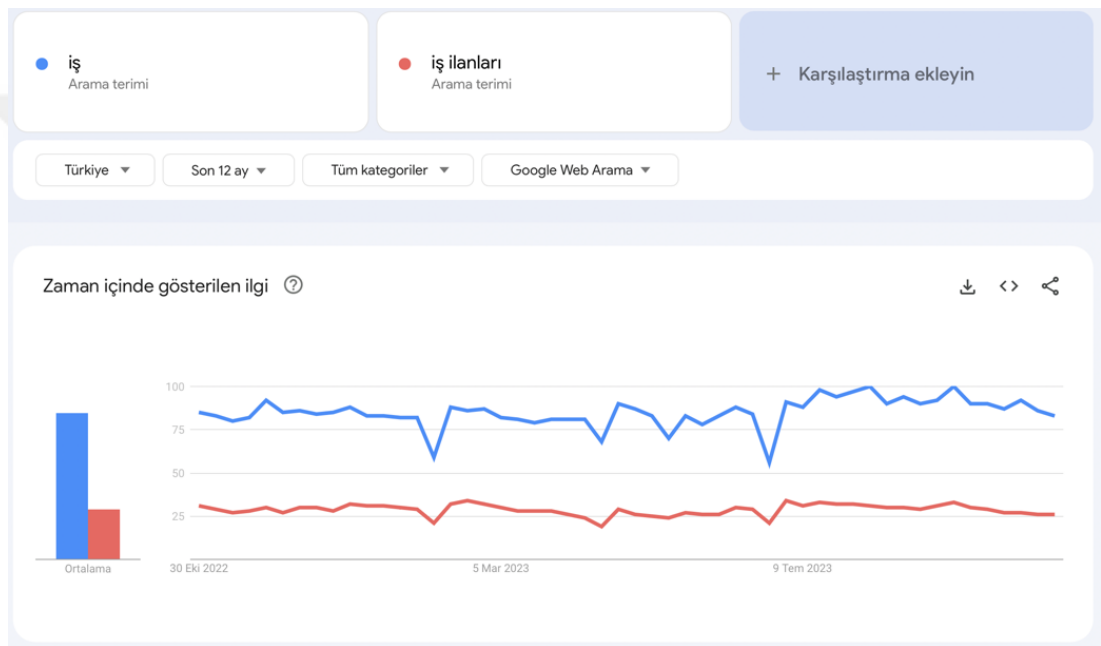
Şekil 24: Google Trends: "işsizlik" Sorgusunun Zaman İçinde Arama Yoğunluğu (23 Ekim 2023 – 30 Ekim 2023)



Kaynak: GT, 2023.

Şekil 25 'de, son bir yıl içinde Türkiye 'de Google Web Arama filtresi altında tüm kategorilerde gerçekleştirilen “iş” ve “iş ilanları” sorgularının grafiğini göstermektedir. Buna göre iki sorgunun zaman içindeki popülerlik değişimlerinin birbiriyle ilişkilendiği gözlenmektedir. Bu durum, Google kullanıcılarının “iş” ve “iş ilanları” gibi sorguları arama sıklığının benzerlik gösterdiğine işaret eder.

Şekil 25: Google Trends: “iş” ve “iş ilanları” Sorgularının Zaman İçinde Arama Yoğunluğu (30 Ekim 2022 – 30 Ekim 2023)



Kaynak: GT, 2023.

Google Trends ile ilgili bilinmesi gereken birçok detay bulunmaktadır. Bu detaylar aşağıda özetlenmiştir.

- GT verileri yalnızca grafiklendirmekle kalmaz csv. dosyası olarak indirip araştırmalarda kullanılabilmesine olanak tanır.
- GT arayüzünden verilere ulaşmak bazen zor ve uğraştırıcı olabilir. Bu nedenle GT verilerine çeşitli programlama dilleriyle erişilebilir. Örneğin R programlama dilinde yer alan “gtrendsR” paketi ile verilere erişilebilir. Ek olarak bu paket ile GT arayüzünün sahip olduğu tüm özelleştirmeler kullanılabilir.

- GT, verileri 0-100 arasında bir ölçekte sunmaktadır. Bu ölçekte 0, belirli bir sorgu için oldukça düşük olan ilgi seviyesini temsil etmektedir. 0 ya da 0 'a yakın bir değer, ilgili sorgunun neredeyse hiç aranmadığı anlamına gelir. 100 değeri ise belirli bir sorgu için oldukça yüksek bir ilgi olduğunu, ilgili dönemde en popüler sorgulardan biri olduğunu ifade etmektedir. 0 ile 100 arasındaki tüm değerler de bu ölçeğe göre değerlendirilmektedir. Örneğin 50 değeri bir sorgunun belirli bir dönemde ortalama seviyede bir popülerliğe sahip olduğu fikrini vermektedir.

Son zamanlarda GT 'in araştırma çalışmalarında kullanımı giderek yaygınlaşmıştır. Zaman içindeki eğilimleri tespit edebilmek için trend analizi, bir bölgedeki davranışları anlamak için coğrafi analiz, beklenmeyen değişiklikler için anomali tespiti, çeşitli işletmelerin pazar payını artırmak ve hedef kitleyi tespit etmek için tercih ettiği rekabet analizi, içerik stratejileri ve otomatik etiketleme gibi pek çok alanda kullanılarak araştırmacıların sık tercih ettiği araçlardan birisi haline gelen GT, geniş uygulama yelpazesi sayesinde makine öğrenmesi ile birleştirildiğinde de analiz çalışmaları için güçlü bir araç olarak kullanılabilir.

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. ÇALIŞMANIN AMACI VE KAPSAMI

Teknolojinin gelişimiyle, geleneksel bilgiye erişim çeşitlerinin tercih edilebilirliği azalmış ve dünyadaki internet kullanım oranı büyük ölçüde artmıştır. Özellikle TÜİK 'in 29 Ağustos 2023 tarihinde yayınladığı Hanehalkı Bilişim Teknolojileri Kullanım Araştırması 'na göre, Türkiye 'de internet erişimi olan hane oranı %95,5 iken, internet kullanıcıların oranı %87,1 'e çıkmıştır (TÜİK, 2023). Bu oran, Türkiye 'de internet kullanımının günlük yaşantıda önemli bir yere sahip olduğunu göstermektedir. İnternetin bilgiye erişim için hedeflere ulaşmayı hızlandıran önemli bir araç oluşu bunun temel nedenidir. Kullanıcıların bilgiye erişim hedefiyle internette gerçekleştirdiği aramaların ve diğer çeşitli aktivitelerin bıraktığı izler büyük çapta verilere dönüşmektedir. Bu nedenle, sürekli güncellenen ve yenilenen bu veriler, kapsamın doğru çizilmesi halinde işgücü göstergelerinin analiz edildiği araştırmalarda tercih edilebilir.

Örneğin iş sahibi olmayan kişiler, iş aramak ya da işsizlik ödeneğinden faydalanabilmek amacıyla kurum ve kuruluşlara fiilen gitmek yerine internette belli başlı aramalar gerçekleştirebilir. Bu nedenle internet kullanımı, işsizlik dönemlerinde çeşitli fırsatlara ulaşabilmek adına önem taşımaktadır. Kullanıcıların arama sorguları için 0-100 ölçeğinde endeks sunan GT, işsiz bireylerin internet aktivitelerinin analiz edilebilmesi yönünde tercih edilebilecek araçlara iyi bir örnektir. Ayrıca, hanehalkı iş gücü anketlerine dayalı olarak hesaplanan ve resmi kaynaklar tarafından detaylı incelemeler sonucu yayınlanan makroekonomik göstergelerle birlikte GT verilerini kullanarak işsizlik oranı için güçlü öngörülerde bulunulabilir (Chadwick ve Şengül, 2015: 4-5; Bolivar ve diğerleri, 2019: 2-3; Şentürk, 2022: 230-233).

Bu çalışmanın temel hedefleri arasında, GT verileri ile oluşturulan modellerin klasik AR(1), AR(2) ve AR(3) otoregresif modellerine kıyasla işsizlik oranını daha etkin bir şekilde tahmin edip etmediğini tespit etmek ve faydalanılan makroekonomik değişkenlerle birlikte makine öğrenme tekniklerinin performansını arttırmada katkı sağlayıp sağlamadığını incelemektir. Diğer benzer çalışmalardan ayrılan

hedeflerinden biri ise farklı frekanstaki veri etkilerini incelemektir. Günlük, haftalık ve aylık ölçeklerdeki GT verilerinin ayrı ayrı veya aynı anda modele dahil edilmesinin, model performansı üzerindeki etkisini değerlendirmektir. Ek olarak, işsizlik oranı çalışmalarında sıklıkla kullanılan makroekonomik göstergelerin de modellere eklenmesiyle GT verilerinin işsizlik oranı tahminindeki etkisinin daha iyi anlaşılması ve bu verilerin, işsizlik oranı tahmin performansına katkı sağlayıp sağlamadığının daha detaylı bir şekilde incelenmesi amaçlanmıştır. Çalışma kapsamında KNN, LASSO, Ridge, ElasticNet, XGBoost, GBDT, SVR, DT ve RF makine öğrenme algoritmaları kullanılmıştır. Bununla birlikte, her bir model için hem mevcut tüm değişkenlerle hem de değişken seçiminin performans üzerindeki etkisini görmek amacıyla LASSO modeliyle değişken seçimi yapılmış ve model performansları bu bağlamda karşılaştırılmıştır.

3.2. LİTERATÜR TARAMASI

Askatas ve Zimmermann (2009) çalışmasında, henüz internetin ekonometrik analizlerde yaygın olarak kullanılmadığı dönemde ekonomik davranışları tahmin etmek için belirli Google verilerini nedensellik yaklaşımı ve hata düzeltme modeli ile kullanmıştır. Bayesyen bilgi kriterine (BIC) dayalı bulgular basit modellerin karmaşık modellere göre çok daha iyi sonuçlar verdiğini tespit etmiştir. Google arama sorguları ve Almanya 'daki işsizlik oranları arasında güçlü bir ilişkinin olduğu tespit edilmiştir. Bu çalışma, işgücü göstergeleri için tercih edilen geleneksel tahmin yöntemlerine alternatif bir yol sunmuş ve değişkenlik gösteren ekonomik koşullarda tahmin yapılmasına yardımcı olmuştur.

Choi ve Varian (2011) çalışmasında Google arama sorgularını kullanarak çeşitli ekonomik göstergeler için tahminlerde bulunmuştur. Yöntemleri detaylandırırken ABD Nüfus Sayım Bürosu 'nun "Aylık Perakende Satışları" raporunda yer alan verileri kullanarak otomotivle ilgili kategorilerdeki Google aramaları otomobil satış tahminleri için kullanmıştır. ABD Çalışma Bakanlığı 'nın her hafta bir önceki hafta için işsizlik yardımına başvuru sayısını açıkladığı raporunda sunulan verilerle işsizlik talepleri tahmini gerçekleştirmiştir. Ek olarak Avustralya için Roy Morgan Tüketici Güven Endeksi kullanılarak tüketici güvenini tahminlemek

hedeflenmiştir. Google trends verileri içeren basit mevsimsel AR modellerinin bu tahminleri içermeyen modellere kıyasla performans artışının sebep olduğu gösterilmiştir. “Şimdiki zamanı tahminlemek” şeklinde nitelendirilen bir yaklaşım sunan bu çalışma, Google sorgu endeksinin ekonomik göstergelerle çoğunlukla pozitif korelasyona sahip olduğunu ve kısa vadeli tahminler için faydalı olduğunu bulmuştur.

Tuhkuri (2015) çalışmasında Google arama sorgularının yakın gelecekte işsizliği tahminleme yeteneğini araştırmıştır. Bu modeli oluştururken Amerika Birleşik Devletleri (ABD) için Google arama sorguları verilerini kullanmıştır. Yaklaşık 35 milyon adet “işsizlik yardımı” kelimesi ile ilgili sorguların kullanılarak geliştirilen bir endeks kullanmıştır. Bu endeks ilgili kelimelerin boolean arama operatörü tekniğiyle birleştirilmesi ile elde edilen “Google Endeksi” değişkenidir. Google endeksinin, resmi olarak yayınlanmakta olan işgücü istatistik verilerini içeren karmaşık bir modelle kıyaslamıştır. Google arama sorguları verileri üzerinde tanımlayıcı çapraz korelasyon analizi gerçekleştirmiş ve Granger nedensellik testleri gerçekleştirmiştir. Elde ettiği bulguların dayanıklılığını tespit etmek için eyalet bazlı değişkenlikleri göz önünde bulundurarak farklı arama terimlerini de çalışmaya dahil etmiştir. Çalışma sonucunda Google arama sorguları ile çalışma kapsamındaki dönemin işsizlik oranı arasında önemli bir korelasyon tespit edilmiş ve Google arama sorgularının kısa vadeli tahminlemede oldukça faydalı olduğu ortaya çıkmıştır.

Chadwick ve Şengül (2015) çalışmasında Türkiye için Google arama sorgularının aylık bazda tarım dışı işsizlik oranını tahminlemeye olan etkisini ve faydalarını araştırmıştır. Türkiye ’de resmi kaynaklar tarafından aylık olarak yayınlanan işsizlik oranının üç aylık bir gecikmeye sahip olma özelliğini göz önünde bulundurarak, Ocak 2005 ile Ekim 2011 arasındaki verileri kapsama dahil ederek, lineer regresyon modelleri ve bayesyen model ortalaması yöntemlerini uygulamışlardır. Temel bir model olarak otoregresif modelle karşılaştırıldığında, Google arama sorgularının ana bileşenlerini içeren en iyi model, RMSE ölçütü göz önünde bulundurularak temel modelden örneklem içi %47.7 ve örneklem dışı %38.4 daha doğru tahminler sağladığı ortaya çıkmıştır. Bu da Google arama sorgularının makroekonomik göstergeleri tahminlerken oldukça değerli sonuçlar ortaya sunduğunu göstermiştir.

Baker ve Fradkin (2017) çalışmasında ABD ’deki internet kullanıcılarının “işler” kelimesi ile ilişkili arama verilerine dayalı “Google İş Arama Endeksi” (GJSI) geliştirmiştir. Geliştirilen bu endeksin yüksek frekansa sahip olması, gerçek zamanlı olması ve coğrafi olarak ayrıştırılabilmesi sayesinde işsizlik sigortası politikalarındaki değişiklikleri 2008-2014 dönemi için inceleme olanağı sunmuştur. Çalışma sonucunda GJSI ile iş piyasası koşulları arasında pozitif ve önemli ölçüde korelasyon elde edilmiştir. Bu da Google aramalarının makroekonomik göstergelere karşı duyarlı olduğu ve ekonomik çalışmalarda daha sık kullanılması gerektiğinin önemini vurgulamıştır.

Castelnuovo ve Tran (2017) çalışmasında ABD ve Avustralya için GT verilerine dayalı belirsizlik endeksi geliştirmişlerdir. Bu endeks belirsizlikle ilişkili olabilecek kelimelerin sıklığına dayalı oluşturulmuştur. Bu kelimeler Federal Rezerv ’in Bej kitabı ve Avustralya Rezerv Bankası ’nın sunduğu Para Politikası Bildirisi gibi ekonomik yayınlarında sıkça yer alan kelimeler içerisinden tercih edilmiştir. GT üzerinden arama sıklıkları tespit edilerek iki ülke için oluşturulan bu endekslerin farklı belirsizlik ölçümleri ile pozitif korelasyona sahip olduğu görülmüştür. Vektör Otoregresyon (VAR) modelleri kullanılarak belirsizlik sıçramalarının makroekonomik göstergelerde sebep olduğu etkiler gözlemlenmiştir. Çalışma sonucunda ABD ’deki belirsizlik sıçramalarının Avustralya ’ya göre daha belirgin olduğu görülmüştür.

Bolívar ve diğerleri (2019) çalışmasında yaklaşık 3 aylık bir gecikmeyle yayınlanan Türkiye işsizlik verilerini anlık tahminlemeyi amaçlamıştır. 2004-2019 periyodu için Google correlate aracılığıyla 75 farklı sorgu elde edilmiş ve bu verileri içeren dinamik bir model oluşturulmuştur. Model geliştirilirken Türkiye Cumhuriyeti Merkez Bankası (TCMB) tarafından aylık olarak yayınlanan sanayi üretimi, elektrik tüketimi artışı ve kapasite kullanım oranı makroekonomik değişkenleri çalışmaya dahil edilmiş, bağımlı değişken olarak işsizlik oranı kullanılmıştır. Bu değişken, mevsim etkisinden arındırılmış olarak TÜİK ’in hane halkı iş gücü istatistikleri raporundan elde edilmiştir. Model belirsizlikleriyle mücadele edebilmek için değişken seçiminde bayesyen model ortalaması kullanılmış, işsizlikle ilişkili çeşitli modellerin anlık tahmin performansı test edilmiştir. 2007-2019 periyodu için bütünleşik otoregresif hareketli ortalama (ARIMA) temel modeli ile karşılaştırılan diğer modeller

içerisinden ridge regresyonun en iyi sonucu sunmuş ve Google arama sorgularından “iş bulma”, “işsizlik yardımı” ve “işsizlik sigortası” değişkenlerinin en yüksek açıklayıcılığa sahip olduğu görülmüştür.

Borup ve Schütte (2020) çalışmasında ABD ’de 2004-2019 yılları için Google arama verileriyle istihdam büyüme oranını yakın ve uzak zaman için tahminlemeyi amaçlamıştır. Ana terim olarak diğer çalışmalara benzer biçimde “işler” kelimesini temel almıştır. Bu ana terim, GT ’nin sunduğu “ilgili sorgular” seçeneği ile çoğaltılarak 172 farklı değişken yani kelime elde edilmiştir. Elde edilen bu değişkenlerle çeşitli modeller oluşturulmuştur. Verilerde karmaşık bir yapı olması nedeniyle katsayıları sıkıştırarak/küçültürerek ElasticNet değişken seçimi tekniğini ve RF algoritmasını tercih etmiştir. Araştırma sonucunda 24 farklı eyalet için oldukça kayda değer R^2 sonuçları elde edilmiştir ve arama terimlerindeki çeşitlilik sayesinde tek arama sorgusuna dayalı veya birden çok makroekonomik göstergesi baz alan modellere kıyasla oldukça başarılı olmuştur. Bu çalışma istihdam büyüme oranını sektörel sınıflara veya eyaletlere bölerek tahminleyebilme olanağı literatüre kazandırılmıştır.

Woloszko (2020) çalışmasında GT arama verilerini kullanarak 46 OECD ve G20 ülkesi için ekonomik faaliyetleri tahminlemeyi hedeflemektedir. Klasik ekonomik göstergelerin yayınlanmasında yaşanan gecikmelerin önüne geçerek güncel bilgilere erişim kolaylığını hedefleyen bu çalışmada GT gibi yüksek frekanslı göstergelerin kullanımını öne çıkarmıştır. Haftalık gayrisafi yurtiçi hasıla (GDP) büyüme oranını tahminlemek için geliştirilen model ile arama sıklıklarının ekonomik hareketlerle olan bağlantısı, doğrusal olmayan ilişkileri yakalayabilen yapay sinir ağları kullanılarak her ülke için modellenmiştir. Pseudo-real time simülasyonlar kullanılarak model performansları karşılaştırılmıştır. GT verilerinin ekonomik hareketleri 46 ülkede de doğru şekilde tahminlediği gözlemlenmiştir.

