

Veri Madenciliği: Makine Öğrenme Algoritmaları ile Türkiye'nin İşsizlik Oranı Tahminini Etkileyen Faktörlerin Tespit Edilmesi

Data Mining: Identifying Factors Affecting Turkey's Unemployment Rate Estimation with Machine Learning Algorithms

DOI:10.33461/uybisbbd.1129013

İncilay YILDIZ¹ 

Öz

Makale Bilgileri

Makale Türü:

Araştırma Makalesi

Geliş Tarihi:

10.06.2022

Kabul Tarihi:

12.09.2022

©2022 UYBİSBBD
Tüm hakları saklıdır.



İşsizlik oranı, bir ülkenin ekonomik verimini ölçen en önemli değişkenlerden biridir. İş gücü ve verimi toplumun gelişmesi açısından çok önemli olup, işsizlik oranının düşürülmesi ve istihdam sağlanması bir devletin büyümesi açısından kilit öneme sahiptir. İşsizlik oranının tahmin edilmesi alınacak önlemler ve planlamalar açısından önem arz etmektedir. Bu çalışmada, işsizlik oranının tahmini için Türkiye İstatistik Kurumunun (TÜİK) İllerde Yaşam Endeksi verilerinden ve TÜİK tarafından oluşturulan veri gruplarından yararlanılmıştır. Tahmini etkileyen değişkenlerin tespiti için veri madenciliği sürecinden ve makine öğrenme algoritmalarından yararlanılmıştır. Bu çalışma kapsamında, işsizlik oranının tahmininde hangi tür veri gruplarının daha etkili olduğu ve hangi makine öğrenme yöntemlerinin daha etkin olduğu sorgulanmıştır. Elde edilen sonuçlar tahmin duyarlılık analizi kapsamında incelenmiş ve hata oranlarına göre yorumlanmıştır. Elde edilen sonuçlara göre işsizlik oranı tahmininde, Genel Regresyon Sinir Ağı (GRNN) yöntemi Destek Vektör Makineleri (SVM) ve Gen İfadesi Programlama (GEP) yöntemlerine göre daha etkin bir yöntemdir. Ayrıca, eğitim konulu veri grubunun diğer veri gruplarından daha etkili olduğu gözlemlenmiştir.

Anahtar Kelimeler: Veri Madenciliği, Makine Öğrenmesi, Tahmin, İşsizlik Oranı.

Article Info

Paper Type:

Research Paper

Received:

10.06.2022

Accepted:

12.09.2022

©2022 UYBİSBBD
All rights reserved.



Abstract

The unemployment rate is one of the most important variables that measure the economic efficiency of a country. Labor force and efficiency are very important for the development of society, reducing the unemployment rate and providing employment are key to the growth of a state. Estimating the unemployment rate is important in terms of measures and plans to be taken. In this study, for the estimation of the unemployment rate, data from the Life Index in Provinces of the Turkish Statistical Institute (TUIK) and data groups created by TUIK were used. Data mining process and machine learning algorithms were used to determine the variables affecting the estimation. Within the scope of this study, it has been questioned which types of data groups are more effective in estimating the unemployment rate and which machine learning methods are more effective. Obtained results were analyzed within the scope of estimation sensitivity analysis and interpreted according to error rates. According to the results obtained, General Regression Neural Network (GRNN) method is a more effective method than Support Vector Machines (SVM) and Gene Expression Programming (GEP) methods in estimating unemployment rate. In addition, it was observed that the education-themed data group was more effective than the other data groups.

Keywords: Data Mining, Machine Learning, Prediction, Unemployment Rate.

Atıf/ to Cite (APA): Yıldız İ., (2022). Veri Madenciliği: Makine Öğrenme Algoritmaları ile Türkiye'nin İşsizlik Oranı Tahminini Etkileyen Faktörlerin Tespit Edilmesi. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 6(2), 78-91

¹ Arş. Gör., Yönetim Bilişim Sistemleri, Uygulamalı Bilimler Fakültesi, Altınbaş Üniversitesi, İstanbul, Türkiye, incilay.yildiz@altinbas.edu.tr.