Caperna ve diğerleri (2020) çalışmasında COVID-19 pandemisinin yol açtığı ekonomik hareketlilikleri ve en önemlisi de işsizlik artışı sorununun etkilerini araştırmıştır. Çalışmada Avrupa Birliği üyesi 27 ülke için GT verileri kullanılmıştır. GT arayüzünün sunduğu “ilgili konular” aracını kullanarak iki binden fazla kelime elde edilmiştir. Bu kelimelerle ülkelerin aylık işsizlik oranlarını tahminlemiştir. Zaman serileri üzerinde şimdi tahmini (nowcasting) gerçekleştirilmiş ve Google verilerinin işsizliği öngörü performansı değerlendirilmiştir. Aynı zamanda RF değişken seçim

yöntemleri kullanarak işsizlik oranının tahmin doğruluğunu artıran kelimeleri tespit etmeye odaklanmıştır. AB hükümetlerince virüsün yayılımını önlemek için alınan karantina kararlarının sonuçlarını tespit etmek için “farklılıklar arasındaki fark” (Difference-in-Differences veya DiD) yöntemi kullanılmıştır. Karantina kararları sonucunda işsizlikle ilgili GT aramalarında pandemi öncesine göre yaklaşık %30 ’luk bir artış gözlemlenmiştir.

Şentürk (2022) çalışmasında Google arama verilerinin Türkiye ’deki işsizlik oranının tahmin modellerine eklenmesinin, bu modellerin tahmin performansını iyileştirip iyileştirmediğini incelemiştir. Tahmin performanslarını karşılaştırmak adına ARIMA ve açıklayıcı değişkenlerle bütünleşik otoregresif hareketli ortalama zaman serisi modelleri kullanılmıştır. Veri erişimi aşamasında gerçekleşebilecek değişkenlikler ve veri kullanımında yaşanabilecek sınırların vurgulandığı bu çalışma ekonomik tahmin yöntemlerinin geliştirilmesine katkı sağlamış ve GT çevrimiçi açık kaynağının alternatif bir veri kaynağı olduğunu öne sürmüştür.

Borup, Rapach ve Schutte (2022) çalışmasında, işsizlikle ilgili 88 terim için günlük GT arama hacmi verilerine dayalı haftalık işsizlik sigortası taleplerini tahminlemeyi hedeflemiştir. İşsizlik sigortası taleplerine ilişkin güncel ve geriye dönük tahminler oluştururken karma frekanslı makine öğrenmesi (MFML) yaklaşımını öne sürmüştür. LASSO, elastik net ve yapay sinir ağlarını baz alarak, doğrusal ve doğrusal olmayan modellerle birçok tahmin oluşturmuştur. Günlük GT verilerini içeren modellerin, içermeyen modellere göre büyük bir farkla daha iyi performans gösterdiğini tespit etmiştir. Özellikle COVID-19 pandemisi esnasında bu verilerin faydası vurgulanmıştır. Çeşitli makine öğrenmesi modelleri ve geleneksel istatistiki yöntemler arasındaki kıyaslamalar sonucu makine öğrenmesi yöntemleri öngörü performansı anlamında öne çıkmıştır.

Grybauskas ve diğerleri (2023) çalışmasında, işsizlik başvurularını tahmin etmek için GT verilerini kullanmayı amaçlamıştır. Sınırlı sayıda kelime yerine altı farklı bağlama sahip anahtar kelimeler (iş arama, ruh sağlığı, şiddet ve istismar, boş zaman araştırmaları, tüketim ve yaşam tarzı, afetler) tercih etmiştir. İşsizlik oranları tahmininde genelde tercih edilen dinamik faktör veya ARIMA gibi yöntemlerin dışına çıkarak uzun-kısa vadeli bellek (LSTM) yapay sinir ağlarına odaklanmıştır. Çalışmada geleceği tahmin etmek yerine mevcut durumu tahminleme amacıyla LSTM ’i baz alan

şimdi tahmin (nowcasting) modeli oluşturulmuştur. Temel LSTM modeli dışında kodlayıcı-çözücü (encoder-decoder) LSTM modeli de kullanmıştır. COVID-19 döneminde “işsizlik”, “online poker” ve “online kumarhane” gibi terimlerin arama sıklığında artış gözlemlenmiştir. Tahmin hatalarının azalmasında farklı boyutlardaki arama kelimelerinin kullanımının önemi vurgulanmıştır.

3.3. DENEY TASARIMI

Araştırma sürecinin yapıtaşı olan deney tasarımının geliştirilmesi, çalışmanın başarıya ulaşması için en temel unsurdur. Kısacası probleme ilişkin optimal çözümü bulmak için iyi bir deney tasarımı sunulmalıdır. Dolayısıyla bu çalışmada, işsizlik oranı öngörüsünde makine öğrenme yöntemlerini farklı frekanslı GT 'den elde edilmiş arama terimleriyle modellemek üzere bir deney tasarımı hazırlanmıştır. Bu kapsamda veri setlerinin elde edilmesinden nihai modele kadar gerçekleşen aşamalar detaylandırılmıştır.

Bu araştırmada, işsizlik oranı öngörüsünde bulunmak amacıyla makroekonomik değişkenlerle modellemeler gerçekleştirilmiş ve GT değişkenlerinin tahmin yeteneğini artırıp artırmadığını tespit etmek adına makroekonomik değişkenlerin, GT değişkenleriyle oluşturulan farklı kombinasyonlarında modellemeler gerçekleştirilmiştir. Tablo 5 'de listelenen 5 farklı model detaylı bir şekilde incelenmiştir. Her model için KNN, LASSO, Ridge, ElasticNet, XGBoost, GBDT, SVR, DT ve RF makine öğrenme tekniklerinden yararlanılmıştır. Son olarak, her model LASSO değişken seçim tekniği ile tekrarlanarak değişken seçim tekniğinin modeller üzerindeki etkisi araştırılmıştır.

Tablo 5: Uygulama Modelleri

Uygulama Modelleri	Çalışmada Kullanılan Veri Setlerine İlişkin Modeller
1. Model	Makroekonomik Değişkenlerle İşsizlik Oranı Öngörüsü
2. Model	Günlük GT Değişkenleri ve Makroekonomik Değişkenlerle İşsizlik Oranı Öngörüsü
3. Model	Haftalık GT Değişkenleri ve Makroekonomik Değişkenlerle İşsizlik Oranı Öngörüsü
4. Model	Aylık GT Değişkenleri ve Makroekonomik Değişkenlerle İşsizlik Oranı Öngörüsü
5. Model	Günlük, Haftalık, Aylık GT Değişkenleri ve Makroekonomik Değişkenlerle İşsizlik Oranı Öngörüsü

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

3.3.1. Veri Toplama

GT platformunun sunduğu veriler 2004 'den itibaren başlamaktadır. Ancak, genişbant internet erişiminin yaygınlaşmaya başladığı, sosyal medya kullanımının arttığı ve kişilerin iş aramak için sıklıkla arama motoru kullandığı senelerin 2010 'dan itibaren yükselişe geçtiği düşünülmektedir. Dolayısıyla GT 'nin sunduğu arama sıklıkları ve resmi kaynakların yayınladığı makroekonomik veriler, 2010 yılının 1. ayı ile 2023 yılının 5. ayına kadar olan dönemini kapsayarak analize dahil edilmiştir.

3.3.1.1. Google Trends Verileri

GT sayesinde internet kullanıcılarının arattığı kelimelerin coğrafi konum, zaman, dil ve kategori kriterlerine dayalı göreceli arama sıklıklarını elde etmek mümkündür. Ancak işsizlik oranını öngörülerken anlamlı sonuçlar elde etmek için temel anahtar kelimeleri doğru seçmek önem taşımaktadır. Çalışma kapsamında işsiz bireylerin arama motorunda en sık aratabileceği temel kelime olarak “iş” seçilmiştir. Ardından GT arayüzünde “iş” kelimesi sorgulanmıştır. GT arayüzünün sunduğu en önemli özelliklerden biri olan “ilgili sorgular” aracılığıyla “iş” kelimesi ile en ilişkili 25 sorgu elde edilmiştir. Ancak bu araç aratılan kelime ile ilişkili fakat çalışma amacına uygun olmayan sorgular da sunabilmektedir. Bu nedenle aracın ürettiği sorgu listesi içerisinde çalışma kapsamında kullanılmayacak olanların elenmesi

gerekmektedir. Örneğin “iş cep giriř” ya da “iş güvenlięi” gibi sorgular işsizlik oranı bağlamında kapsamın dışında kalmaktadır. Bu nedenle listelenen 25 sorgu içerisinde “işkur”, “iş ilanı”, “iş ilanları” ve “iş başvurusu” kelimeleri çalışmaya en uygun olan kelimeler olarak seçilmiştir. Seçilen sorgular içerisinde yer alan “iş ilanı” ve “iş ilanları” aynı bağlama sahip olmalarına rağmen biçimsel olarak farklı kelimelerdir; bu nedenle tekil ya da çoęul ek almış hallerinin, tahmin performansı üzerinde etkisi olabileceęi için her iki kelime de araştırmaya dahil edilmiştir (Borup ve Schütte, 2022:186–200).

Anahtar kelime sayısını artırmak amacıyla Bolívar ve diğerlerinin (2019) Türkiye ’de işsizlik verilerini GT ile tahminlemeyi amaçladığı çalışmasında yer alan 75 kelimeden uygun görülen 12 adet kelime seçilmiştir. Bu kelimeler řu şekildedir: “işsizlik sigortası”, “işsizlik maaşı”, “yenibiriş”, “cv”, “part time”, “iş arıyorum”, “full time”, “iş bulma”, “örnek cv”, “özgeçmiş”, “eleman.net”, “kariyer.net” ve “secretcv”.

Kelimeler içerisinde yer alan “işsizlik maaşı”, GT arayüzünde sorgulanarak “ilgili sorgular” aracılığıyla benzer kelimeler elde edilmiştir. Bu kelimeler řu şekildedir: “işsizlik kredi”, “işsizlik desteęi”, “işsizlik ödeneęi”, “işsizlik başvurusu”, “işsizlik yardımı” ve “işsizlik fonu”. Son olarak Türkiye ’de oldukça popüler olan iş arama sitesi “linkedin” de listeye eklenerek Tablo 6 ’da yer alan 25 anahtar kelime ile sürece devam edilmiştir.

Tablo 6: Google Trends Arama Sorguları

<i>iş</i>	<i>iş başvurusu</i>
<i>işsizlik kredi</i>	<i>işsizlik fonu</i>
<i>işsizlik desteęi</i>	<i>iş ilanları</i>
<i>işsizlik ödeneęi</i>	<i>iş ilanı</i>
<i>işsizlik yardımı</i>	<i>iş bulma</i>
<i>işsizlik başvurusu</i>	<i>örnek cv</i>
<i>işsizlik sigortası</i>	<i>özgeçmiş</i>
<i>işsizlik maaşı</i>	<i>işkur</i>
<i>yenibiriş</i>	<i>eleman.net</i>
<i>cv</i>	<i>kariyer.net</i>
<i>part time</i>	<i>secretcv</i>
<i>iş arıyorum</i>	<i>linkedin</i>
<i>full time</i>	

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Arama sorgularının ölçeklendirilmiş frekanslarını elde etmek için R programlama dilinin “gtrendsR” kütüphanesi tercih edilmiştir. Bu paket belirli sorgular için verileri almaya, analiz etmeye ve görselleştirmeye yardımcı olmaktadır. Araştırma kapsamında yalnızca verileri arayüze göre daha hızlı çekebilmek adına yararlanılmıştır. “gtrendsR” kütüphanesinde yer alan “gtrends” metodu sonuçları coğrafi konum, dil, zaman gibi unsurlar bakımından spesifikleştirerek sunmaktadır. Ancak “gtrends” metodunun en fazla 5 kelime sorgulama limiti nedeniyle 25 kelime Tablo 7 ’deki gibi 5 farklı gruba ayrılmıştır. Bu kapsamda, 5 farklı kelime grubu için günlük, haftalık ve aylık ölçeklenmiş arama sıklığı serileri elde edilmiştir.

Tablo 7: Veri Çekme İşlemi için Oluşturulan Değişken Grupları

Kelime Grubu	<i>Değişken 1</i>	<i>Değişken 2</i>	<i>Değişken 3</i>	<i>Değişken 4</i>	<i>Değişken 5</i>
1. Grup	<i>iş</i>	<i>işsizlik kredi</i>	<i>işsizlik desteği</i>	<i>işsizlik ödeneği</i>	<i>işsizlik yardımı</i>
2. Grup	<i>işsizlik başvurusu</i>	<i>işsizlik sigortası</i>	<i>işsizlik maaşı</i>	<i>iş başvurusu</i>	<i>işsizlik fonu</i>
3. Grup	<i>iş ilanları</i>	<i>iş ilanı</i>	<i>iş bulma</i>	<i>örnek cv</i>	<i>özgeçmiş</i>
4. Grup	<i>işkur</i>	<i>yenibiriş</i>	<i>eleman.net</i>	<i>cv</i>	<i>kariyer.net</i>
5. Grup	<i>part time</i>	<i>secretcv</i>	<i>linkedin</i>	<i>iş arıyorum</i>	<i>full time</i>

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

3.3.1.1.1. Günlük Verilerin Elde Edilmesi

Günlük frekansları elde etmek için gtrends metoduna en fazla 8 aylık periyotlar girilmelidir. 8 aydan fazla bir periyot girildiğinde gtrends metodu günlük veriler yerine haftalık ya da aylık veriler verecektir. Bu nedenle çalışma kapsamındaki dönem 8 aylık dilimlere ayrıştırılarak 5 kelime grubu, 25 kelime için sorgulamalar yapılmış ve bu kelimelerin günlük frekansları elde edilmiştir.

3.3.1.1.2. Haftalık Verilerin Elde Edilmesi

Haftalık frekansları elde etmek için gtrends metoduna en fazla 36 aylık periyotlar girilmelidir. 36 aydan fazla bir periyot girildiğinde gtrends metodu haftalık veriler yerine aylık veriler verecektir. Bu nedenle çalışma kapsamındaki dönem 36 aylık dilimlere ayrıştırılarak 5 kelime grubu, 25 kelime için sorgulamalar yapılmış ve bu kelimelerin haftalık frekansları elde edilmiştir.

3.3.1.1.3. Aylık Verilerin Elde Edilmesi

Aylık frekansları elde etmek için gtrends metoduna 36 aydan fazla periyotlar girilmelidir. Bu nedenle çalışma kapsamındaki dönemi tamamen kapsayan bir sorgulama gerçekleştirerek 5 kelime grubu, 25 kelime için sorgulamalar yapılmış ve bu kelimelerin aylık frekansları elde edilmiştir.

3.3.1.1.4. Günlük, Haftalık, Aylık Verilerin Konsepti

Beş kelime grubu için sorgulama işleminde günlük, haftalık ve aylık veriler periyotlara ayrılırken her bir periyodun bir önceki periyodun son gözlemine sahip olmasına dikkat edilmiştir. Çünkü belirli bir gün için farklı zamanlarda elde edilen frekans değerleri kendi içinde küçük değişiklikler gösterebilmektedir. Fakat verilerin sağlıklı sonuç üretebilmesi için bu değişkenlik göz önünde bulundurularak çeşitli hesaplamalarla verilerin düzenlenmesi gerekmektedir. Örneğin günlük veri çekiminde ilk periyot 1 Ocak 2010 – 1 Eylül 2010 tarih aralığını kapsamaktadır. Bu nedenle ikinci periyot kesişen bir tarihe sahip olmalı ve 1 Eylül 2010 – 1 Mayıs 2011 tarih aralığını kapsamalıdır. Bu sayede belirli bir gün için elde edilen bu farklı frekans değerleri oranlanarak veri setleri uç uca birleştirilebilir ve daha sağlıklı veri setine ulaşılabilir.

Bu bağlamda belirlenen veri periyotları aşağıdaki Tablo 8 'de yer almaktadır. GT verileri ilgili zaman aralığındaki coğrafi bölgede gerçekleştirilen tüm Google aramaları arasından bir örnek olarak hesaplandığından, 0 ile 100 arası bir ölçekte sunulan GT verileri indirildikleri gün ve internet protokol adresine (IP) göre değişiklik gösterebilmektedir. Bu nedenle örneklemeden kaynaklı olarak endeks değişkenliğini

minimize etmek amacıyla veri çekme işlemi her değişken için on gün boyunca tekrarlanmıştır. Kesişim tarihi belirlenmesi ve veri çekme işleminin on farklı gün tekrarlanması “Verilerin Hazırlanması” başlığında detaylandırılmıştır.

Tablo 8: Farklı Frekanslı Arama Terimlerini Oluşturmak İçin Veri Çekme Periyotları

Günlük Veri Periyotları	Haftalık Veri Periyotları	Aylık Veri Periyotları
<i>01.01.2010 - 01.09.2010</i>	<i>01.01.2010 - 01.01.2013</i>	<i>01.01.2010 - 30.04.2023</i>
<i>01.09.2010 - 01.05.2011</i>	<i>01.01.2013 - 01.01.2016</i>	
<i>01.05.2011 - 01.01.2012</i>	<i>01.01.2016 - 01.01.2019</i>	
<i>01.01.2012 - 01.09.2012</i>	<i>01.01.2019 - 30.04.2023</i>	
<i>01.09.2012 - 01.05.2013</i>		
<i>01.05.2013 - 01.01.2014</i>		
<i>01.01.2014 - 01.09.2014</i>		
<i>01.09.2014 - 01.05.2015</i>		
<i>01.05.2015 - 01.01.2016</i>		
<i>01.01.2016 - 01.09.2016</i>		
<i>01.09.2016 - 01.05.2017</i>		
<i>01.05.2017 - 01.01.2018</i>		
<i>01.01.2018 - 01.09.2018</i>		
<i>01.09.2018 - 01.05.2019</i>		
<i>01.05.2019 - 01.01.2020</i>		
<i>01.01.2020 - 01.09.2020</i>		
<i>01.09.2020 - 01.05.2021</i>		
<i>01.05.2021 - 01.01.2022</i>		
<i>01.01.2022 - 01.09.2022</i>		
<i>01.09.2022 - 01.05.2023</i>		
<i>01.05.2023 - 01.01.2023</i>		
<i>01.01.2023 - 30.04.2023</i>		

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

3.3.1.2. Makroekonomik Veriler

Sağlıklı bir işsizlik oranı tahmin modeli geliştirmek ve iş piyasasındaki eğilimleri anlamak için makroekonomik veriler de büyük önem taşımaktadır. Çünkü işsizlik kavramı pek çok ekonomik değişkenle ilişkili olabilmektedir. Ancak makroekonomik değişkenlerin kapsama dahil edilmesinin bir diğer nedeni, farklı kaynaklardan elde edilen verilerin farklı bilgi türleri sunarak analizi zenginleştirmesidir. Böylelikle tahmin modellerinin performansının artacağı düşünülmüştür. İşsizlik oranı öngörüsünde kullanılabilecek gayrisafi milli hasıla (GSMH) yüzde değişimi, reel efektif kur oranı, tüketici fiyat endeksi (TÜFE), reel faiz oranı, imalat sektöründeki reel ücret endeksi, nüfus sayısı, ekonomi güven endeksi (EGE), tüketici güven endeksi (TGE), istihdam oranı ve TÜFE bazlı reel efektif döviz kuru gibi pek çok değişken örnek gösterilebilir. Ancak araştırma kapsamında resmi kaynaklarca aylık dilimlerle sunulan mevsimsellikten arındırılmış işsizlik oranı, tüketici fiyat endeksi (TÜFE), ekonomik güven endeksi, 15+ yaş nüfus sayısı, istihdam sayısı ve TÜFE bazlı reel efektif döviz kuru değişkenleri tercih edilmiştir.