1. GİRİŞ

Günümüzde büyük hacimlerdeki veriler bankalar, işletmeler, Türkiye İstatistik Kurumu (TÜİK) ve bazı kuruluşlar tarafından depolanmaktadır. Bu veriler içindeki değerli bilgileri ortaya çıkarmak ya da yeni bir bilgi keşfetmek işletme karlılığı açısından oldukça önemlidir. Ayrıca, hedeflenen bir planlamanın yapılması için elde edilen bilgiler yol gösterici olarak kullanılabilir. Bu verilerden en doğru şekilde faydalanmak için, üzerinde doğru analizlerin yapılması gerekmektedir.

Verileri veri tabanlarında saklayan kurumların en önemli amaçlarından biri ham verileri bilgiye dönüştürmektir. Veriyi bilgiye dönüştürme işlemi veri madenciliği olarak adlandırılmaktadır. Veri toplama araçları ve veri tabanı teknolojilerindeki gelişmeler, bilgi depolarında çok miktarda bilginin depolanmasını ve çözümlenmesini gerektirmektedir (Dener vd., 2009).

Tahmin (Öngörü) gelecekte gerçekleşecek olan olayların önceden saptanması, hesaplanması ya da kestirimidir. Bir tahmin çalışmasının duyarlılığı, yapılan tahminlerle gerçek değerler arasındaki sapmaların analizi sonucu ortaya çıkmaktadır. Duyarlılık analizi gerçek değerler bilindikten sonra yapılabilir. Bu çalışmada TÜİK veri havuzu kullanılarak tahmin çalışması yapılmıştır.

Veri madenciliği ve uygulamaları tahminleme çalışmalarında en sık faydalanılan çalışma alanlarından biridir. Veri madenciliği uygulamasının verimini arttırmak için, çalışılan verilerin çeşitliliği, veri madenciliği algoritmasının etkinliği, sonuçların yararlılık düzeyi, kesinlik ve anlamlılık unsurlarının sağlanması, elde edilen örüntülerin çeşitli biçimlerde gösterimi, gizliliğin ve veri güvenliği özelliklerinin sağlanması gereklidir. Veri madenciliği bilgi keşfi sürecinin bir parçası şeklinde düşünülebilir (Dener vd., 2009).

Bir veri madenciliği sürecinin çeşitli fonksiyonları vardır. Bu çalışmada ilişkilendirme, regresyon ve sınıflandırma fonksiyonlardan yararlanılmıştır.

İlişkilendirme: bir veri tabanındaki verilerin mevcut ya da potansiyel ilişkilerini ortaya çıkarmayı hedefler (Gong, vd., 2012) İstatistikler ve apriori algoritmaları örnek olarak gösterilebilir (Ngai, vd., 2009).

Sınıflandırma: veri tabanı kayıtlarını belirli kriterlere göre önceden tanımlanmış sınıflara ayırmaktır. Bu sınıflarda verileri toplayarak tahmine dayalı modelleri keşfetmeye yarar (Chen ve Liu, 2004). Sinir ağları ve karar ağaçları örnek olarak gösterilebilir (Ngai, vd., 2009).

Regresyon, iki ya da daha çok nicel değişken arasındaki ilişkiyi ölçmek için kullanılan analiz metodudur. Sık görülen model eğilimlerini yakalamak için kullanılır. Doğrusal regresyon ve lojistik regresyon analizi örnek olarak gösterilebilir.

Veri madenciliği kullanılarak yapılan tahmin çalışmalarında hem hızlı hem de yetkin bir metot olan makine öğrenme yöntemleri sıkça kullanılmıştır. Makine öğrenmesi (yapay öğrenme) bir problemi başka bir probleme ait girdi ve çıktı verilerine göre modelleyen bilgisayar algoritmaları olarak tanımlanabilir. Yapay zekânın ana kollarından biri olan makine öğrenmesi, birçok akademik alandaki çalışmalara entegre edilmiştir. Makine öğrenmesi, bilgisayar sisteminin kendi kendine anlayabilme ve öğrenebilme kabiliyetini sağlamayı amaçlar. Bilimsel hedefi, alternatif öğrenme mekanizmalarını araştırmak ve bulmaktır. Ayrıca makine öğrenmesi sisteminden, öğrenici ile iletişim kurması ve çalışma verilerinin sisteme uygulanabilmesi beklenmektedir. Sistemin kurulması ve test edilmesiyle bu süreçlerin etkinliği ve limitleri belirlenmektedir (Carbonell ve Gil, 1987).