TÜFE verileri, her ayın üçünde saat 10:00 'da bir ay öncesi için yayınlanan TÜİK 'in "Tüketici Fiyat Endeksi ve Oranları" raporundan, ekonomik güven endeksi ay sonunda ilgili ayı da kapsayacak şekilde yayınlanan TÜİK 'in "Ekonomik güven endeksi ve alt endeksler" raporundan elde edilmiştir. 15+ yaş nüfus, istihdam sayısı ve işsizlik oranı değişkenleri ise yine TÜİK 'in ilgili ayı takip eden 40. gün, her ayın onunda (eğer haftasonuna denk gelirse ilk iş günü) saat 10:00 'da yayınladığı ve hanehalkı anketleri ile elde ettiği "İşgücü İstatistikleri" raporundan alınmıştır. TÜFE bazlı reel efektif döviz kuru değerleri ise her ay bir önceki ay için Türkiye Cumhuriyeti Merkez Bankası'nın yayınladığı raporlardan elde edilmiştir.

3.3.2. Veri Ön işleme

Verilerin analizde sağlıklı sonuçlara ulaşmaya yardımcı olabilmesi için çalışmaya uygun hale getirilmeleri gerekmektedir. Bu nedenle çalışmada elde edilen veriler, etkin bir hale getirilmek için pek çok aşamadan geçmiştir. Bu aşamalar bu başlıkta detaylandırılmıştır.

3.3.2.1. Veri Temizleme

Çalışma kapsamında kullanılan GT verilerinin analize uygun, sağlıklı bir hale getirilmesi araştırmanın geçerli ve doğru sonuçlar üretebilmesi açısından oldukça kritiktir. Gerçekleştirilen sorguya istinaden elde edilen arama sıklıkları 0 ile 100 arası bir ölçekte sunulmaktadır. Burada 0 olarak ifade edilen değer, ilgili tarihte ilgili konunun hiç popüler olmadığını ifade etmektedir. Bu nedenle elde edilen veriler içerisinde göz ardı edilemeyecek ölçüde 0 değerine sahip değişkenler, analizin istatistiksel anlamlılığını azaltarak çıktının güvenilirliğini sarsabilir. Bu, veri temizleme aşamasında dikkate alınması gereken ve çalışma amacına uygun olmayan sonuçlar doğurabilecek kritik bir durumdur. Veri hazırlama sürecinde bu tür değişkenler uygun bir biçimde ele alınmalıdır. Eksik, tutarsız ya da gereksiz değişkenlerin düzenlenmesi veya kapsamdan çıkarılması gerekmektedir. Bu nedenle verilerin doğru yorumlanabilmesi için, %5 ve üzerinde 0 değerine sahip olan “işsizlik başvurusu”, “full time”, “işsizlik kredisi”, “işsizlik desteği”, “işsizlik ödeneği”, “işsizlik yardımı”, “işsizlik fonu”, “iş bulma”, “örnek cv”, “yenibiriş”, “eleman.net”, “kariyer.net” değişkenleri kapsamdan çıkarılmıştır. Böylece Tablo 9 ’da verilen 13 değişken için elde edilen verilerle çalışmaya devam edilmiştir.

Tablo 9: Nihai GT Değişkenleri

<i>iş</i>
<i>iş ilanları</i>
<i>iş ilanı</i>
<i>özgeçmiş</i>
<i>işkur</i>
<i>iş başvurusu</i>
<i>işsizlik sigortası</i>
<i>işsizlik maaşı</i>
<i>cv</i>
<i>part time</i>
<i>iş arıyorum</i>
<i>secretcv</i>
<i>linkedin</i>

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Makroekonomik göstergelere dayalı veriler geniş çaplı değerleri ifade etmektedir ve sağlam kaynaklar tarafından toplanmış bilgilerden elde edilmiştir. Özellikle TÜİK 'den elde edilen ya da TÜİK 'den elde edilerek TCMB tarafından hesaplanmış verilerdir. Dolayısıyla veriler düzenli ve sistematik olarak toplanmaktadır. Toplanan veriler sürekli olarak güncel kontrollere tabi tutulmaktadır ve yenilenmektedir. Bu nedenle bu verilerde eksik, gereksiz ya da tutarsız herhangi bir veri veya veri grubu gözlemlenmemiştir. Dolayısıyla temizleme ya da düzeltme gereksinimi duyulmamıştır.

3.3.2.2. Verilerin Hazırlanması

İlk etapta her değişken için periyotlar halinde 10 gün boyunca çekilen verilerin düzenlenmesi gerekmektedir. Çünkü her bir periyot için onar adet veri setimiz bulunmaktadır. Yani yalnızca bir değişken için günlük, haftalık ve aylık ölçeklerde toplamda 27 periyot bulunması nedeniyle 27×10 veri seti söz konusudur. Birleştirme işlemi için periyot başına düşen 10 veri setinin ortalaması elde edilerek 1 veri seti haline getirilmiştir. Nihai durumda çalışmada kullanılacak GT veri seti adeti değişken başına 27 'de sabitlenmiştir.

Ayrıca, günlük ve haftalık verilerdeki periyot kısıtlamasından kaynaklanan parçalı veri çekme işlemi de zaman ve IP sorunundan etkilenmektedir. Bu nedenle verilere erişirken Tablo 8 'de görüldüğü gibi her bir periyot, kesişim tarihleri olacak şekilde kurgulanmıştır. Bu zaman aralıklarının birbirine bağlanması için belirlenen kesişim tarihi ise kesişim gözlemi elde etmeyi sağlamıştır. Kesişen bu gözlemler birbirine oranlanarak elde edilen değer, bir sonraki küme ile çarpılmış ve bu işlem periyotlar birbirine eklenip tamamlanana kadar tekrarlanmıştır. Böylece günlük ve haftalık ölçekte farklı zaman aralıklarındaki verilerin birleştirilmesi sağlanmıştır. Aylık verilerin tek periyot halinde indirilebilmesi nedeniyle bir önceki paragrafta bahsedildiği gibi yalnızca ortalamasının alınması yeterli olmuştur. Sonuç olarak günlük, haftalık, aylık veriler bir set haline getirilmiştir ve her değişken için üç farklı frekansta oluşturulmuş arama sıklığı serisi elde edilmiştir. Örnek olarak günlük ölçekte çekilen “iş” değişkeni için uygulanan adımlar aşağıda detaylandırılmıştır:

1. adım: Günlük verileri elde etmek için, 8 aylık periyodu geçmeyecek ve kesişim tarihi olacak şekilde 22 tarih aralığı belirlenmiştir.

2. adım: 22 tarih aralığı için 10 farklı gün veri çekme işlemi tekrarlanmıştır.

3. adım: 10 farklı çekimde elde edilen veri setlerinin ortalaması alınmıştır ve 22 periyot için bir veri kümesine ulaşılmıştır.

4. adım: Elde edilen 22 periyodun verileri kesişim tarihi baz alınarak birleştirilmiştir. Birleştirme işleminde birinci tarih aralığının son gözlemi ile ikinci tarih aralığının ilk gözlemi oranlanmıştır. Ardından ikinci tarih aralığının tüm gözlemleri elde edilen bu oranla çarpılmıştır. Ulaşılan bu ikinci tarih aralığı ile üçüncü tarih aralığı arasında da aynı hesaplamalar gerçekleştirilmiş ve her birleştirmede elde edilen küme üzerinden ilerlenmiştir. Bu işlem 22 periyot birleşene kadar tekrarlanmıştır.

3.3.3. Veri Setlerinin İstatistiksel Özellikleri

3.3.3.1. Google Trends Veri Seti Tanımlayıcı İstatistikleri

Tablo 10 'da GT aracılığıyla elde edilen günlük, haftalık ve aylık ölçeklerdeki değişkenlerin frekans değerlerinin istatistiksel bir özeti sunulmuştur. “N” sütunu veri setindeki toplam gözlem sayısını, “ORT” sütunu değişkenler için elde edilen ortalama frekans değerlerini, “SS” sütunu değişkenlerin standart sapmasını ve “MIN”, “MAX” sütunları ise gözlemler içerisinde yer alan en küçük ve en büyük frekansları ifade etmektedir. Aynı zamanda değişkenlerin Q1, Q2 ve “MED” olarak ifade edilen medyan değerleri sunulmuş, çarpıklık değerleri “ÇARP” ve son olarak basıklık değerleri “BAS” olarak ifade edilmiştir.

Tablo 10: Google Trends Değişkenlerinin Tanımlayıcı İstatistikleri

DEĞİŞKEN	N	ORT	SS	MIN	MAX	Q1	MED	Q3	ÇARP	BAS
günlük - cv	4868	10,58	3,99	0	42,14	7,56	10,08	12,58	0,90	1,66
haftalık - cv	697	13,68	3,74	4,99	31	10,50	13,31	14,98	0,76	1,03
aylık - cv	161	7,78	2,17	3	15,80	6,50	7	8,10	1,24	1,16
günlük - işkur	4868	35,86	22,10	2,11	230	17,60	37,54	49,53	0,90	1,66
haftalık - işkur	697	59,19	28,52	23,36	382,92	46	52,86	63,60	0,76	1,02
aylık - işkur	161	39,53	13,05	20,20	99,40	31,35	36	43,85	1,26	1,23
günlük - iş arıyorum	4868	5,52	4,41	0	30,30	2,44	4,21	7,13	0,90	1,66
haftalık - iş arıyorum	697	6,43	4,81	0,90	25,30	2,94	5,08	8,34	0,76	1,03
aylık - iş arıyorum	161	7,25	5,23	1	22,70	3,50	6,00	9	1,29	1,38
günlük - iş başvurusu	4868	25,43	12,89	11,18	114,24	36,12	44,16	53,68	0,90	1,66
haftalık - iş başvurusu	697	51,15	12,59	23,61	100	41,87	48,73	56,96	0,77	1,04
aylık - iş başvurusu	161	59,79	13,04	35,80	100	50,70	56,90	65,25	1,32	1,55
günlük - iş ilanı	4868	4,89	1,36	0,40	10,60	4	4,90	5,88	0,90	1,66
haftalık - iş ilanı	697	7,56	2,54	2,90	16,67	5,61	6,50	9,37	0,77	1,04
aylık - iş ilanı	161	3,77	1,12	2	7	3	3,50	4,70	1,35	1,67
günlük - iş ilanları	4868	65,29	15,69	18,10	106	53,81	65,69	76,15	0,90	1,66
haftalık - iş ilanları	697	76,98	14,80	25,56	113,78	68,03	76,70	86,56	0,77	1,05
aylık - iş ilanları	161	56,62	15,99	29,40	99,20	42,30	54,70	67,30	1,38	1,81
günlük - iş	4868	56,92	13,49	21,42	100	45,32	58,47	67,59	0,91	1,66
haftalık - iş	697	76,30	10,07	40,75	99,54	70,57	77,31	83,62	0,78	1,06
aylık - iş	161	79,59	8,98	51,50	99,40	74,20	80,60	86	1,42	2
günlük - part time	4868	8,39	3,21	0	25,58	6,20	8,22	10,39	0,91	1,67
haftalık - part time	697	10,27	3,33	2,56	22,36	7,90	10,03	12,42	0,78	1,08
aylık - part time	161	11,63	3,57	3	22,40	9	11	13,80	1,46	2,21
günlük - secretcv	4868	20,81	21,21	0,24	104,35	4,62	10,87	32,67	0,91	1,67
haftalık - secretcv	697	23,55	23,34	1,66	100	5,11	12,04	37,92	0,79	1,09
aylık - secretcv	161	26,43	26,15	2	98,80	5	13	43,15	1,50	2,42
günlük - işsizlik maaşı	4868	20,81	12,80	0	170	11,72	18,70	27,82	0,91	1,67
haftalık - işsizlik maaşı	697	24,56	12,98	5,80	125,74	13,55	23,64	32,23	0,79	1,12
aylık - işsizlik maaşı	161	28,80	13,79	9,80	89,30	15,85	29	38,40	1,54	2,67
günlük - işsizlik sigortası	4868	1,70	1,38	0	14,56	0,70	1,40	2,30	0,91	1,67
haftalık - işsizlik sigortası	697	1,96	0,68	0	5	1,49	1,80	2,25	0,80	1,15
aylık - işsizlik sigortası	161	1,95	0,76	1	4,10	1	2	2,10	1,58	2,88
günlük - linkedin	4868	33,47	15,62	0	85,55	23,48	35,12	42,93	0,91	1,67
haftalık - linkedin	697	43,93	18,35	2,5	87,61	38,25	45,90	52,78	0,80	1,17
aylık - linkedin	161	49,38	20,90	4	99	41,60	49,70	59,35	1,60	2,91
günlük - özgeçmiş	4868	2,54	1,25	0	14,75	1,63	2,28	3,26	0,91	1,67
haftalık - özgeçmiş	697	3,82	1,17	1,25	12,50	2,80	4	4,25	0,81	1,20
aylık - özgeçmiş	161	1,39	0,48	1	2	1	1	2	1,63	3,08

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Değişkenler için elde edilen tanımlayıcı istatistikler incelendiğinde, belirli bir değişkenin günlük, haftalık ve aylık ölçeklerdeki frekans değerlerinin farklılaşabildiği dikkat çekmektedir. Bu durumun temel nedeni, kullanıcıların mevsimsel etkiler veya toplumsal olaylar gibi faktörler altında arama alışkanlıklarının ve ilgi düzeylerinin zaman aralıklarına göre değişebilmesidir. Özellikle tabloda yer alan değişkenler içerisinde “aylık - iş” değişkeninin 79,59 ile en yüksek ortalama frekans değerine sahip olması “iş” kelimesinin aylık ölçekte Google arama motorunda diğer kelimelere göre belirgin bir biçimde popüler olduğunu ifade etmektedir. Ek olarak “haftalık-iş ilanları” ve “ haftalık - iş” değişkenleri de sırasıyla 76,98 ve 76,30 frekans ortalaması ile en yüksek değerlere sahiptir. “iş” kelimesinin haftalık ve aylık ölçekte frekans değerlerinin oldukça yüksek olması, genel olarak sürekli ve istikrarlı bir şekilde ilgi gördüğünü, internet kullanıcılarının bu konuda düzenli olarak bilgi almaya veya araştırma yapmaya ihtiyaç duyduklarını göstermektedir. Özellikle “haftalık - iş” ve “aylık - iş” değişkenlerinin Q1 değeri sırasıyla 70,57 ve 74,20 olarak elde edilmiştir. Değişkenlerin veri seti içerisindeki alt çeyrek değerlerinin yüksek olması ve medyan ile Q3 değerlerinin bu durumu takip etmesi, bu değişkenlerin popüler arama sorguları olduğunu göstermektedir. Ayrıca oldukça düşük Q3 değerine sahip olan, örneğin “aylık-özgeçmiş” ve “aylık-işsizlik sigortası” değişkenlerinin internet kullanıcıları için pek popüler olmayan arama sorguları olduğu anlaşılmaktadır.

Ancak ilginç bir bulgu da verileri çekme işleminde, farklı biçimlerde fakat aynı bağlama sahip olan ve tekil-çoğul kelimelerin farklı sonuçlar sunup sunmadığını görme amacıyla analize dahil edilen “iş ilanı” ve “iş ilanları” değişkenlerinin ortalama frekans değerleri arasındaki büyük farktır. Örneğin “haftalık - iş ilanı” değişkeninin ortalama frekans değeri 7,56 iken “haftalık - iş ilanları” değişkeninin 76,98 ’dir. Bu durum, iki değişkenin günlük ve aylık ölçekteki değerleri için de geçerlidir. Aradaki bu tartışılmaz fark, arama terimlerindeki küçük sayılabilecek biçimsel farklılıkların bambaşka frekans sonuçları doğurabileceğini ispatlar niteliktedir. Bu da internet kullanıcılarının sorgulama yaparken, spesifik bir iş ilanı aramak yerine birden çok iş ilanına ulaşmak istediği anlamını verebilir.

Öte yandan, ortalama frekans değeri 2 ’nin altında kalan “günlük-işsizlik sigortası”, “haftalık-işsizlik sigortası” ve “aylık-işsizlik sigortası” değişkenleri, “işsizlik sigortası” kelimesinin tüm ölçeklerde pek fazla popüler olmadığını

göstermektedir. “aylık – özgeçmiş” değişkeni de 2 ’nin altında kalarak en az popüler değişkenler içerisinde dördüncü sırayı almaktadır.

Minimum ve maksimum frekans değerleri incelendiğinde, minimum frekans değeri 0 olan değişkenlerin bazı gün, hafta veya aylarda hiç popüler olmadığını gösterir. Bunlara “günlük - özgeçmiş”, “günlük - linkedin” ve “günlük - cv” örnek gösterilebilir. Ancak aynı değişkenlerin maksimum frekans değerlerinin sırasıyla 14,75, 85,55, 42,14, olduğu da görülmektedir. Bu durum, bazı değişkenlerin zaman içerisindeki çeşitli toplumsal dinamiklerden, coğrafi konumdan ve mevsimsel faktörlerden etkilenerek arama sıklığının artabileceğini ifade etmektedir. Örneğin, “iş ilanı” veya “cv” terimleri için arama frekansındaki artışlar, iş piyasasındaki mevsimsel değişikliklerle veya ekonomik olaylarla ilişkili olabilir. Ayrıca, “haftalık - işkur”, “haftalık - işsizlik maaşı” gibi maksimum frekans değeri 100 ’ün üstünde elde edilen gözlemler, veri birleştirme esnasında gerçekleştirilen oran hesaplamasından etkilenmiş gözlemlerdir. Bu gözlemler ilgili değişkenin ilgili ölçekte olağanüstü popüler dönemlerinden birinde olduğunu ifade etmektedir.

Elde edilen standart sapma değerleri, veri setlerinin içindeki değişkenlerin homojenlik veya heterojenlik seviyelerini anlamaya yardımcı olabilir. Örneğin, “haftalık - işkur” ve “aylık-secret cv” gibi 20 ’nin üstünde standart sapma değerine sahip değişkenler, diğerlerine kıyasla daha heterojen bir yapıya işaret etmektedir. Bununla birlikte, “aylık - özgeçmiş” ve “haftalık - işsizlik sigortası” gibi 1 ’in altındaki standart sapma değerleri, diğer değişkenlere göre daha homojen bir yapıya sahip olduğunu gösterebilir. Dolayısıyla “haftalık-işkur” gibi yüksek standart sapmaya sahip değişkenler için arama frekansı zaman içerisinde önemli ölçüde değişebilir.

Çarpıklık değerleri incelendiğinde simetrik dağılıma sahip olan bir değişken söz konusu değildir ancak genel olarak pozitif çarpıklık değerlerinin olması değişkenlerin sağa doğru bir eğilim gösterdiğini ifade etmektedir. Bu durum ise, veri seti içerisindeki aykırı değerlerin, dağılımın sağ tarafında yoğunlaştığını ifade etmektedir. Son olarak 3,08 basıklık değerine sahip olan “aylık - özgeçmiş” değişkeni göze çarpmaktadır. Bu değer baz alınarak veri seti içerisinde yer alan gözlemlerin ortalamaya yakın noktalarda yoğunlaştığı ve kuyruk kısımlarında daha az veri olduğu söylenebilir. Bu durum, veri seti içerisinde aykırı değerlerin olma ihtimalinin yüksek olduğunu ifade ederken, diğer değişkenlerin genel olarak 3 ’ün altında basıklık

değerlerine sahip olması, veri setlerinin genel olarak dengeli olduğu anlamını vermektedir.

3.3.3.2. Makroekonomik Dışsal Veri Setinin Tanımlayıcı İstatistikleri

2010–2023 dönemi için resmi kaynaklar tarafından yayınlanan makroekonomik değişkenlerin istatistiksel özellikleri Tablo 11 'de yer almaktadır. Bu tabloda yer alan “N” sütunu veri setindeki toplam gözlem sayısını, “ORT” sütunu değişkenler için elde edilen ortalama değerleri, “SS” sütunu değişkenlerin standart sapmasını ve “MIN”, “MAX” sütunları ise gözlemler içerisinde yer alan en küçük ve en büyük değerleri ifade etmektedir. Aynı zamanda değişkenlerin Q1, Q2 ve “MED” olarak ifade edilen medyan değerleri sunulmuş, çarpıklık değerleri “ÇARP” ve son olarak basıklık değerleri “BAS” olarak ifade edilmiştir.

Tablo 11: Makroekonomik Değişkenlerin Tanımlayıcı İstatistikleri

DEĞİŞKEN	N	ORT	SS	MİN	MAX	Q1	MED	Q3	ÇARP	BAS
EGE	161	101,2	7,6	54,8	113,4	97,6	102,9	106,1	-2,2	9,8
REEL EFEKTİF DÖVİZ KURU	161	89,5	21,2	47,6	126,4	69,7	95,78	105,6	-2,2	9,7
TÜFE	161	381,85	250,8	174,07	1300,04	219,8	282,4	444,9	-2,2	9,7
NÜFUS (bin kişi)	161	58960,4	3713,9	52376	65305	55773,5	58874,5	61991,5	-2,2	9,6
İSTİHDAM (bin kişi)	161	26556,8	2578,3	21315	31698	24447,2	26972	28356,7	-2,2	9,6
İŞSİZLİK ORANI (%)	161	10,7	1,6	8	14,2	9,7	10,4	11,6	-2,2	9,5

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo incelendiğinde ortalama değeri 101,2 olan EGE 'nin minimum değeri 54,85 iken maksimum değeri 113,4 'dür. Standart sapma değeri de 7,6 ile oldukça düşük olan bu endeks diğer değişkenlere göre daha homojen bir yapıdadır. Bu, EGE endeksinin ölçümlerinde genellikle daha az değişkenlik veya dalgalanma olduğunu ve endeks değerlerinin tutarlı olduğu bilgisini verebilir. Öte yandan, 89,5 ortalama değerine sahip döviz kuru, 21,2 standart sapma değeri ile EGE 'ye göre zaman

içerisinde daha fazla dalgalanmalar yaşamaktadır. Minimum ve maksimum değerleri arasındaki büyük fark ise araştırma dönemi içerisindeki önemli döviz kuru hareketlilikleri hakkında fikir verir. Ayrıca, 381,8 ortalama ve 250,8 standart sapma değerlerine sahip olan TÜFE 'nin minimum ve maksimum değerleri (174,07-1300,04) arasındaki büyük fark, ilgili zaman aralığında şiddetli değişikliklere uğradığına işaret etmektedir. Bu değişikliklerin nedeni yüksek enflasyon oranlarıdır.

Büyük veri setlerinin farklı dağılımlara veya aşırı değerlere sahip olabilmesi, standart sapmanın genel özellikleri tam anlamıyla yansıtamayabileceği anlamına gelmektedir. Özellikle 15 yaş ve üzeri nüfus ile istihdam sayısı gibi değişkenlerde çok yüksek standart sapma ve ortalama değerler gözlemlenmiştir. Bu durum, nüfus ve istihdam değişkenlerinde önemli değişiklikler yaşandığı fikrini verebilir. Ayrıca nüfus ve istihdam değişkenlerinin yüksek Q1, medyan ve Q3 değerleri de veri setinin çoğunlukla yüksek değerlerden oluştuğunu ifade etmektedir. İşsizlik oranının ortalama değeri 10,7 ve standart sapma değeri 1,6 olarak görünmektedir. Bu, işsizlik oranının gözlemlenen dönem boyunca diğer değişkenlere kıyasla daha dengeli kaldığını gösterir. Minimum ve maksimum işsizlik oranları (%8 ve %14,2) arasındaki anlaşılabilir fark, ekonomik koşullardaki değişikliklere bağlı olarak işsizlik oranında gerçekleşen dalgalanmalara işaret etmektedir.

Son olarak tamamen negatif olan çarpıklık değerleri, değişkenlerin sola doğru bir eğilim gösterdiğini yani veri setlerinde düşük değerlerin yüksek değerlere göre daha fazla sayıda olabileceğini gösterirken, basıklık değerlerinin tamamen 3 'den yüksek oluşu da değişkenlerin genel olarak çok sivri bir dağılıma sahip olduğunu ifade etmektedir. Bu durum, veri seti içerisinde aykırı değerlere daha az rastlandığını ifade ediyor olabilir.

3.3.4. Değişkenlerin Düzenlenmesi

Tablo 5 'de belirtilen modeller üzerinde analiz sonuçları elde edebilmek için GT değişkenleri ve makroekonomik dışsal değişkenlerin bu modellerin gerekliliklerine uygun hale getirilmesi gerekmektedir. Bu nedenle bağımsız değişkenlerin zaman serisi analizine adapte edilme detayları bu başlıkta açıklanmıştır.

Bağımsız değişkenlerin zaman serisi analizine dahil edilmesinde, temel MIDAS regresyonundan yola çıkarak belirsizlikleri dikkate alan UMIDAS tekniğinden yararlanılmıştır. Farklı zaman dilimlerindeki verileri birleştirme işlemi ağırlıkları kullanan ve belirsizliği göz önünde bulunduran UMIDAS tekniği “Karma Veri Örneklemesi” başlığında detaylandırılmıştır.

Geçmiş dönem verilerinin etkisini değerlendirmek ve gelecekteki eğilimlerini belirlemek adına her modelde t dönemi için gecikmeli değişkenler belirlenmiş ve bu gecikmeler modelleme sürecine dahil edilmiştir. Gecikme sayıları belirlenirken veri noktalarının ne kadar geriye gittiğine bağlı düzenlemeler yapılmıştır. Gecikmeli gözlemleri modellere ekleyerek, tahmin yeteneğini artırmak ve bu doğrultuda işsizlik oranı öngörüsü için geçmiş dönem verilerinin etkisini daha sağlıklı bir şekilde modellemek hedeflenmiştir.

1. Model için düzenleme: Makroekonomik değişkenlerle işsizlik oranı öngörüsü için belirlenen 5 makroekonomik dışsal değişkenin ve işsizlik oranı değişkeninin 3 dönem ($t-1$, $t-2$ ve $t-3$) gecikmesi alınmıştır ve yalnızca makroekonomik değişkenlerle kurgulanan bu model için 18 bağımsız değişken elde edilmiştir. İşsizlik öngörüsüne istinaden modeller bu bağımsız değişkenler üzerinden kurulmuştur.

2. Model için düzenleme: Günlük GT değişkenleri ve makroekonomik değişkenlerle işsizlik oranı öngörüsü için günlük değişkenlerde frekans sayısının artması nedeniyle gecikme alınırken ayın son 7 günü baz alınmıştır. Buna göre 13 değişken için t zamanda 7 dönem ($t-1$, $t-2$, $t-3$, $t-4$, $t-5$, $t-6$ ve $t-7$) gecikme alınarak 91 bağımsız değişken elde edilmiştir. Modelde kurgulanan yapı itibarıyla 1. modelde belirlenen işsizlik oranı ve makroekonomik dışsal değişkenlerinin 3 dönem gecikmesi de eklenmiş ve toplamda 109 bağımsız değişken üzerinden modeller kurulmuştur.

3. Model için düzenleme: Haftalık GT değişkenleri ve makroekonomik değişkenlerle işsizlik oranı öngörüsü için haftalık değişkenlerde gecikme alınırken, son 4 haftada yer alan gözlemler baz alınmıştır. Buna göre, her gözlemin ilgili ayın kaçınıcı haftasında olduğu tespit edilmiş ve haftalık gözlemler için gecikme alma işlemi gerçekleştirilmiştir. Dolayısıyla, 13 değişken için 4 dönem ($t-1$, $t-2$, $t-3$ ve $t-4$) gecikme alınarak 52 bağımsız değişken elde edilmiştir. Modelde kurgulanan yapı itibarıyla 1. modelde belirlenen işsizlik oranı ve makroekonomik dışsal değişkenlerinin 3 dönem

gecikmesi de eklenmiş ve toplamda 70 bağımsız değişken üzerinden modeller kurulmuştur.

4. Model için düzenleme: Aylık GT değişkenleri ve makroekonomik değişkenlerle işsizlik oranı öngörüsü için aylık GT değişkenlerinin 3 dönem (t-1, t-2 ve t-3) gecikmeleri alınarak 39 değişken elde edilmiştir. Modelde kurgulanan yapı itibarıyla 1. modelde belirlenen işsizlik oranı ve makroekonomik dışsal değişkenlerinin 3 dönem gecikmesi de eklenmiş ve toplamda 57 bağımsız değişken üzerinden modeller kurulmuştur.

5. Model için düzenleme: Günlük, haftalık, aylık GT değişkenleri ve makroekonomik değişkenlerle işsizlik oranı öngörüsü için gecikme alınarak elde edilen tüm değişkenlerin modele dahil edildiği bir yapı oluşturulmuştur. Makroekonomik dışsal bağımsız değişkenleri ve serinin kendi gecikmelerini temel olarak belirlenen 18 bağımsız değişken ile günlük GT değişkenleri üzerinden elde edilen 91, haftalık GT değişkenleri üzerinden elde edilen 52 ve aylık GT değişkenleri üzerinden elde edilen 39 değişken birleştirilmiş, toplamda 200 değişken üzerinden modeller kurulmuştur.

3.3.5. Verilerin Durağanlaştırılması

Analiz ve tahmin gerçekleştirilen serilerde trend veya yapısal değişikliklerden dolayı durağan olmama durumu söz konusuysa sonuçlarda tutarsızlık ve yanlışlık söz konusu olabilmektedir. Bu nedenle zaman serilerini modelleyebilmek için serilerin durağanlığının test edilmesi ve durağan olmayan değişkenlere durağanlaştırma işleminin uygulanması gerekmektedir. Durağanlaştırma işlemiyle, zaman serisinin ortalama ve varyans değerlerini zaman boyunca sabitleyecek bir yapı kurulmaktadır.

Çalışma kapsamında yapısal değişikliklerin önüne geçme amacıyla tüm değişkenlere Kwiatkowski-Phillips-Schmidt-Shin (KPSS) testi uygulanmış ve durağan olmayan değişkenler tespit edilmiştir. Durağan olmayan değişkenleri durağanlaştırmak için çeşitli araçlar içerisinde fark alma yöntemi tercih edilmiştir. Mevcut zaman serisinden, önceki zaman noktalarının değerlerini çıkararak birinci dereceden farklar alınmış ve birinci derecede durağanlaşmayan değişkenler tespit edilerek bu değişkenlerin ikinci dereceden farkları alınmıştır. İkinci dereceden fark

alma işlemi sonucu tüm değişkenler durağan hale gelmiştir. Böylelikle tutarsızlık ve yanlışlık problemini engellemek hedeflenmiştir.

3.3.6. Verilerin Standartlaştırılması

Bağımsız değişken veri setlerinde daha homojen bir dağılım elde etmek, aykırı ve yüksek değerlere sahip değişkenlerin tahmin üzerindeki etkisini azaltmak gibi nedenlerden ötürü ölçeklendirme işlemi gerçekleştirmek gerekmektedir. Bu işleme veri normalleştirme ya da veri standartlaştırma adı verilir. Çalışma kapsamında standartlaştırma işlemi tercih edilmiştir ve Scikit-Learn kütüphanesinin “preprocessing.StandardScaler()” fonksiyonu kullanılmıştır. Böylelikle değişken değerlerinin 0 ortalamasının etrafında dağıldığı ve varyansın 1 olduğu bir yapı elde edilmiştir.

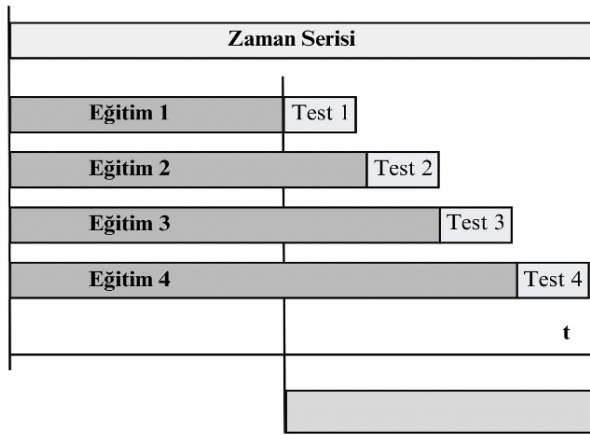
3.3.7. Eğitim ve Test Veri Setlerinin Belirlenmesi

Makine öğrenmesi modellerinin eğitilmesi ve test edilmesi amacıyla gerçek veri kümesi, eğitim ve test seti olmak üzere iki parçaya ayrılır. Veri setinin büyük bir çoğunluğu modelleri eğitmek için kullanılır ve bu verilere “eğitim seti” denir. Veri setinin kalan kısmı ise modelin tahmin performansının testi için kullanılır ve bu verilere ise “test seti” denir. Eğitim setinin, test setinden daha çok veriden oluşması, oluşturulan modelin problemi daha iyi tanımasına olanak tanır. Böylelikle model, veri setinin desenini çözümleyerek daha sağlam tahminler yapar. Çünkü modelin daha sağlam performans göstermesi için veri setindeki örüntülerin iyi anlaşılması gerekmektedir.

Değişkenlerin gecikmelerinin alınması ve durağanlaştırma süreçlerinin uygulanması, seri içindeki gözlem sayısını 161 'den 156 'ya indirmiştir. 2020 'nin 1. ayından 2023 'ün 5. ayına kadar olan dönemdeki 41 gözlem, test seti olarak ayrılmıştır. Geriye kalan 115 gözlem ise eğitim veri seti olarak kullanılmıştır. Ek olarak, zaman serisi verileri üzerinde gerçekleştirilen tahminleme için genişleyen pencere (expanding window) tekniğinden yararlanılmıştır. Bu teknik sayesinde model için oluşturulan eğitim veri seti sürekli bir şekilde artarak genişleyecek ve test veri seti zaman

içerisinde bir adım ilerleyecek şekilde oluşturulmuştur. Yani, zaman serisi verileri belirli bir zaman aralığı boyunca parçalara ayrılarak, her adımda modelin daha fazla veriyle güncellenip eğitilmesi sağlanmıştır. Şekil 26 'daki gibi, ilk adımda model yalnızca serinin başlangıç verileriyle eğitilmiş ve sonrasında pencere olarak ifade edilen küme adım adım genişletilmiştir. Böylelikle modelin her adımda daha fazla geçmiş veriyle güncellenmesi sağlanmıştır. Sonuç olarak zamanla değişen veriye uyum artırılmış ve değişen örüntülerin anlaşılması sağlanmıştır.

Şekil 26: Genişleyen Pencere Yöntemi



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Son olarak, zaman serisi çalışmalarında eğitim ve değerlendirme aşamasında model performansını tespit etmeyi sağlayan k-fold çapraz doğrulama (cross-validation) işlemi gerçekleştirirken dikkatli olmak gerekmektedir. Çünkü klasik çapraz doğrulama yöntemleri zaman serilerinde veri aralıklarının yitirilmesine neden olabilir. Bu nedenle, verinin zamansal yapısını kaybetmemek ve rastgeleliğin neden olabileceği sorunların önüne geçmek adına Python dilinde veri işleme ve model değerlendirme araçları içeren Scikit-learn kütüphanesinin “`model_selection.TimeSeriesSplit()`” aracı kullanılmıştır. `TimeSeriesSplit` fonksiyonu ile verinin sıralı zamansal parçalara bölünmesi sağlanmış ve bu parçalar k-fold çapraz doğrulama için kullanılmıştır. Her bölünmede test verileri eğitim verilerinden ileri bir tarihten seçilmiş olup, bu sayede

gelecekteki verilerin geçmiş verilere dayalı tahmin yapılması durumu gözlemlenmiştir.

3.3.8. Değişken Seçimi

Değişken seçimi, hangi değişkenin model performansına daha fazla katkısı olduğunu tespit etmek ve buna istinaden modelin ne gibi durumlarda yüksek performans sergilediğini karşılaştırmak açısından oldukça önemlidir. Model karmaşıklığını azaltarak veri setinde belli başlı özellikleri kullanmayı sağlar ve böylelikle gerekli olmayan değişkenlerin kapsama dahil edilmemesine yardımcı olur. Bu nedenle değişken seçimi olmadan oluşturulan tüm tahmin modelleri değişken seçim işlemi gerçekleştirilerek tekrarlanmıştır. Bu işlem gerçekleştirilirken hem tahmin üreten hem de değişken seçim tekniği olarak kullanılan, değişken sayısının çok ancak veri sayısının az olduğu durumlarda tercih edilen LASSO modelinden yararlanılmıştır. LASSO, ceza terimi kullanarak seyrek katsayılar sağlar ve değişken seçimine yardımcı olur (Aras ve Lisboa, 2022: 3). Test setindeki 41 gözlemin her birini tahmin ederken, değişken seçimi iteratif bir biçimde tekrarlanmış ve tahmin modelleri kurulmuştur. Dolayısıyla 1-adım sonrası (one-step ahead) tahmin edecek şekilde öngörüler oluşturulmuştur. Tablo 12 'de 41 adımda en çok seçilen değişkenler listelenmiştir. Söz konusu tablo oluşturulurken 10 adım ve üstünde seçilen değişkenler eklenmiştir. Buna göre, en çok seçilen değişkenlerin ait olduğu veri grubu ve toplam seçilme sayısı tabloda gösterilmiştir.

Tablo 12: LASSO Değişken Seçimi ile En Çok Seçilen Değişkenler

<i>En Çok Seçilen Değişkenler</i>	<i>Seçilme Sayısı</i>
<i>makro - issizlik oranı (t-1)</i>	12
<i>makro - istihdam (t-1)</i>	10
<i>makro - Tufe (t-2)</i>	14
<i>makro - 15 yas nufus (t-2)</i>	28
<i>makro - issizlik oranı (t-3)</i>	36
<i>makro - EGE (t-3)</i>	34
<i>makro - 15 yas nufus (t-3)</i>	31
<i>makro - istihdam (t-3)</i>	36
<i>günlük - issizlik maası (t-2)</i>	32
<i>günlük - issizlik sigortası (t-3)</i>	22
<i>günlük - is ilanı (t-4)</i>	20
<i>günlük - linkedin (t-4)</i>	10
<i>günlük - linkedin (t-5)</i>	17
<i>günlük - is (t-6)</i>	15
<i>günlük - part time (t-6)</i>	30
<i>günlük - linkedin (t-6)</i>	20
<i>günlük - ozgecmis (t-6)</i>	10
<i>günlük - is basvurusu (t-7)</i>	27
<i>günlük - is ilanları (t-7)</i>	12
<i>haftalık - is ilanları (t-1)</i>	14
<i>haftalık - issizlik sigortası (t-1)</i>	14
<i>haftalık - ozgecmis (t-1)</i>	27
<i>haftalık - cv (t-2)</i>	11
<i>haftalık - is basvurusu (t-2)</i>	10
<i>haftalık - is ilanları (t-2)</i>	20
<i>haftalık - cv (t-3)</i>	14
<i>haftalık - linkedin (t-3)</i>	15
<i>haftalık - cv (t-4)</i>	22
<i>haftalık - is ilanları (t-4)</i>	20
<i>aylık - is basvurusu (t-1)</i>	17
<i>aylık - issizlik maası (t-1)</i>	24
<i>aylık - ozgecmis (t-1)</i>	18
<i>aylık - issizlik maası (t-2)</i>	23
<i>aylık - is basvurusu (t-3)</i>	12
<i>aylık - issizlik maası (t-3)</i>	31

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 12 'ye göre, “makro - issizlik oranı (t-3)”, “makro - EGE (t-3)”, “makro - 15 yas nufus (t-3)”, “makro - istihdam (t-3)”, “günlük - issizlik maası (t-2)”, “günlük - part time (t-6)” ve “aylık - issizlik maası (t-3)” değişkenleri 41 adım içerisinde 30 kez ve üstü seçilen değişkenlerdir. Bu durumda bu değişkenlerin kurulan modellere en çok katkıda bulunan ve tahmin gerçekleştirme kabiliyeti en yüksek olan değişkenler

olduğu söylenebilir. Özellikle makroekonomik değişkenlerin işsizlik oranı üzerinde büyük etkiye sahip olması ve bu değişkenlerin tahmin performansına yüksek katkısının oluşu oldukça anlaşılır bir durumdur.