Bu çalışmada pazar araştırması, iş gücü araştırması, insan kaynakları alanı ve eğitim araştırmalarında kullanılabilecek bir veri olan Türkiye'deki işsizlik oranının, tahminini etkileyen değişkenlerin analizi makine öğrenme algoritmalarından faydalanılarak yapılmıştır. İşsizlik oranı tahmininde kullanılan bağımlı değişken TÜİK verilerinden elde edilen "işsizlik oranı" değeridir. Bu çalışmada, işsizlik oranı tahmininde etkin olan girdi değişkenlerini (bağımsız değişkenleri) saptamak ve etkin olan makine öğrenmesi yöntemini tespit etmek hedeflenmiştir.

2. LİTERATÜR TARAMASI

Literatür araştırması kapsamında 2015-2020 yılları arasında Türkiye'deki işsizlik oranının tahmini yapan çalışmalar incelenmiştir.

İçen ve Günay (2015), Türkiye'deki işsizlik oranı tahmini için LINGO programını kullanmışlardır. Analiz sonucunda bulanık regresyon modeli kullanılarak daha güvenilir tahminler elde edildiği gözlemlenmiştir. Çalışmada kullanılan bulanık model parametrelerinin önemi bulanık hipotez testi ile test edilmiştir. Bulanık regresyon analizi doğrusal programlama yaklaşımı ile yapılmış olup farklı h düzeyleri ile çalışılmıştır. Verilen amaç fonksiyonu bulanık katsayıların toplam yayılımlarını minimize etmek için kullanılmış ve regresyon katsayılarının merkez ve yayılım değerleri elde edilmiştir. Elde edilen model parametrelerinin önemli olup olmadığı Buckley (2004) çalışmasında önerilen bulanık hipotez testi ile test edilmiştir

Gülenay Chadwick ve Şengül (2015) Google tarama sonuçlarının verilerini kullanarak Türkiye'de üç aylık bir gecikme ile yayınlanan "aylık tarım dışı işsizlik" dönem içi tahminini geliştirip geliştirilmediği araştırılmıştır. Çalışmanın yöntemi olarak, lineer tahmin modelleri ve Bayes tabanlı model ortalaması kullanılmıştır. Veri analizinde Ocak 2005 ve Kasım 2011 tarihleri arası baz alınmıştır. Elde edilen sonuçlara göre tarım dışı işsizlik oranı dönem içi tahminlerini iyileştirmektedir. Tahmin modellerinin kare hataların ortalamalarının karekökü karşılaştırıldığında en iyi performansı gösteren modelin Google tarama sonuçlarının ana bileşenlerini içermekte olduğu tespit edilmiştir. Google tarama verilerini içeren model örneklem içinde %47,7, örneklem dışında ise %38,4 daha doğru tahminler verdiği gözlemlenmiştir. Google tarama sonuçları verilerini içermeyen en iyi model örneklem içinde %34,1, örneklem dışında ise %29,4 daha doğru tahminler verdiği gözlemlenmiştir.

Çiççi (2016) çalışmasında, Markov zincirlerini stokastik süreç temelli bir tahminleme aracı olarak kullanmıştır. Bu çalışmada, Markov zincirleri yöntemi kullanılarak Türkiye'nin 2016 yılı kurumsal olmayan işsizlik oranı tahminlemesi yapılmıştır. Veri seti olarak, TÜİK'in yayınladığı 1996 ile 2014 yıllarına ait kurumsal olmayan işsizlik verileri kullanılmıştır. Elde edilen sonuçlara göre 2016 yılında Türkiye'de işsizlik oranı %94 olasılıkla 6,5 puanın üstünde bir değer alacağı, %79 olasılıkla 9,9 puanın altında bir değer alacağı, %80 olasılıkla 10 puan altı bir değer olacağı ve %88 olasılıkla 11 puan altı bir değer olacağı öngörülmüştür. Genel anlamda 2016 yılı işsizlik oranının tek haneli bir değer alacağı tahmin edilmektedir. Bu değerlerin 6,5 ile 9,9 aralığında kalması beklenmektedir.

Yolcu ve Baş (2016) çalışmasında işsizlik oranını ve işgücüne katılım oranını tahmin etmiştir. Polonya ve Türkiye'de farklı bulanık zaman serileri yöntemleriyle işsizlik oranları analiz edilmiştir. Tüm analizler sonucunda, bulanık zaman serileri tahmin modellerinin işsizlik oranı ve işgücüne katılım oranı gibi zaman serilerinin tahmininde üstün tahmin performansına sahip olduğu ve bu tür tahminler için etkili bir tahmin aracı olarak kullanılabilirliği belirlenmiştir. Ayrıca, önerilen bulanık zaman serisi yöntemleri geliştirildiğinde daha da iyi tahmin sonuçları verebilecekleri söylenmiştir.

Tüzemen ve Yıldız (2017) Türkiye işsizlik oranını, Toplamsal ve Çarpımsal-Mevsimsel Holt-Winters yöntemleri ile tahmin edilmiştir. Çalışmada merkez bankasının veri setinden yararlanılmıştır. Veriler, 01/2005-10/2016 dönemi Türkiye işsizlik oranlarından oluşturulmuştur. Yapılan analizlerin performans değerlendirmeleri hata kareler ortalaması ve ortalama mutlak yüzde hata değerleri ile değerlendirilmiştir. Yapılan analizde, Toplamsal-Mevsimsel HW yöntemi ile $MSE = 0,073711$ ve $MAPE=0,021695$, Çarpımsal-Mevsimsel HW yöntemi ile de $MSE = 0,8618929$ ve $MAPE = 0,02217202$ değerleri saptanmıştır. Ayrıca, Toplamsal-Mevsimsel HW yöntemi ile yapılan analizde işsizlik oranlarının 2017 yılında artarak devam edeceği öngörülmüştür.

Yıldırım ve Başgeçmez (2017) bu çalışmada, tek değişkenli zaman serisi verilerini tahmin etmede yararlı olan teknikleri kullanmışlardır. Bu çalışmanın amacı, Türkiye İşsizlik Oranını tahmin eden en etkin yöntemi saptamaktır. Türkiye İşsizlik oranına ilişkin veriler, 1996-2015 tarihleri arası baz alınarak TÜİK'den alınmıştır. Öncelikle yıllık işsizlik oranlarının karakterizasyonu için otokorelasyon analizi kullanılmıştır. Otokorelasyon analizi sonuçlarına göre Türkiye işsizlik oranları

arasında korelasyon olmadığı gösterilmiştir. Ayrıca bu serilerin mevsimselliğinin olduğu da tespit edilmiştir. Bu nedenle en iyi tahmin modellerinin basit hareketli ortalama, ağırlıklı hareketli ortalama, basit üstel düzeltme, uyarlanabilir yanıt oranı üstel düzeltme, Holt Lineer Tahmin ve Holt-Winter trend ve mevsimsellik yöntemi olduğuna karar verilmiştir. Tüm modeller için tahmine en uygun tekniğin seçilmesinde kullanılan Theil'in U istatistikleri hesaplanmıştır. Bu istatistiklere göre, Türkiye işsizlik oranı için en uygun yöntem Holt'un Doğrusal Tahmin yöntemidir. Bu tekniğe göre, Türkiye işsizlik oranınının 10,5 olacağı öngörülmüştür. Elde edilen sonuçlara göre, Uyarlanabilir Tepki Oranı Üssel Düzeltme (ARRES) yönteminin basit hareketli ortalamalar için 4.016, ağırlıklı hareketli ortalamalar için 2.653, çift hareketli ortalamalar için 4.940 olan en düşük Ortalama Kareler Hatası değerini hesapladığı için Türkiye işsizlik oranı için en uygun yöntem olduğu sonucuna varılabileceği belirtilmiştir.