3.3.9. Model Kurulumu

Probleme istinaden çözüm gerçekleştirilirken doğru modeli oluşturmak en önemli adımlardan birisidir. Model kurulumunun sağlıklı ilerlemesi için ise doğru teknik ve doğru parametreler tespit edilmelidir. Model parametrelerine ek olarak, model davranışını kontrol etmeyi sağlayan hiperparametre (hyperparameter) vardır. Tahmin sonuçlarını elde ederken, modellerin aşırı öğrenme (overfitting) ya da eksik öğrenme (underfitting) gibi sorunlarının önüne geçmek için parametreler doğru belirlenmelidir. Hiperparametre, model eğitim aşamasından önce belirlenir ve modelin karmaşıklığını düzenlemeye çalışır. Deneme-yanılma ya da optimizasyon yöntemleriyle belirlenen hiperparametre, verilere bağlı değişkenlik gösterir ve modelin eğitim süreci boyunca sabit kalır.

Model optimizasyonu adımımda çeşitli hiperparametre seçme teknikleri vardır. Ancak bu çalışmada ızgara arama (grid-search) hiperparametre optimizasyon tekniği tercih edilmiştir. Bu yöntemle belirlenen çeşitli parametre kombinasyonları denenmiş ve sonuçlar çeşitli performans metrikleriyle ölçümlenmiştir. Sonuç olarak en iyi performansı sergileyen hiperparametre kombinasyonu seçilmiştir. Ardından bu çeşitli kombinasyonlarla model eğitilerek model performansı ölçümlenmiştir. Çalışma kapsamında kullanılan; GBDT, RF, SVR, KNN, XGboost, LASSO, ElasticNet ve Ridge regresyon modellerinin optimizasyonuna istinaden belirlenen hiperparametre ve değer aralıkları Tablo 13 'de yer almaktadır.

Tablo 13: Modellerin Hiperparametre Dizini

Model	Hiper Parametre	Belirlenen Aralıklar
GBDT	min_samples_leaf	[1,3,5,7,10,15,30,50]
	learning_rate	[0.001,0.005,0.01,0.03,0.05,0.07,0.1,0.2]
	min_samples_split	[2,3,5,7,10,15,30,50]
	n_estimators	[30,50,100,200,300]
	max_depth	[1,3,5,7,10]
	loss	[squared_error,absolute_error, huber, quantile]
RF	criterion	[squared_error, absolute_error, friedman_mse]
	min_samples_leaf	[1,3,5,7,10,15,30,50]
	max_depth	[1,3,5,7,10]
	min_samples_split	[2,3,5,7,10,15,30,50]
	n_estimators	[50,100,200,300]
SVR	C	[10000,5000,1000,500,100,10,1]
	ϵ	[0,0.1,0.01,0.001,0.0001]
	γ	[0.1,0.01,0.001,0.0001,0.00001]
KNN	n_neighbours	[1,50]
	metric	[manhattan, minkowski]
	weights	[uniform, distance]
XGboost	max_depth	[1,3,5,7,10,15]
	learning_rate	[0.005,0.01,0.03,0.05,0.07,0.1,0.2,0.3]
	γ	[0,0.1,0.3,0.7,1,3,5,10,15,30]
	n_estimators	[30,50,100,300,500,1000]
	min_child_weight	[0,0.1,0.3,0.5,1,3,5,10,30,50,100]
	colsample_bytree	[0.6,0.7,0.8,0.9,1]
	reg_lambda	[0,0.1,0.3,0.5,1,3,5,10,30,50,100]
	reg_alpha	[0,0.1,0.3,0.5,1,3,5,10,30,50,100]
	objective	[reg:squarederror,reg:pseudohubererror,reg:linear]

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

3.3.10. Model Performanslarının Değerlendirilmesi

Bu çalışma kapsamında, model öngörü performanslarının değerlendirilmesinde kullanılan çeşitli metrikler içerisinde; MAE, MSE, R^2 ve MAPE performans istatistikleri tercih edilmiştir. Kullanılan ölçütlere ait detaylı bilgi

ikinci bölümde bulunan “Regresyon Modellerinde Performans İstatistikleri” başlığında detaylandırılmıştır.

Ayrıca, yalnızca işsizlik oranı serisinin 3 dönem gecikmesine (t-1, t-2, t-3) kadar olan otoregresif modeller oluşturulmuştur. Bu modellerin performans ölçüleri Tablo 14 'de verilmiştir.

Tablo 14: Otoregresif Modellerin Performans İstatistikleri

Otoregresif Model	MAE	MSE	R ²	MAPE
AR(1)	0,3962	0,3314	0,8224	0,0341
AR(2)	0,4082	0,3458	0,8147	0,0351
AR(3)	0,4083	0,3199	0,8285	0,0353

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Elde edilen kıyaslama modelleri arasından AR(3) modelinin performans ölçüleri, tahmin modellerinin değerlendirilmesinde tercih edilmiştir. AR(3) modelinin performans ölçüleri, tahmin modellerinin performans ölçüleriyle oranlanarak karşılaştırılmıştır.

3.4. BULGULAR VE BULGULARIN YORUMLANMASI

Bu bölümde, makroekonomik değişkenler ve serinin kendi gecikmelerini içeren veri setleriyle farklı kombinasyonlarda günlük, haftalık ve aylık bazda elde edilen tahmin modellerinin performans istatistikleri incelenmiştir. LASSO değişken seçimi gerçekleştirilmiş ve değişken seçimi gerçekleştirilmemiş durumlar için elde edilen performans ölçüleri, AR(3) kıyaslama modelinin performans ölçüleriyle oranlanarak Tablo 16 ve Tablo 25 arasında sunulmuştur. Karşılaştırma işlemi gerçekleştirilirken gerçekleştirilen hesaplamalar Tablo 15 'de ifade edilmiştir. Ardından oluşturulan dokuz modelin en iyi performansı sergilediği senaryolar Tablo 26 'da açıklanmıştır.

Tablo 15: Performans Hesaplamaları

Ölçü	Hesaplama Yöntemi
MAE*	Tahmin Modeli MAE Değeri/AR(3) Modeli MAE Değeri
MSE*	Tahmin Modeli MSE Değeri/AR(3) Modeli MSE Değeri
R ² *	AR(3) Modeli R ² * Değeri/ Tahmin Modeli R ² * Değeri
MAPE*	Tahmin Modeli MAPE Değeri/AR(3) Modeli MAPE Değeri

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 15 'de, R² ölçüsü için gerçekleştirilen hesaplamaların diğer ölçülere göre farklılık gösterdiği görülmektedir. Normal şartlarda, diğer ölçülere istinaden gerçekleştirilen oranlama işlemi sonrası elde edilen değer 1 'den küçük olması gözetilirken R² 'de bu durum tam tersidir. Bu nedenle bu ölçü için tam tersi işlem uygulanarak yorumlama kolaylığı amaçlanmıştır.

Tablo 16 ve Tablo 25 arasında sunulan sonuçlar içerisinde işsizlik oranını tahmin etmek için oluşturulan modellerin performansını değerlendirmek için koyulaştırılmış değerler dikkate alınmıştır. Bu değerler, her değişken kombinasyonu için en iyi modelin performans sonuçlarını ifade etmektedir. Oluşturulan bu tablolar sonucunda tüm tahmin modellerinin çalışma kapsamında en iyi sonuç verdiği grup Tablo 26 'da verilmiştir. Sunulan bu tabloda her bir modelin en iyi performansı sergilediği durum raporlanmış ve son olarak değişken seçiminin ne derece etkili olduğu gözlemlenmiştir.

Tablo 16: Değişken Seçimi Olmadan: Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	1,0393	1,0615	1,0129	1,0186
LASSO	1,0246	1,0213	1,0044	1,0070
KNN	0,9199	0,9039	0,9805	0,9178
SVR	1,0354	0,9933	0,9986	1,0167
DT	0,9707	0,9208	0,9839	0,9709
RF	0,9190	0,8652	0,9729	0,9129
GBDT	0,8861	0,8671	0,9732	0,8832
XGboost	0,8874	0,8904	0,9778	0,8909
Ridge	1,0615	1,1001	1,0212	1,0408

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 16 'da değişken seçimi olmadan oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken makroekonomik dışsal değişkenlerle işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlara göre MSE* ve R²* değerleri sırasıyla 0,8652 ve 0,9729 elde edilen RF modeli ile MAE* ve MAPE* değerleri sırasıyla 0,8861 ve 0,8832 olarak elde edilen GBDT modellerinin performans sonuçları 1 'in altında kalan diğer modellere göre en iyi performansı sergileyen modeller oldukları gözlemlenmiştir. Bu durumda her iki modelin de diğer modellere göre işsizlik oranı öngörüsünde AR(3) modelinden daha iyi performans sergilediği söylenebilir. Ayrıca XGBoost modelinin de performans sonuçları GBDT ve RF modellerine oldukça yakındır. Bu durumda XGBoost'un da AR(3) modeline göre oldukça iyi sonuçlar elde ettiği söylenebilir. Öte yandan performans sonuçları 1 'in üstünde elde edilen ElasticNet, LASSO, SVR ve Ridge modellerinin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadıkları görülmektedir.

Tablo 17: Değişken Seçimi Olmadan: Aylık GT + Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	1,0021	0,9803	0,9959	0,9922
LASSO	1,0213	0,9964	0,9992	1,0107
KNN	0,9121	0,8838	0,9765	0,9051
SVR	0,9823	0,9491	0,9896	0,9831
DT	1,0514	1,0293	1,0061	1,0544
RF	0,9048	0,8593	0,9717	0,9018
GBDT	0,9074	0,8772	0,9752	0,9005
XGboost	0,8764	0,8751	0,9748	0,8740
Ridge	1,0909	1,1512	1,0323	1,0742

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 17 'de değişken seçimi olmadan oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken aylık GT değişkenleri, makroekonomik dışsal değişkenler ve işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre MSE* ve R²* değerleri sırasıyla 0,8593 ve 0,9717 olarak elde edilen RF modeli ve MAE* ve MAPE* değerleri sırasıyla 0,8764 ve 0,8740 olarak elde edilen XGBoost modelinin, performans sonucu 1 'in altında kalan diğer modellere göre en iyi performansı sergiledikleri gözlemlenmiştir. Bu durumda her iki modelin diğer modellere göre işsizlik oranı öngörüsünde AR(3) modelinden daha iyi performans sergilediği söylenebilir. Öte yandan performans sonuçları çoğunlukla 1 'in üstünde elde edilen ElasticNet, LASSO, DT ve Ridge modellerinin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadıkları görülmektedir.

Tablo 18: Değişken Seçimi Olmadan: Haftalık GT + Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	0,9931	0,9581	0,9914	0,9840
LASSO	1,0104	0,9760	0,9951	1,0005
KNN	0,9280	0,9399	0,9877	0,9246
SVR	0,9611	0,9279	0,9853	0,9596
DT	0,8990	0,8957	0,9789	0,8998
RF	0,9204	0,9010	0,9799	0,9164
GBDT	0,9326	0,9119	0,9821	0,9315
XGboost	0,9254	0,9550	0,9908	0,9220
Ridge	1,1045	1,1269	1,0270	1,0970

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 18 'de değişken seçimi olmadan oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken haftalık GT değişkenleri, makroekonomik dışsal değişkenler ve işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre MAE*, MSE*, R²* ve MAPE* değerleri sırasıyla 0,899, 0,8957, 0,9789 ve 0,8998 olarak elde edilen DT modelinin performans sonucu 1'in altında kalan diğer modellere göre en iyi performansı sergilediği gözlemlenmiştir. Bu durumda DT modelinin diğer modellere göre işsizlik oranı öngörüsünde AR(3) modelinden daha iyi performans sergilediği söylenebilir. Öte yandan performans sonuçları çoğunlukla 1 'in üstünde elde edilen LASSO ve Ridge modellerinin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadıkları görülmektedir. Ek olarak, bu performans sonuçları Tablo 16 ve Tablo 17 'de elde edilen sonuçlarla kıyaslandığında, çoğu model için elde edilen değer 1 'in altında kaldığı dolayısıyla AR(3) modelinden daha iyi performans sergilediği görülmektedir.

Tablo 19: Değişken Seçimi Olmadan: Günlük GT + Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	1,0235	1,0577	1,0121	1,0168
LASSO	1,0293	1,0759	1,0160	1,0197
KNN	0,9263	0,9458	0,9889	0,9215
SVR	0,9804	0,9715	0,9941	0,9651
DT	0,9617	0,9294	0,9856	0,9592
RF	0,9325	0,9133	0,9824	0,9276
GBDT	0,9441	0,9421	0,9882	0,9404
XGboost	0,8995	0,9058	0,9809	0,8948
Ridge	1,0943	1,2016	1,0435	1,0961

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 19 'da değişken seçimi olmadan oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken günlük GT değişkenleri, makroekonomik dışsal değişkenler ve işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre MAE*, MSE*, R²* ve MAPE* değerleri sırasıyla 0,8995, 0,9058, 0,9809 ve 0,8948 olarak elde edilen XGBoost modelinin performans sonucu 1 'in altında kalan diğer modellere göre en iyi performansı sergilediği gözlemlenmiştir. Bu durumda XGBoost modelinin diğer modellere göre işsizlik oranı öngörüsünde AR(3) modelinden daha iyi performans sergilediği söylenebilir. Öte yandan performans sonuçları 1 'in üstünde elde edilen ElasticNet, LASSO ve Ridge modellerinin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadıkları görülmektedir.

Tablo 20: Değişken Seçimi Olmadan: Tüm Veriler ile Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	0,9711	0,9574	0,9913	0,9721
LASSO	0,9811	0,9717	0,9942	0,9817
KNN	0,8999	0,9015	0,9800	0,8922
SVR	0,9515	0,9268	0,9851	0,9442
DT	0,9707	0,9662	0,9931	0,9822
RF	0,9339	0,9182	0,9834	0,9278
GBDT	0,9135	0,9030	0,9803	0,9038
XGboost	0,8697	0,8711	0,9740	0,8599
Ridge	1,1359	1,2436	1,0531	1,1500

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 20 'de değişken seçimi olmadan oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken günlük, haftalık ve aylık GT değişkenleri, makroekonomik dışsal değişkenler ile işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre MAE*, MSE*, R²* ve MAPE* değerleri sırasıyla 0,8697, 0,8711, 0,9740 ve 0,8599 olarak elde edilen XGBoost modelinin diğer modellere göre en iyi performansı sergilediği gözlemlenmiştir. Bu durumda XGBoost modelinin performans sonucu 1 'in altında kalan diğer modellere göre işsizlik oranı öngörüsünde AR(3) modelinden daha iyi performans sergilediği söylenebilir. Öte yandan performans sonuçları 1 'in üstünde elde edilen model yalnızca Ridge modeli olmuştur. Bu durumda tüm değişkenlerin birleştirildiği bu senaryoda, yalnızca Ridge modelinin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadığı görülmektedir.

Son olarak değişken seçimi olmadan gerçekleştirilen tüm modeller içerisinde en iyi performansı sergileyen çalışma, tüm değişkenlerin birleştirildiği durumda ortaya çıkmıştır ve ilgili değişken kombinasyonları içerisinde genellikle en iyi performansları GBDT, XGBoost ve DT ağaç tabanlı modellerin sergilediği görülmüştür. Ayrıca

Ridge modelinin tüm deęişken kombinasyonlarında AR(3) modeline kıyasla işsizlik oranını daha kötü açıkladığı gözlemlenmiştir.

Tablo 21: LASSO Deęişken Seçimi ile: Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	1,0393	1,0615	1,0129	1,0186
LASSO	1,0246	1,0213	1,0044	1,0070
KNN	0,9251	0,8898	0,9777	0,9237
SVR	1,0798	1,0784	1,0165	1,0601
DT	0,9886	0,9395	0,9876	0,9924
RF	0,9294	0,8889	0,9775	0,9275
GBDT	0,9191	0,8905	0,9778	0,9159
XGboost	0,9159	0,8964	0,9790	0,9174
Ridge	1,0615	1,1001	1,0212	1,0408

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 21 'de LASSO deęişken seçimi ile oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken makroekonomik dışsal deęişkenler ile işsizlik oranı serisinin gecikmeli deęişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre en iyi MAE* deęeri 0,9159 ile XGBoost modelinde elde edilirken, en iyi MSE* deęeri 0,8889 ve R²* deęeri 0,9775 ile RF modelinde elde edilmiştir. Öte yandan en düşük MAPE* deęeri ise GBDT modelinde gözlemlenmiştir. Bu durumda bu üç modelin, performans sonucu 1 'in altında kalan diğer modellere kıyasla AR(3) modelinden daha iyi performans sergilediği söylenebilir. Öte yandan Tablo 16 'da aynı verilerle deęişken seçimi olmadan elde edilen sonuçlar gibi deęişken seçimi ile elde edilen bu sonuçlarda da performans sonuçları 1 'in üstünde elde edilen modeller; ElasticNet, LASSO, SVR ve Ridge modelleri olmuştur. Bu modellerin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadığı görülmektedir.

Tablo 22: LASSO Değişken Seçimi ile: Aylık GT + Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	1,0021	0,9803	0,9959	0,9922
LASSO	1,0213	0,9964	0,9992	1,0107
KNN	0,9234	0,8990	0,9795	0,9259
SVR	0,9906	0,9806	0,9960	0,9775
DT	1,0066	0,9250	0,9847	1,0105
RF	0,9210	0,8954	0,9788	0,9169
GBDT	0,8838	0,8325	0,9665	0,8819
XGboost	0,9158	0,9121	0,9821	0,9186
Ridge	1,0909	1,1512	1,0323	1,0742

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 22 'de LASSO değişken seçimi ile oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken aylık GT değişkenleri, makroekonomik dışsal değişkenler ve işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre en iyi MAE*, MSE*, R²* ve MAPE* değerleri sırasıyla 0,8838, 0,8325, 0,9665, 0,8819 olarak GBDT modelinde gözlemlenmiştir. Bu durumda GBDT modelinin, performans sonucu 1 'in altında kalarak AR(3) modelinden daha iyi performans gösteren diğer modeller içerisinde en iyi sonucu veren model olduğu söylenebilir. Öte yandan Tablo 17 'de aynı verilerle değişken seçimi olmadan elde edilen sonuçlar gibi değişken seçimi ile elde edilen bu sonuçlarda da performans sonuçları çoğunlukla 1 'in üstünde elde edilen modeller; ElasticNet, LASSO, DT ve Ridge modelleri olmuştur. Dolayısıyla bu modellerin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadığı görülmektedir.

Tablo 23: LASSO Değişken Seçimi ile: Haftalık GT + Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	0,9931	0,9581	0,9914	0,9840
LASSO	1,0104	0,9760	0,9951	1,0005
KNN	0,9184	0,9487	0,9895	0,9169
SVR	1,0196	0,9813	0,9961	1,0159
DT	0,9886	0,9868	0,9973	0,9850
RF	0,9146	0,9240	0,9845	0,9108
GBDT	0,9197	0,8745	0,9747	0,9206
XGboost	0,9326	0,9937	0,9987	0,9328
Ridge	1,1045	1,1269	1,0270	1,0970

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 23 'de LASSO değişken seçimi ile oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken haftalık GT değişkenleri, makroekonomik dışsal değişkenler ve işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre en iyi MAE* ve MAPE* değerleri sırasıyla 0,9146 ve 0,9108 ile RF modeli için elde edilirken, en iyi MSE* ve R²* değerleri sırasıyla 0,8745 ve 0,9747 ile GBDT modeli için elde edilmiştir. Bu durumda bu senaryo için performans sonucu 1 'in altında kalarak AR(3) modeline göre daha iyi performans sergileyen diğer modeller içerisinde RF ve GBDT modellerinin en iyi modeller oldukları söylenebilir. Öte yandan performans sonuçları çoğunlukla 1 'in üstünde elde edilen modeller; LASSO, SVR ve Ridge modelleri olmuştur. Dolayısıyla bu modellerin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadığı görülmektedir.