Gil-Alina vd. (2019) Türkiye işsizlik oranı serilerini uzun hafızalı modeller ve kısmi bütünleşik teknikler kullanarak incelemiştir. Çalışmanın amacı, literatürde var olan hipotezlerden (NAIRU, yapısalıcı, kalıcılık ve histeresis) hangi hipotezin incelenen veriler için daha uygun olduğunu belirlemek amacıyla zaman serilerinin entegrasyon sırasını tahmin etmektir. Çalışmada, 1988 yılı 3.çeyrek ile 2013 yılı 4.çeyrek arası Türkiye işsizlik oranı verileri kullanılmıştır. Sonuçlarımız, Türkiye'de işsizliğin yüksek oranda kalıcı olduğunu ve vakaların çoğunda entegrasyon sıralarının 1'e eşit veya daha yüksek olduğunu göstermektedir. Ampirik sonuçlarından önemli politika çıkarımları ortaya çıkmıştır; işgücü ve makroekonomik politikaların Türkiye'deki işsizlik oranları üzerinde uzun süreli etkilerinin olması beklenmektedir.

Bağcı (2020) Türkiye'deki işsizlik oranını tahmin etmek için GM (1.1) modeli kullanıldıktan sonra, Markov zincirleri GM (1.1) modeline entegre edilerek Gri Markov modeli oluşturulmuş ve bu yöntemde işsizlik oranı yeniden tahmin edilmiştir. Başlangıç olarak, Gri sistem modellerinden biri olan GM (1,1) modeli kullanılarak işsizlik oranlarına ilişkin istatistikler hesaplanmıştır. Daha sonra her bir gözlem değerinin hata oranları hesaplanarak belirlenmiştir. Hem hata oranları hem de verilerin orijinal durumu kullanılarak Markov analizi yapmak için durum geçişleri matrisi elde edilmiştir. Tüm gözlemler, bu matris ve durum geçişleri yardımıyla Gri Markov analizi kullanılarak yeniden tahmin edilmiştir. İki GM arasında karşılaştırma yapabilmek için (1,1) yeni tahmin değerlerine ait hata oranları da hesaplanmıştır ve Gri Markov analizinin daha doğru tahmin sonuçları ürettiği tespit edilmiştir.

Bu çalışmada, literatürden farklı olarak tahmin modelleri oluşturulurken veri madenciliğinin sınıflandırma fonksiyonundan yararlanılmıştır. TÜİK tarafından yapılan sınıflandırma yardımıyla tahmin modelleri oluşturulmuş ve işsizlik oranı tahmininde hangi özelliklere sahip girdi değişkenlerinin yani hangi veri gruplarının daha etkin olduğu araştırılmıştır. Ayrıca işsizlik oranı 3 farklı (GRNN, SVM, GEP) makine öğrenme algoritması ile tahmin edilerek yöntemlerin performansları karşılaştırılmıştır. Regresyon ve sinir ağı tabanlı bir yöntem olan GRNN yönteminin, SVM ve GEP yöntemlerinden daha etkin sonuçlar verdiği tespit edilmiştir.

3. VERİ SETİ VE YÖNTEM

3.1 Veri Seti

Bu çalışmada kullanılan veri seti TÜİK tarafından ilk defa 2016 yılında yayımlanan (2015 verilerini içermektedir) İllerde Yaşam Endeksi (İYE) çalışmasından alınmıştır. TÜİK tarafından veri seti hakkında verilen bilgiler şu şekildedir; “Veri seti, endeks objektif ve subjektif göstergeleri içermektedir. Subjektif göstergelerinin temel veri kaynağı il düzeyinde uygulanması planlanan Yaşam Memnuniyeti Araştırmasının (YMA) sonuçlarıdır. Objektif göstergeler ise il düzeyinde TÜİK verileri ve diğer kurum ya da kuruluşlardan derlenen verilerden oluşmaktadır.” (Resmi İstatistik, 2016).