Tablo 24: LASSO Değişken Seçimi ile: Günlük GT + Makro + AR(3) Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	1,0235	1,0577	1,0121	1,0168
LASSO	1,0293	1,0759	1,0160	1,0197
KNN	0,8736	0,9488	0,9895	0,8736
SVR	1,1354	1,2249	1,0488	1,1208
DT	1,0334	1,1002	1,0212	1,0491
RF	0,9348	0,9433	0,9884	0,9349
GBDT	0,9061	0,9368	0,9871	0,9035
XGboost	0,8799	0,8752	0,9748	0,8765
Ridge	1,0943	1,2016	1,0435	1,0961

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 24 'de LASSO değişken seçimi ile oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken günlük GT değişkenleri, makroekonomik dışsal değişkenler ve işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre en iyi MAE* ve MAPE* değerleri 0,8736 olarak elde edilen KNN ile en iyi MSE* ve R²* değerleri sırasıyla 0,8752 ve 0,9748 olarak elde edilen XGBoost modelleri ön plana çıkmaktadır. Bu durumda bu senaryo için 1 'in altında kalan modeller içerisinde AR(3) modeline göre en iyi performansı KNN ve XGBoost modellerinin sergilediği söylenebilir. Öte yandan performans sonuçları çoğunlukla 1 'in üstünde elde edilen modeller; ElasticNet, LASSO, SVR, DT ve Ridge modelleri olmuştur. Dolayısıyla bu modellerin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadığı görülmektedir.

Tablo 25: LASSO Değişken Seçimi ile: Tüm Veriler ile Model Performansları

Model	MAE*	MSE*	R ² *	MAPE*
ElasticNet	0,9711	0,9574	0,9913	0,9721
LASSO	0,9811	0,9717	0,9942	0,9817
KNN	0,9650	0,9441	0,9886	0,9612
SVR	1,1092	1,1198	1,0254	1,0950
DT	0,9379	0,9443	0,9886	0,9434
RF	0,8654	0,8382	0,9676	0,8589
GBDT	0,8596	0,8171	0,9635	0,8480
XGboost	0,9239	0,9789	0,9957	0,9212
Ridge	1,1359	1,2436	1,0531	1,1500

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 25 'de LASSO değişken seçimi ile oluşturulan modeller için elde edilen performans sonuçları yer almaktadır. Bu modeller oluşturulurken günlük, haftalık ve aylık GT değişkenleri, makroekonomik dışsal değişkenler ile işsizlik oranı serisinin gecikmeli değişkenleri kullanılmıştır. Modellerin performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak elde edilen sonuçlarına göre 1 'in altında kalarak AR(3) modeline göre daha iyi performans sergileyen modeller içerisinde, MAE*, MAPE*, MSE* ve R²* değerleri sırasıyla 0,8596, 0,8171, 0,9635 ve 0,8480 olarak elde edilen GBDT, en iyi performansı sergilemiştir. Öte yandan performans sonuçları çoğunlukla 1 'in üstünde elde edilen modeller SVR ve Ridge modelleri olmuştur. Dolayısıyla bu modellerin AR(3) modeline kıyasla daha kötü performans sergilediği yani işsizlik oranını açıklamakta AR(3) modelinden daha iyi olmadığı görülmektedir.

Son olarak LASSO değişken seçimi ile gerçekleştirilen tüm modeller içerisinde en iyi performansı sergileyen çalışma, tüm değişkenlerin birleştirildiği durumda ortaya çıkmıştır ve ilgili değişken kombinasyonları içerisinde genellikle en iyi performansları GBDT ve RF ağaç tabanlı modellerin sergilediği görülmüştür. Ayrıca Ridge modelinin tüm değişken kombinasyonlarında AR(3) modeline kıyasla işsizlik oranını daha kötü açıkladığı gözlemlenmiştir.

Tablo 26: Modellerin En İyi Değişken Kombinasyonları

	Değişken Kombinasyonları	Model	MAE*	MSE*	R ² *	MAPE*
Değişken Seçimsiz	Makro+ AR(3)	Ridge	1,0615	1,1001	1,0212	1,0408
	Haftalık GT + Makro + AR(3)	DT	0,8990	0,8957	0,9789	0,8998
	Tümü	SVR	0,9515	0,9268	0,9851	0,9442
	Tümü	XGboost	0,8697	0,8711	0,9740	0,8599
LASSO Değişken Seçimli	Günlük GT + Makro + AR(3)	KNN	0,8736	0,9488	0,9895	0,8736
	Tümü	ElasticNet	0,9711	0,9574	0,9913	0,9721
	Tümü	LASSO	0,9811	0,9717	0,9942	0,9817
	Tümü	RF	0,8654	0,8382	0,9676	0,8589
	Tümü	GBDT	0,8596	0,8171	0,9635	0,8480

Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Tablo 26 'da çalışma kapsamında tercih edilen dokuz modelin en iyi performans sergilediği değişken kombinasyonu verilmiştir. Buna göre ilgili modelin en iyi performans sergilediği kombinasyonun değişken seçimli olup olmadığı da belirtilmiştir. Ek olarak modellere istinaden AR(3) modeli ile oranlanarak elde edilen performans istatistikleri de tabloda sunulmuştur. İlgili değişken kombinasyonları 3. bölüm “Değişkenlerin Düzenlenmesi” başlığında detaylandırılmıştır.

Değişken seçimi olmadan oluşturulan modeller için elde edilen sonuçlara göre, çalışma kapsamında Ridge modelinin en iyi performans gösterdiği model “Makro + AR(3)”, DT modelinin “Haftalık + Makro”, SVR ve XGBoost modellerinin ise “Tümü” yani tüm değişkenlerin modele eklendiği modellerdir. Buna göre, bu tahmin modellerinin, değişken seçimi gerçekleştirilmeden daha iyi sonuçlar verdiği gözlemlenmiştir.

Ridge tahmin modeli için en iyi performans değişken seçimi olmadan gerçekleştirilen ve makroekonomik dışsal değişkenlerle işsizlik oranı serisinin gecikmeli değişkenleri ile kurulan modelde gerçekleşmiştir. Ancak modelin kendi performans istatistiklerinin, AR(3) performans metrikleri ile oranlanarak elde edilen

sonuçlarına göre MAE*, MSE*, R²* ve MAPE* değerleri sırasıyla 1,0615, 1,1001, 1,0212 ve 1,0408 olarak elde edilmiştir. Buna göre 1'in üstünde olan bu sonuçlar, Ridge modelinin optimal performansının, AR(3) modelinin gerisinde kaldığını yani AR(3) modelinden daha kötü performans sergilediğini göstermektedir. Bu nedenle Ridge tahmin modelinin kıyaslama modeline göre daha iyi performans sergilediği söylenememektedir.

DT tahmin modelinin çalışma kapsamında en iyi performans gösterdiği model, değişken seçimi olmadan gerçekleştirilen ve haftalık GT değişkenlerinin “Makro + AR(3)” değişkenleri ile birleştirilerek oluşturulduğu modeldir. DT performans istatistiklerinin, AR(3) performans istatistikleri ile oranlanarak ulaşılan sonuçlara göre tüm değerler 1 'in altındadır. Performans ölçülerinin 1 'in altında olması, DT tahmin modelinin kıyaslama modeline göre daha iyi performans sergilediğini ifade etmektedir.

SVR ve XGBoost tahmin modelleri için en iyi performans, değişken seçimi olmadan gerçekleştirilen ve tüm değişkenlerin birleştirilerek modellere eklendiği modelde elde edilmiştir. İki model için de performans ölçüleri 1 'in altında kalarak AR(3) kıyaslama modeline göre daha iyi performans göstermiştir. Bu durum, iki modelin de iyi ve anlamlı sonuçlar ürettiğini ifade etmektedir. Ancak bu iki modeli yer aldıkları “Tümü” grubu açısından değerlendirdiğimizde, XGBoost modelinin örneğin MAE* değerinde elde ettiği 0,8697 oranı ile, SVR modeline göre nispeten daha iyi performans sergilediği görülmektedir. Öte yandan XGBoost değişken seçimi olmadan elde edilen ve değişken seçimli varyasyonuna göre daha iyi performans sergileyen tüm modeller içerisinde, AR(3) modeline göre en iyi performansı sergileyen model olmuştur.

LASSO değişken seçimi ile oluşturulan modeller için elde edilen sonuçlara göre, çalışma kapsamında KNN 'in değişken seçimi ile “Günlük GT + Makro” grubunda en iyi performans sergilediği gözlemlenirken ElasticNet, LASSO, RF ve GBDT modellerinin en iyi sonuçları “Tümü” grubunda vermiştir. Buna göre bu modeller, tüm değişkenlerin dahil edildiği modelde LASSO değişken seçimi ile daha iyi performans sergilemektedir. Bu modellerin, örneğin MSE* değerleri sırasıyla 0,9488, 0,9574, 0,9717, 0,8382 ve 0,8171 'dir. Bu durumda, modellerin hepsi AR(3) modeline göre daha iyi performans göstermektedir. Öte yandan LASSO değişken

seçimi ile elde edilen ve değişken seçimi gerçekleştirilmeyen varyasyonuna göre daha iyi performans sergileyen tüm modeller içerisinde, AR(3) modeline göre en iyi performansı sergileyen model, GBDT olmuştur.

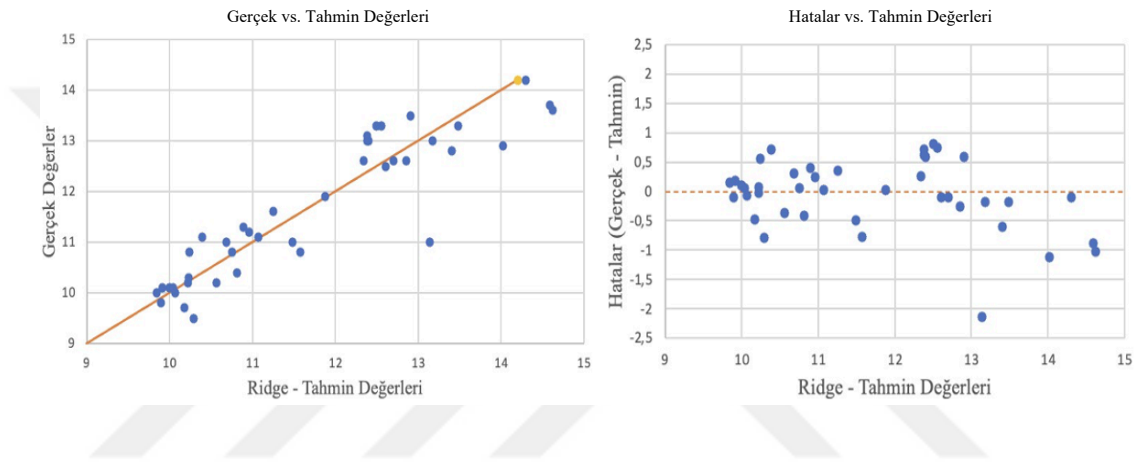
Özetle, tüm modeller ve farklı değişken kombinasyonları değerlendirildiğinde, oluşturulan 9 tahmin modeli içerisinde SVR, XGBoost, ElasticNet, LASSO, RF ve GBDT modellerinin, tüm değişkenlerin birleştirildiği “Tümü” grubunda daha iyi performans sergilediği görülmüştür. Bu modeller içerisinde SVR ve XGBoost değişken seçimi olmadan daha iyi sonuçlar üretirken, ElasticNet, LASSO, RF ve GBDT modelleri ise değişken seçimi ile daha iyi sonuçlar üretmiştir. Ek olarak tüm modeller içerisinde en iyi performansı değişken seçiminin tüm değişkenlerde kullanıldığı GBDT modeli gösterirken; RF, XGBoost ve DT modellerinin de iyi sonuçlar elde etmesi ağaç tabanlı modellerin çalışma kapsamında iyi sonuçlar ürettiğini göstermektedir. Ayrıca analiz sonuçları Ridge regresyon modelinin GT verilerine ihtiyaç duymadığını ve tüm modeller içerisinde en kötü performansa sahip olduğunu göstermektedir. Bu durumda Ridge tahmin modelinin AR(3) modelinden bile kötü olduğu anlaşılmıştır. Ek olarak, oluşturulan modellerin çoğunlukla tüm değişkenlerin birleştirildiği modelde daha iyi performans sergilemesi beklenen bir şeydir. Çünkü değişken gruplarının birleştirilmesi sayesinde daha geniş ve kapsamlı veri seti elde edilmiştir. Böylelikle model daha geniş bir özellik yelpazesinde çalışmıştır. Bu durum, daha kompleks ilişkileri açıklama yeteneğini artırabilir ve farklı değişken gruplarından gelen bilgiler sayesinde, bir değişken grubunun eksik kaldığı noktaları güçlü olan değişken grubunun tamamlayabilme ihtimalini ortaya çıkarır. Bu sayede daha güçlü tahminler yapılabilir.

Son olarak, “Tümü” grubunda yer alan ve iyi performans sergileyen bu modellerin çoğunlukla değişken seçimi ile elde edilen sonuçlarda daha anlamlı sonuçlar üretmesi, LASSO değişken seçiminin model içerisinde yer alan gereksiz ya da etkisiz değişkenleri elemesinden kaynaklanmaktadır. Gereksiz değişkenlerin elenmesiyle birlikte modelin daha basit ve daha genelleştirilebilir olmasını sağlayabilir. Ek olarak modelin aşırı uyum sorununun önüne geçerek model performansını iyileştirebilir.

Aşağıda her modelin en iyi performans sergilediği modelde elde edilen tahmin değerleri ve hata (artık) değerleri ile oluşturulan dağılım grafikleri yer almaktadır.

Buna göre 9 en iyi model için oluşturulan şekiller içerisinde, solda yer alan “Gerçek vs. Tahmin Değerleri” şekillerinde $x = y$ referans doğrusu baz alınarak yorumlamalar gerçekleştirilmiştir. Bu sayede işsizlik oranının gerçek değerleri ile tahmin modeli ile elde edilmiş tahmin değerlerinin ilişkileri elde edilmiştir. Sağda ise hata terimleri yatay ekseninde gösterilen model tahmin değerlerine göre konumlandığı grafikler yer almaktadır.

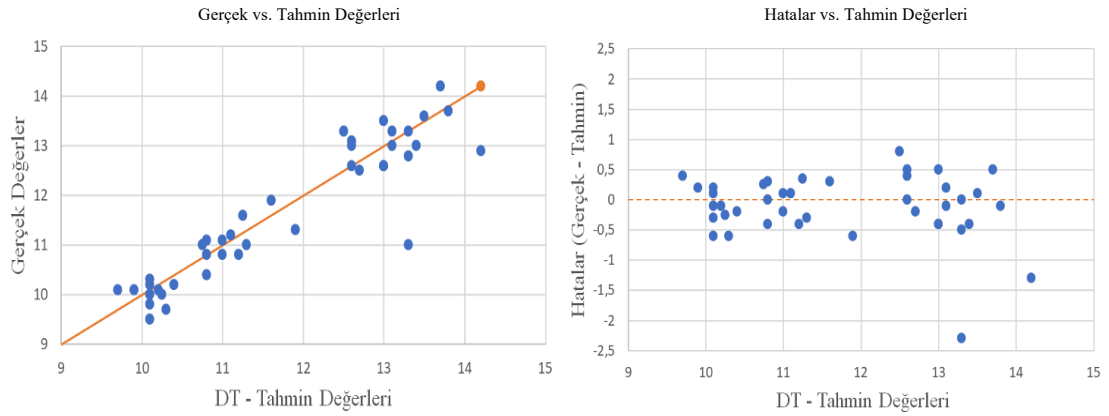
Şekil 27: Değişken Seçimi Olmadan “Makro + AR(3)” Ridge Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 27 'de çalışma sonucunda “Makro + AR(3)” grubunda en iyi performans gösteren Ridge modeline istinaden oluşturulmuş iki grafik yer almaktadır. Buna göre solda yer alan dağılım grafiği, işsizlik oranının gerçek değerleri ve Ridge modeli ile elde edilmiş tahmin değerleri arasındaki ilişkiyi göstermektedir. Grafikte yer alan noktalar genellikle referans doğrusuna yakın bir konumdadır. Bu da model ile elde edilen tahmin değerlerinin çoğunlukla gerçek değerlere yakın olduğu fikrini vermektedir. Ancak referans doğrusuna uzak olan bazı noktalar, tahmin değerlerinin her zaman tutarlı olamayacağını gösteriyor olabilir. Sağda ise hata değerlerinin, yatay ekseninde gösterilen Ridge modeli tahmin değerlerine göre konumlandığı bir grafik yer almaktadır. Hata değerlerinin dikey ekseninde -1 ve +1 arasında konumlanması Ridge modelinin, genel olarak gerçek değerlerden kabul edilebilir bir sapmaya sahip olduğunu göstermektedir. Ancak şekilde yer alan bazı uç değerler, modelin beklenenden daha yüksek veya düşük tahminler yapma eğilimini göstermektedir.

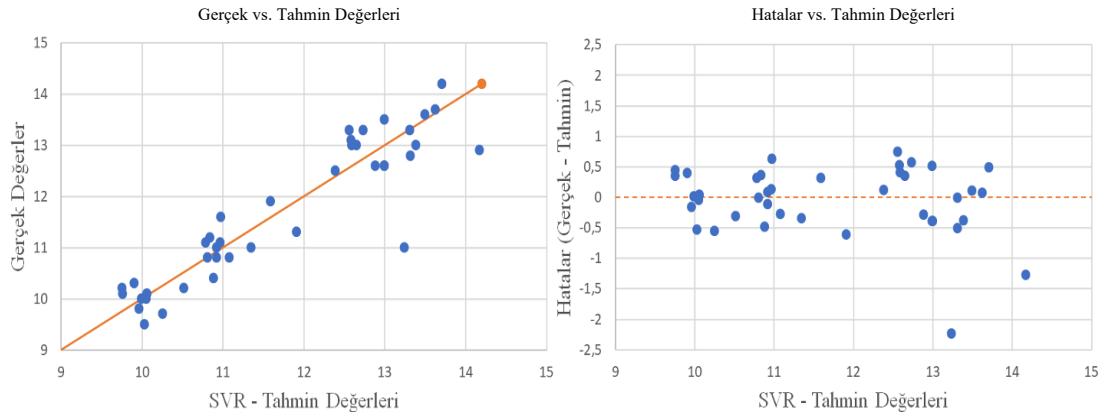
Şekil 28: Değişken Seçimi Olmadan “Haftalık GT + Makro + AR(3)” DT Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 28 'de çalışma sonucunda “Haftalık GT + Makro + AR(3)” grubunda en iyi performanslardan birini gösteren DT modeline istinaden oluşturulmuş iki grafik yer almaktadır. Sağdaki grafikte referans doğrusuna yakın ve doğrusal trende sahip noktalar, modelin değişkenler arasındaki ilişkileri iyi bir şekilde yakalayarak doğru tahminler yaptığı bilgisini vermektedir. Sağdaki grafiğe göre, çoğunlukla -1 ve +1 arasında toplanan hata değerleri, modelin kabul edilebilir bir hata payına sahip olduğunu gösterir. Hataların geniş yayılımı, modelin bazı noktalarda gerçek değerlerden önemli ölçüde sapabileceğine işaret eder. Bu durum, özellikle karar ağaçlarının aşırı dallanma eğiliminde olduğu ve verinin bazı bölümlerinde aşırı öğrenme yapabileceği durumlar için geçerlidir.

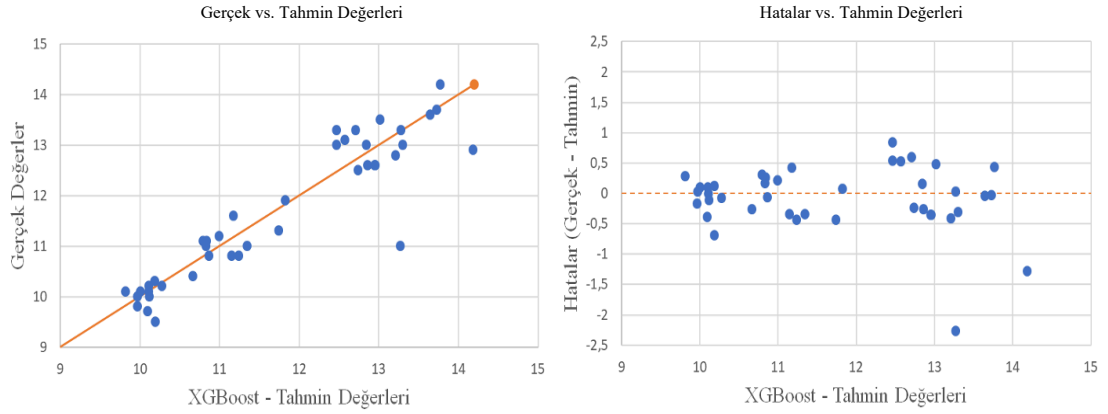
Şekil 29: Değişken Seçimi Olmadan “Tümü” SVR Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 29 'da çalışma sonucunda “Tümü” grubunda en iyi performanslardan birini gösteren SVR modeline istinaden oluşturulmuş iki grafik yer almaktadır. Veri noktalarının genel olarak referans doğrusu etrafında birikmesi SVR modelinin iyi bir doğruluk seviyesine sahip olduğunu göstermektedir ve modelin farklı değer aralıklarında geçerli tahminlerde bulunduğunu ifade etmektedir. Böylelikle, SVR modelinin marjinal hataları iyi bir şekilde optimize ettiği çıkarımı yapılabilir. Marjinal hataları iyi bir şekilde optimize etmek, SVR modelinin güçlü olduğu özelliklerinden birisidir. Uç nokta sayısındaki azlık, SVR modelinin yüksek değerlerde kabul edilebilir bir performans sergilediğini göstermektedir. Sağda ise, hataların çoğu -1 ile +1 arasında, yani makul bir hata marjı içinde yoğunlaşmıştır. Bu, modelin çoğu durumda güvenilir tahminler yaptığını göstermektedir. Nispeten simetrik dağılan hatalar, modelin ne aşırı tahmin (overestimation) ne de eksik tahmin (underestimation) eğiliminde olduğunu göstermektedir. Genel olarak, SVR modelinin veri setinde iyi bir genelleme başarısına sahip olduğu söylenebilir.

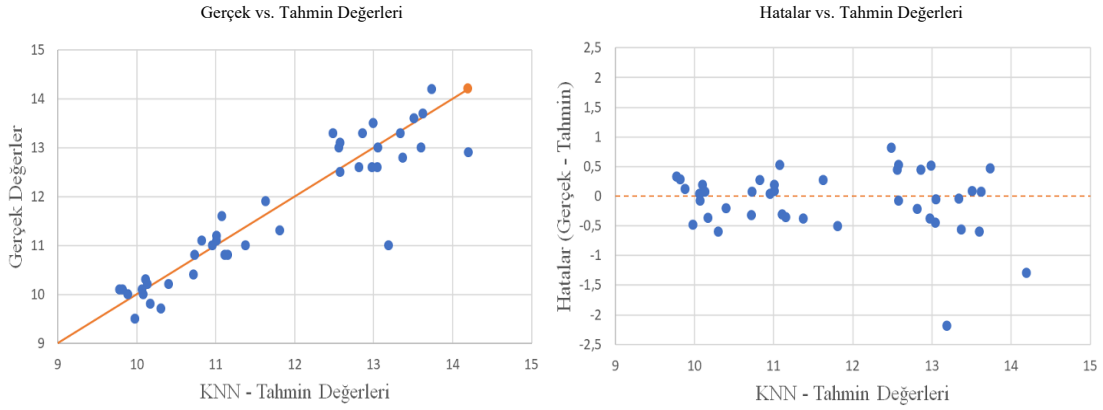
Şekil 30: Değişken Seçimi Olmadan “Tümü” XGBoost Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 30 'da çalışma sonucunda “Tümü” grubunda en iyi performanslardan birini gösteren XGBoost modeline istinaden oluşturulmuş iki grafik yer almaktadır. Noktaların referans doğrusuna yakın bir yerde yoğunlaşması, modelin verilerin geniş bir aralığında genelleştirme kabiliyetinin iyi olduğunu gösterir. Özellikle kompleks verilerde güçlü performansı ile bilinen XGBoost, bu grafikte de genel eğilimi yansıtıyor gibi görünmektedir. Sağdaki grafikte ise, hataların büyük bir kısmı -1 ile +1 arasında toplanmıştır, bu da modelin çoğunlukla kabul edilebilir bir hata marjı içinde çalıştığını gösterir. Hataların yatay eksende eşit olarak yayılması, modelin değişken tahmin seviyelerinde tutarlı bir hata oranını sabit tuttuğunu göstermektedir. Bu durum, modelin üzerinde çalıştığı veri setinin çeşitli yönlerini iyi bir şekilde yakaladığı anlamına gelmektedir.

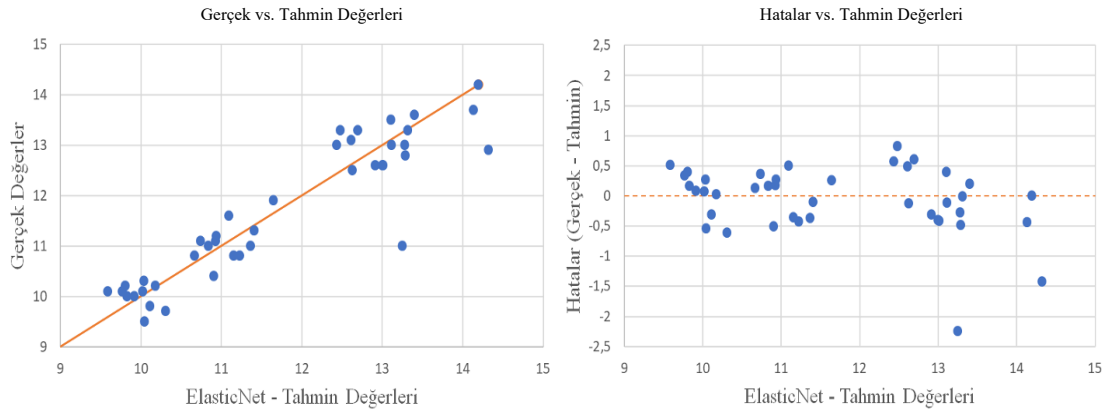
Şekil 31: LASSO Değişken Seçimi ile “Günlük GT + Makro + AR(3)” KNN Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 31 'de çalışma sonucunda “Günlük GT + Makro + AR(3)” grubunda en iyi performanslardan birini gösteren KNN modeline istinaden oluşturulmuş iki grafik yer almaktadır. Buna göre soldaki grafikte, noktalar çoğunlukla $x = y$ referans doğrusuna yakındır. Bu da modelin genel olarak iyi performans gösterdiğini ifade etmektedir. Şekle göre artan tahmin değerlerine yönelik gerçek değerlerin daha düşük olma eğilimi, modelin yüksek değerlerde tahmin performansının düşebileceği hakkında bilgi vermektedir. Sağdaki grafikte ise hata değerleri çoğunlukla -1 ve +1 arasındadır ve dolayısıyla model genel olarak kabul edilebilir sapmalara sahiptir. KNN modelinin yerel özelliklere (yani veri noktalarının yakınındaki komşu noktalara benzer olma durumu) duyarlılığı bazı durumlarda tutarsız tahminlere yol açabilir.

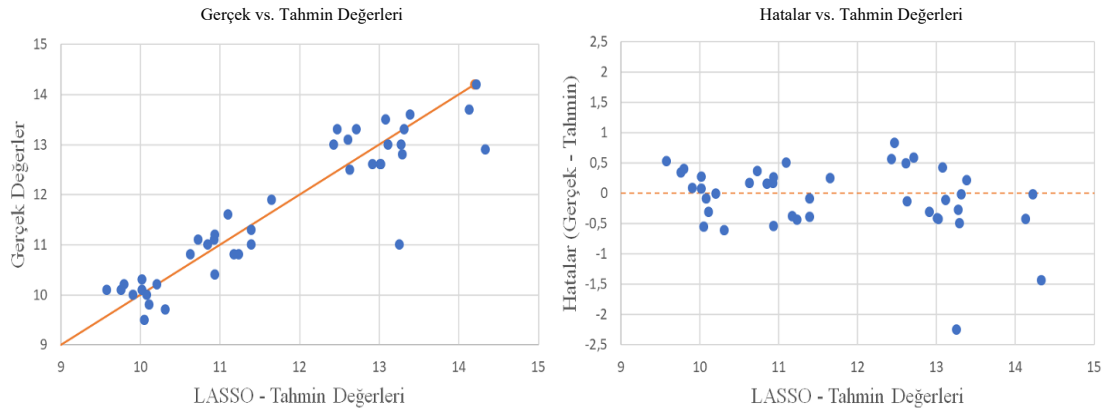
Şekil 32: LASSO Değişken Seçimi ile “Tümü” ElasticNet Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 32 'de çalışma sonucunda “Tümü” grubunda en iyi performanslardan birini gösteren ElasticNet modeline istinaden oluşturulmuş iki grafik yer almaktadır. ElasticNet modeli tarafından yapılan tahminler gerçek değerlerle karşılaştırıldığında, noktaların genel olarak referans doğrusuna yakın olduğu göze çarpmaktadır. Bu, modelin veri seti üzerinde iyi bir genelleme yapabildiğini gösterir. Bu durum, modelin çoğu zaman gerçek değerleri yakından takip edebildiğini göstermektedir. Bazı tahminlerde gerçek değerlerden sapmalar olsa da bu durum modelin bazı veri noktalarında mükemmel olmadığını ancak genel olarak iyi bir performans sergilediğini gösterir. Sağda ise, hata değerleri, çoğunlukla -1 ile +1 arasında toplanmıştır, bu da modelin çoğu tahminde kabul edilebilir bir hata marjı içinde çalıştığını gösterir. Hata değerlerinin simetrik bir dağılım göstermesi, modelin ne aşırı tahmin yapma ne de eksik tahmin yapma eğiliminde olduğunu gösterir. Genel olarak, ElasticNet modelinin LASSO değişken seçimi ile birleştirildiğinde, gereksiz değişkenleri modelden çıkararak daha etkili tahminler yaptığı söylenebilir.

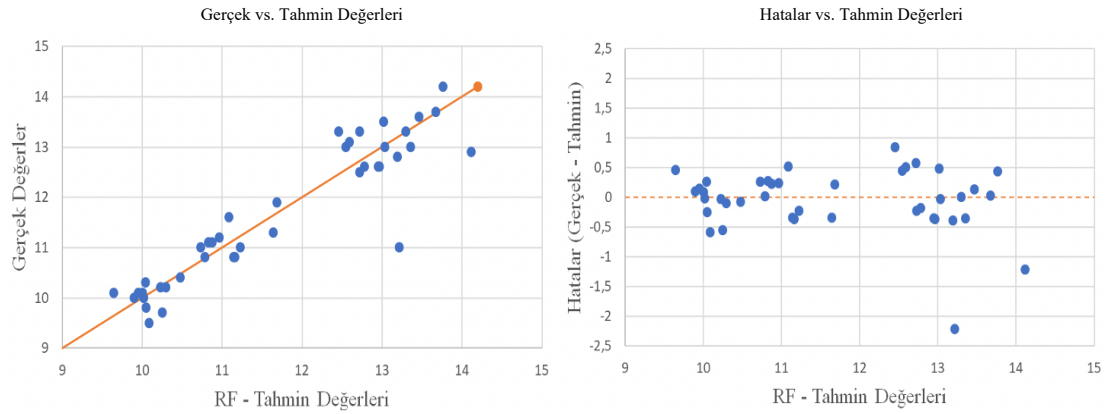
Şekil 33: LASSO Değişken Seçimi ile “Tümü” LASSO Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 33 'de çalışma sonucunda “Tümü” grubunda en iyi performanslardan birini gösteren LASSO modeline istinaden oluşturulmuş iki grafik yer almaktadır. LASSO modeli kullanılarak yapılan tahminlerin gerçek değerlere oldukça yakın olduğu görülmektedir. Bu, modelin doğruluk açısından iyi bir performans sergilediğini göstermektedir. Noktaların çoğu referans doğrusuna yakındır. Böylece modelin veri setini iyi bir şekilde genelleştirdiği ve tutarlı tahminler yaptığı söylenebilir. Noktalar arasında, özellikle daha yüksek tahmin değerlerinde, doğrudan sapmalar gözlemlenmektedir. Bu durum modelin belirli değer aralıklarındaki tahminlerde hatalar yapabileceğini göstermektedir. Sağda ise, hata değerlerinin yine -1 ve +1 arasında yoğunlaşmış durumda olması, modelin çoğu tahminde kabul edilebilir bir hata marjı içinde olduğunu gösterirken, modelin farklı tahmin seviyelerinde tutarlı bir hata oranını koruduğuna işaret etmektedir. Bu durumda modelin veri setinin çeşitli yönlerini iyi bir şekilde yakaladığı anlamı çıkarılabilir.

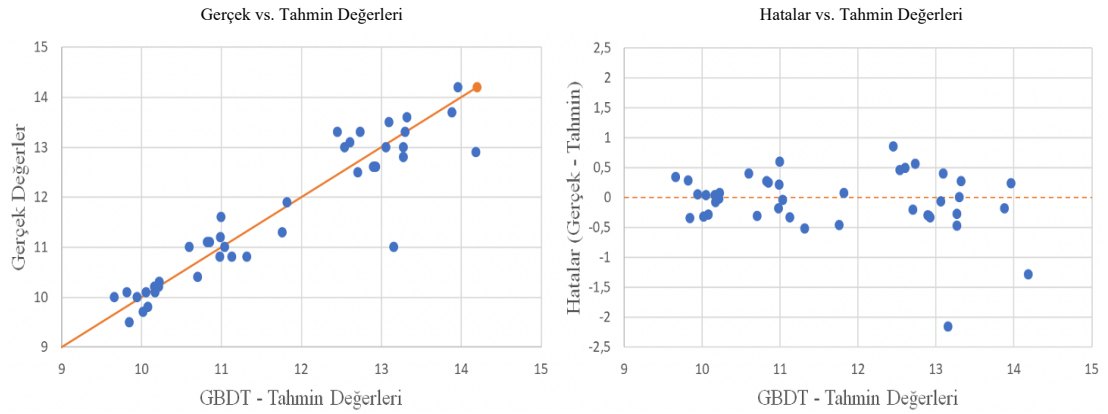
Şekil 34: LASSO Değişken Seçimi ile “Tümü” RF Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 34 'de çalışma sonucunda “Tümü” grubunda en iyi performanslardan birini gösteren RF modeline istinaden oluşturulmuş iki grafik yer almaktadır. RF modelinin tahminleri ile gerçek değerler arasında genellikle iyi bir uyum olduğu görülmektedir. Birçok veri noktası, referans doğrusuna yakın bir konumda yer almaktadır, bu da modelin yüksek doğrulukta tahminler yaptığına işaret etmektedir. Veri noktalarının referans doğrusuna yakınlığı, modelin farklı değer aralıklarında iyi genelleştirme yapabildiğini göstermektedir. Yüksek tahmin değerleri için bile modelin gerçek değerlere oldukça yakın tahminler ürettiği görülmektedir. RF modeli, genellikle veri setlerindeki gürültüye ve aşırı değerlere karşı direnç gösterir ve bu grafikte de tahminlerin sağlamlığı gözlemlenmiştir. Sağda ise, hata değerlerinin yine çoğunlukla -1 ile +1 arasında bir aralıkta yoğunlaşmış olduğu görülmektedir. Hataların yayılımı yatay eksenle tutarlı bir şekilde dağılmıştır. Yani model farklı tahmin seviyelerinde benzer hata düzeylerini korumuştur. Bu durum, modelin genel olarak güvenilir olduğunu göstermektedir.

Şekil 35: LASSO Değişken Seçimi ile “Tümü” GBDT Modeli



Kaynak: Tez Yazarı Tarafından Hazırlanmıştır.

Şekil 35 'de çalışma sonucunda “Tümü” grubunda en iyi performanslardan birini gösteren GBDT modeline istinaden oluşturulmuş iki grafik yer almaktadır. GBDT modeli tahmin değerlerinin gerçek değerlere oldukça yakın olduğu görülmektedir. Böylece modelin veri seti üzerinde iyi bir genelleme yapabildiği ve yüksek doğrulukta tahminler ürettiği anlamı çıkarılmaktadır. Noktaların referans doğrusu boyunca düzenli olarak artması, modelin çeşitli tahmin seviyelerinde tutarlı ve doğru tahminler yaptığını göstermektedir. Bu durum, GBDT 'nin karmaşık doğrusal olmayan ilişkileri modelleyebilme kapasitesini yansıtmaktadır. Genellikle, veri setlerinin karmaşık yapılarını ve yüksek boyutluluklarını iyi işlemesiyle bilinen GBDT modelinin, verilere gayet iyi uyum sağladığı görülmektedir. Sağda ise, hata değerleri yine -1 ile +1 arasında toplanmıştır. Hatalar yatay eksende nispeten düzgün bir şekilde dağılmıştır. Bu nedenle, modelin farklı tahmin seviyelerinde tutarlı hata oranlarına sahip olduğu anlamı çıkarılabilir. Sonuç olarak modelin veri setinin çeşitli özelliklerini iyi bir şekilde yakaladığı söylenebilir. Hataların genel olarak simetrik bir dağılım göstermesi ve çok az sayıda uç değer olması, modelin ne aşırı tahmin yapma ne de eksik tahmin yapma eğiliminde olduğunu göstermektedir. Genel olarak, LASSO değişken seçimi ile desteklenen GBDT modeli hem grafiklere hem de performans metriklerinde elde ettiği değerlere göre veri seti üzerinde iyi bir performans sergilemiştir.

SONUÇ

Bu çalışma kapsamında işsizlik oranı öngörüsü gerçekleştirilirken Google Trends verileri ve makroekonomik göstergeleri bir arada kullanarak farklı frekanslı değişkenlerin Türkiye 'deki işsizlik oranı öngörümlemesine olan etkisi incelemek amaçlanmıştır. Tahminleme işlemi gerçekleştirilirken denetimli makine öğrenmesinin regresyon tabanlı tekniklerinden yararlanılmıştır. Çeşitli veri kaynaklarının entegrasyonunun, ekonomik analizlerdeki önemi araştırılarak hem ekonomik karar alma süreçlerine katkıda bulunmak hem de araştırmacılara farklı ve kapsamlı bir bakış açısı sunmak hedeflenmiştir.

Analizin başarılı bir şekilde gerçekleştirilmesi için belirli bir bağlamın netleştirilmesi ve farklı frekanstaki verilerin doğru bir biçimde entegre edilmesi büyük önem taşır. Bu sebeple, zaman serisi analizini yürütebilmek adına işsizlik oranı bağımlı değişkeninin çalışma içindeki tarih aralığının doğru bir şekilde belirlenmesi kritik bir adımdır. Bu belirlemenin doğru yapılabilmesi için çalışmada kullanılacak tüm verilere erişim sağlanabilir olmalıdır. Çünkü, tüm değişkenler için aynı zaman diliminde veri elde etmek önemlidir. Ancak, bazı durumlarda belirli bir tarihten önce bazı değişkenlerin verilerine erişmek mümkün olmayabilir. Bu durumda, analizde kullanılacak verilere erişilebilirlik, analizde kullanılacak tarih aralığının belirlenmesinde kritik bir rol oynar. Ayrıca, Google Trends verileri çalışmada yer aldığı için, toplumun yıllar içindeki internet kullanım oranlarının da dikkate alınması gerekmektedir. Google Trends arayüzünden alınan arama sıklıklarının çalışmanın kapsamında anlamlı sonuçlar üretebilmesi için, ilgili dönemde toplumdaki internet kullanım oranının kayda değer olması gerekmektedir. Bu bağlamda, verilere erişilebilirlik ve internet kullanım oranları dikkate alınarak, 2010 1. ayı ile 2023 5. ayına kadar olan dönemde gerçekleşen işsizlik oranları üzerinde aylık ölçekte TÜİK verileri kullanılarak zaman serisi analizi yapılmıştır.

Çalışmada, işsiz bireylerin internet arama davranışlarının analize dahil edilmesi için, internette aratabilecekleri “iş” anahtar kelimesinden yola çıkarak, Türkiye 'de yapılan benzer çalışmalardan ve Google Trends platformunun benzer bağlamli sorguları öneren “ilgili sorgular” aracından yararlanılmış ve 13 farklı anahtar arama sorgusu belirlenmiştir. Bu sorguların farklı frekans değerleri günlük, haftalık ve

aylık ölçeklerde olmak üzere elde edilmiştir. Ancak, Google Trends 'in uzun tarih aralıklarındaki verileri tek seferde çekmeye izin vermemesi nedeniyle, veriler parçalı olarak çekilmiş ve ardından belirli tekniklerle birleştirilmiştir. Google Trends arama sıklığına ait veriler 0 ile 100 arasında ölçekli ve aynı zamanda örnekleme yapılarak elde edildiğinden bu verilerin direkt olarak kullanılması örneklem değişkenliğine ait gürültüyü analize katabileceğinden sağlıklı olmayabilir. Bu nedenle, veri çekme işlemi 10 günlük periyotlarla tekrarlanmış, çekilen bu tekrarlı verilerin ortalamaları alınmış ve birleştirilerek günlük, haftalık ve aylık ölçeklerde veri setleri oluşturulmuştur. Ayrıca, çalışmanın kapsamını genişletmek ve farklı değişkenlerin etkisini değerlendirebilmek amacıyla analize uygun makroekonomik dışsal göstergeler belirlenmiştir. TÜFE, ekonomik güven endeksi, döviz kuru, istihdam sayısı ve 15+ yaş nüfus sayısından oluşan bu göstergeler; TÜİK ve TCMB üzerinden aylık ölçeklerde elde edilmiştir ve araştırmanın temel veri setine eklenmiştir.

Çalışma, beş temel model üzerinde yoğunlaşmıştır. Her model için ElasticNet, LASSO, KNN, SVR, DT, RF, GBDT, XGBoost ve Ridge regresyon modelleri kurulmuştur. İlk modelde, 18 bağımsız değişken; makroekonomik değişkenlerin ve işsizlik oranının gecikmeleri alınarak incelenmiştir. İkinci modelde, makroekonomik değişkenlerin, işsizlik oranının ve günlük Google Trends (GT) değişkenlerinin gecikmeleri alınarak 109 bağımsız değişken incelenmiştir. Üçüncü modelde, makroekonomik değişkenlerin, işsizlik oranının ve haftalık GT değişkenlerinin gecikmeleri alınarak 70 bağımsız değişken değerlendirilmiştir. Dördüncü modelde, makroekonomik değişkenlerin, işsizlik oranının ve aylık GT değişkenlerinin gecikmeleri alınarak 57 bağımsız değişken analiz edilmiştir. Son olarak, beşinci modelde, makroekonomik değişkenlerin, işsizlik oranının, günlük, haftalık ve aylık GT değişkenlerinin gecikmeleri birleştirilerek 200 bağımsız değişken üzerinden modeller oluşturulmuştur. Daha sonra, tüm modeller LASSO değişken seçim tekniği ile tekrarlanarak, değişken seçiminin model performansına etkisi değerlendirilmiştir.

Analiz sonuçları, aynı değişkene ait farklı frekanslı verilerin ayrı bir değişken olarak model içerisinde birleştirildiği kombinasyonlarda en üstün sonuçların elde edildiğini ortaya koymuştur. Günlük, haftalık ve aylık ölçekteki Google Trends verilerinin, makroekonomik dışsal değişkenlerle ve serinin gecikmeleriyle birleştirildiği modelde, değişkenler en iyi performansı sergilemiş ve en düşük hatalı

tahminler elde edilmiştir. Değişken seçimi denemeleri sonucunda, dokuz denetimli makine öğrenme tekniğinden beş tanesinin özellikle etkili olduğu görülmüştür. Tüm modellerin en iyi performans sergilediği gruplar belirlenmiş ve bu modeller içerisinde GBDT modeli en yüksek performans değerlerine ulaşan model olduğu belirlenmiştir. GBDT modeli hem farklı frekanslı Google Trends hem de makroekonomik verilerden oluşan veri setinde en düşük hata oranını göstermiştir. Bu sonuç farklı frekanslı Google Trends verilerinin birbirinden farklı bilgiler taşıyarak öngörü performansını arttırmada fayda sağlayacağı gerçeğine yönelik olarak bir bulgu niteliği taşımaktadır. Aylık gibi daha düşük frekanslı verilerin genel trend hakkında daha fazla bilgi sağlarken, günlük arama terimlerinden oluşan daha yüksek frekanslı verilerin ise gerçekleşen yeni durumlara modelin daha hızlı uyum sağlama konusunda fayda sağlamış olabileceği düşünülmektedir.

Genel olarak, bu çalışma farklı frekanslardaki değişkenlerin işsizlik oranı öngörüsünde etkili bir strateji olabileceğini ortaya koyarken, tüm değişkenlerin birleştirilirken önemli değişkenlerin seçimi yapılarak GBDT modelinin kullanılmasının en iyi sonuçları verdiği sonucuna ulaşmıştır. Ayrıca, işsizlik öngörüsünde, işsiz bireylerin sıklıkla arayabileceği sorgu sayılarından elde edilen Google Trends verilerinin, makroekonomik göstergelerle birlikte kullanımının tahmin performansını artırdığı ve işsizlik oranı üzerinde daha doğru tahminler yapılmasına olanak tanıdığını göstermiştir.

Gelecekte gerçekleştirilecek çalışmalarda, Google Trends verilerinde daha geniş bir anahtar kelime yelpazesi incelenebilir. Değişken seçimi teknikleri arasında LASSO yerine ElasticNet veya SVM tabanlı alternatif yöntemler değerlendirilebilir. Ya da birden çok değişken seçim tekniği kullanılarak çalışma kapsamında oluşturulmuş modeller tekrarlanabilir ve hangi değişken seçim tekniğinin model performansı üzerinde daha etkili olduğu araştırılabilir. Ayrıca, bu çalışmada kullanılan makine öğrenme algoritmalarına, sığ ve derin öğrenme tabanlı yapay sinir ağıları modelinin dahil edilmesi düşünülebilir. Son olarak, çalışma işsizlik oranları üzerinde odaklanmış olsa da farklı makroekonomik göstergelerin, ekonomik değişkenlerin ve sosyal faktörlerin bir araya getirilmesi, daha kapsamlı bir analiz sunabilir. Bu faktörlerin birleştirilmesiyle, daha geniş bir tahmin yapma kapasitesi ve ekonomik trendlerin daha ayrıntılı bir şekilde anlaşılması mümkün olabilir.

KAYNAKÇA

Akkaya, T. (2017). *Türkiye 'de İşsizlik Olgusu Bağlamında İş Beğenmeme Nedenleri İstanbul İli Örneği*. (Uzmanlık Tezi). Ankara: T.C. Çalışma ve Sosyal Güvenlik Bakanlığı Türkiye İş Kurumu Genel Müdürlüğü.

Alpaydın, E. (2010). *Introduction to Machine Learning* (Second Edition.). USA: The MIT Press.

ARAS, S. ve Lisboa, P.J.G. (2022). Explainable inflation forecasts by machine learning models. *Expert Systems with Applications*. 207: 117982.

Armesto, M. ve Engemann, K. (2010). Forecasting with Mixed Frequencies. *Federal Reserve Bank of St.Louis Review*, 92(3): 521-536.

Arya, N. (2023). *K-Means Clustering for Unsupervised Machine Learning*, <https://www.ejable.com/tech-corner/ai-machine-learning-and-deep-learning/k-means-clustering/>, (20.10.2023).

Askatas, N. ve Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*. 55(2): 107-120.

Awad, M. ve Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Springer.

Ay, S. (2012). Türkiye 'de İşsizliğin Nedenleri: İstihdam Politikaları Üzerine Bir Değerlendirme. *Yönetim ve Ekonomi Dergisi*. 19(2): 321-341.

Aydın, N. (2012). İşsizlik, Beveridge Eğrisi ve Türkiye İşgücü Piyasasına İlişkin Bir Değerlendirme. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*. 2(32): 119-134.

Bağcı, E. (2020). *Türkiye 'de İşsizlik sorunu ve İstihdam Politikaları*. İstanbul: DER Yayınları.

Baker, S. R. ve Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from Google search data. *Review of Economics and Statistics*. 99(5): 756-768.

Barro, R.J. (1997). *Macroeconomics* (Fifth Edition.). London: The MIT Press.

Bolivar, F., Ortiz, A. ve Rodrigo, T. (2019). Nowcasting Turkish Unemployment Using Real Time Data From Google. *BBVA Research*.

Borup, D. ve Schütte, E. C. M. (2022). In search of a job: Forecasting employment growth using Google Trends. *Journal of Business and Economic Statistics*. 40(1): 186–200.

Borup, D., Rapach, D. E. ve Schütte, E. C. M. (2023). Mixed-frequency machine learning: Nowcasting and backcasting weekly initial claims with daily internet search volume data. *International Journal of Forecasting*. 39(3): 1122-1144.

Borup, D. ve Schütte, E. C. M. (2022). In search of a job: Forecasting employment growth using Google Trends. *Journal of Business & Economic Statistics*. 40(1): 186-200.

Breiman, L. (1996). Arching the edge. *Statistics Department University of California Technical Report 486*. 1-14.

Breiman, L. (2001). Random forests. *Machine learning*. 45(1): 5-32.

Brownlee, J. (2016). *Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, And Work Projects End-To-End* (Edition: v1.4.).

Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Quebec City, QC, Canada: Andriy Burkov.

Caperna, G., Colagrossi, M., Geraci, A. ve Mazzearella, G. (2020). Googling unemployment during the pandemic: Inference and nowcast using search data. *JRC Working Papers in Economics and Finance*. 1-19.

Castelnuovo, E. ve Tran, T. D. (2017). Google it up! a Google Trends-based uncertainty index for the United States and Australia. *Economics Letters*. 161: 149-153.

Chadwick, M. G. ve Sengül, G. (2015). Nowcasting the unemployment rate in Turkey: let 's ask google. *Central Bank Review*. 15(3): 1-24.

Choi, H. ve Varian, H. (2012). Predicting the present with Google Trends. *Economic record*. 88(s1): 2-9.

Chollet, F. (2021). *Deep Learning with Python* (Second Edition.). USA: Manning Publications.

Dangeti, P. (2017). *Statistics for Machine Learning: Build supervised, unsupervised, and reinforcement learning models using both Python and R*. Birmingham, UK.: Packt Publishing Ltd.

Deng, H., Zhou, Y., Wang, L. ve Zhang, C. (2021). Ensemble learning for the early prediction of neonatal jaundice with genetic features. *Bmc Medical Informatics and Decision Making*. (21)338: 1-11.

Eğilmez, M. (2023). *Ekonominin Temelleri (Kavramlar ve Kurumlar)* (7. Baskı.). İstanbul: Remzi Kitabevi.

Eyüboğlu, D. (2003). *2001 Krizi Sonrasında İşsizlik ve Çözüm Yolları*. Ankara: Milli Prodüktivite Merkezi Yayınları.

Friedman, J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Institute of Mathematical Statistics*. 29(5):1189-1232.

Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd Edition.). Canada: O'Reilly Media.

Ghysels, E., Santa-Clara, P. ve Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models.

Grybauskas, A., Pilinkienė, V., Lukauskas, M., Stundžienė, A. ve Bruneckienė, J. (2023). Nowcasting Unemployment Using Neural Networks and Multi-Dimensional Google Trends Data. *Economies*. 11(5): 1-23.

GT. (2023). *Keşfet*. <https://trends.google.com/trends/> , (26.10.2023).

Gunn, S. R. (1998). Support vector machines for classification and regression. *Isis technical report*. 14(1): 1-54.

Gupta, A. (2023). *Semi-Supervised Learning in ML*. <https://www.geeksforgeeks.org/ml-semi-supervised-learning/>, (18.09.2023).

Gündoğan, N. (2022). *İşsizlik*. <https://ansiklopedi.tubitak.gov.tr/ansiklopedi/issizlik>, (04.09.2023).

Gündoğan, N., Biçerli, M.K. ve Özkaplan, N. (2019). Çalışma Ekonomisinin Temel Kavramları, İşsizlik ve Emek Piyasası Politikaları. *Çalışma Ekonomisi* (ss. 2-19, 164-191). Eskişehir: Anadolu Üniversitesi.

Hallman, J. (2019). *A Comparative Study On Linear Regression And Neural Networks For Estimating Order Quantities Of Powder Blends*. (Degree Project). Stockholm, Sweden: Kth Royal Institute Of Technology.

ILO. (2023), *Report: Global Employment Policy Review 2023: Macroeconomic Policies For Recovery and Structural Transformation*.
https://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_882222.pdf, (05.09.2023).

IMF. (2022), *World Economic Outlook Report October 2022: Countering The Cost-Of-Living Crisis*.
<https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022>, (04.09.2023).

İŞKUR. (2023), *Aylık İstatistik Bültenleri*. <https://www.iskur.gov.tr/kurumsal-bilgi/istatistikler/>, (15.09.2023).

Jo, T. (2021). *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Cham, Switzerland: Springer International Publishing.

Julian, D. (2016). *Designing Machine Learning Systems With Python: Design efficient machine learning systems that give you more accurate results*. UK: Packt Publishing Ltd.

Kanca, O.C. (2012). Türkiye 'de İşsizlik ve İktisadi Büyüme Arasındaki Nedenselliğin Ampirik Bir Analizi. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*. 21(2): 1-18.

Khanzode, G. (30.08.2015). *Machine Learning*.
<https://www.slideshare.net/GirishKhanzode/supervised-learning-52218215>, (18.09.2023).

Koç, E. (07.02.2022). *Gelişmiş Ağaç Yöntemleri*. <https://medium.com/@kocelifk/geli%C5%9Fmi%C5%9F-a%C4%9Fa%C3%A7-y%C3%B6ntemleri-340445460f8b>, (26.10.2023).

Kor, K. (2015). *Penetration Rate Optimization With Support Vector Regression Method*. (M.Sc. Thesis). İstanbul: Istanbul Technical University Graduate School Of Science Engineering And Technology.

Lantz, B. (2019). *Machine Learning With R: Expert techniques for predicting modeling* (Third Edition.). UK: Packt Publishing.

Mali, K. (2021). *Everything you need to Know about Linear Regression!*. <https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/>, (21.09.2023).

Mcculloch, W.S. ve Pitts, W. (1990). A logical calculus of the ideas immanent in nervous activity. *Department of Psychiatry at the Illinois Neuropsychiatric Institute*. 52(No.1/2): 99-115.

Nacar, E.N. ve Erdebilli, B. (2021). Makine Öğrenmesi Algoritmaları ile Satış Tahmini. *Endüstri Mühendisliği Dergisi*. 32(2): 307-320.

OECD. (2023), *Unemployment Rate*. <https://data.oecd.org/unemp/unemployment-rate.htm>, (09.10.2023).

Otker Robe, I. ve Podpiera, A.M. (2013). The Social Impact of Financial Crises Evidence from the Global Financial Crisis. *The World Bank Policy Research Working Papers*. No. 6703: 1-34.

Pekin, T. (1996). *Makro Ekonomi* (4. Baskı.), İzmir: Bilgehan Basımevi.

Rafifaghihi, N. (2022). *From Theory to Practice: Implementing Support Vector Regression for Predictions in Python*. <https://medium.com/@niiousha.rf/support-vector-regressor-theory-and-coding-exercise-in-python-ca6a7dfda927>, (18.10.2023).

Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *2019 International Conference On Machine Learning, Big Data, Cloud And Parallel Computing (Com-It-Con)*.

Saini, A. (2021). *Decision Tree Algorithm – A Complete Guide*. <https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>, (26.10.2023).

Sharma, H. ve Kumar, S. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *International Journal Of Science And Research (IJSR)*, 5(4): 2094-2097.

Smola, A. (1996). *Regression Estimation with Support Vector Learning Machines*.

Şahin, L. (2007). Türkiye İşgücü Piyasasının Yapısal Özellikleri ve İşsizlik Sorunu. *Sosyal Siyaset Konferansları Dergisi*. (53): 543-575.

Şentürk, G. (2022). Can Google search data improve the unemployment rate forecasting model? An empirical analysis for Turkey. *Journal of Economic Policy Researches*. 9(2): 229-244.

Tansel, A. (2012). *2050 'ye Doğru Nüfusbilim ve Yönetim: İşgücü Piyasasına Bakış*. İstanbul: Sis Matbaacılık.

Taş, H.Y. (2018). Dördüncü Sanayi Devrimi 'nin (Endüstri 4.0) Çalışma Hayatına ve İstihdama Muhtemel Etkileri. *Uluslararası Toplum Araştırmaları Dergisi*. 9(16): 1817-1836.

Tuhkuri, J. (2015). *Big Data: Do Google Searches Predict Unemployment?*. (Master's Thesis). Helsinki: University of Helsinki Faculty of Social Sciences.

TÜİK. (10.08.2023). *Hanehalkı İşgücü İstatistikleri*.
<https://data.tuik.gov.tr/Kategori/GetKategori?p=istihdam-issizlik-ve-ucret-108&dil=1#data1>, (17.09.2023).

TÜİK. (17.11.2023). *Mevsim etkisinden arındırılmış işgücüne ilişkin tamamlayıcı göstergeler*. <https://data.tuik.gov.tr/Kategori/GetKategori?p=istihdam-issizlik-ve-ucret-108&dil=1>, (17.11.2023).

Unay, C. (1996). *Makro Ekonomi* (6. Baskı.), Bursa: Ekin Kitabevi.

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*. New York: Springer Science+ Business Media, Inc.

Woloszko, N. (2020). Tracking activity in real time with Google Trends. *Economics Department Working Papers No. 1634*. 1-54.

Yolasiğmaz, İ.B. (2015). *A support vector regression method for conceptual cost estimate of construction projects*. (Master's Thesis). Ankara: The Graduate School Of Natural And Applied Sciences Of Middle East Technical University.

Zafari, A., Abdolazadeh, R. ve Kianmehr, M.H. (2013). Modeling the effect of extrusion parameters on density of biomass pellet using artificial neural network. *International Journal Of Recycling of Organic Waste in Agriculture*. (2)9: 1-11.

Zaim, S. (1997). *Çalışma Ekonomisi* (10. Baskı.), İstanbul: Filiz Kitabevi.

Zengin, Ş. (2000). *Türkiye 'de İşsizlik Sigortasının Uygulanabilirliği*. (Yayınlanmamış Yüksek Lisans Tezi). Eskişehir: Anadolu Üniversitesi Sosyal Bilimler Enstitüsü.