YMA verileri, 4500 cevaplı örnek hanede 9397 cevaplı bireyle gerçekleştirilen anket çalışmasından elde edilmiştir. Anketin amacı TÜİK tarafından “Türkiye'deki bireylerin öznel mutluluk algısı, sağlık, sosyal güvenlik, örgün eğitim, çalışma hayatı, gelir, kişisel güvenlik ve adalet

hizmetleri, kişisel gelişim gibi temel yaşam alanlarındaki memnuniyetleri ölçmek ve bunların zaman içindeki değişimi takip etmek” şeklinde açıklanmıştır (TÜİK, 2016a).

Veri anketinin kapsamı, “Türkiye Cumhuriyeti sınırları içinde bulunan hanelerde yaşayan 18 ve daha yukarı yaştaki T.C. vatandaşları ile yabancı uyruklu kişilerdir. Araştırmada, kurumsal nüfus (üniversite öğrenci yurdu, huzurevi, bakımevi, ceza ve tutukevi, ıslahevi, yetiştirme yurdu, askeri birlik ve kışla vb.) kapsam dışı tutulmaktadır.” Anket sonuçları Türkiye’nin 81 ilini kapsamakta ve oranlar illere göre ayrılmaktadır (TÜİK, 2016b).

Bu çalışmada TÜİK tarafından yayımlanan İYE göstergelerinin; işsizlik oranı % (IO) iller (ID), Kişi başına düşen tasarruf mevduatı TL (KDT), Orta ve üstü gelir grubundaki hanelerin oranı % (OUG), Temel ihtiyaçlarını karşılayamadığını beyan eden hanelerin oranı % (TIK), Hekim başına düşen müracaat sayısı (HDM), Sağlığından memnuniyet oranı % (SMO), Kamunun sağlık hizmetlerinden memnuniyet oranı % (KSM), Okul öncesi eğitimde (3-5 yaş) net okullaşma oranı % (OOE), TEOG sistemi yerleştirmeye esas puan ortalaması (TEOG), YGS puan ortalaması (YGS), Fakülte veya yüksekokul mezunlarının oranı % (FYM), Kamunun eğitim hizmetlerinden memnuniyet oranı (%) (EHM), Cinayet oranı (CO), Ölümlü ve yaralanmalı trafik kazası sayısı (TKO), Gece yalnız yürürken kendini güvende hissedenenlerin oranı (GGO), Kamunun asayiş hizmetlerinden memnuniyet oranı (KAM), Mahalli idareler seçimlerine katılım oranı (SKO), Siyasi partilere üyelik oranı (SPU), Sendika/dernek faaliyetleri ile ilgili olanların oranı (SDU), İnternet abone sayısı (IAS), Kanalizasyon ve şebeke suyuna erişim oranı % (SEO), Havalimanına erişim oranı (HEO), Belediyenin toplu taşıma hizmetlerinden memnuniyet oranı (BHM), Sinema ve tiyatro seyirci sayısı (STS), Bin kişi başına düşen alışveriş merkezi alanı (m²) (AMO), Sosyal ilişkilerinden memnuniyet oranı (SIM) ve Sosyal hayatından memnuniyet oranı (SHM) verileri kullanılmıştır. IO’yu tahmin etmek için IO bağımlı değişken çıktı verisi olarak, diğer 27 bağımsız değişken ise girdi verisi olarak kullanılmıştır.

TÜİK tarafından hazırlanan İYE çalışmasının verileri niteliklerine göre sınıflara ayrılmıştır. Bu sınıflandırma sonucunda oluşan veri grupları ile tahmin modelleri oluşturulmuştur. Oluşturulan 8 tahmin modeli ve içerdiği değişkenler Tablo 1’de sunulmuştur. Veri setinin istatistiği Tablo 2’de sunulmuştur.

Tablo 1. Tahmin Modellerinin Veri Grupları

Model No	Model İsmi	Model Değişkenleri
1	Gelir ve Servet	ID IO KDT OUG TIK
2	Sağlık	ID HDM SMO KSM
3	Eğitim	ID OOE TEOG YGS FYM EHM
4	Güvenlik	ID CO TKO GGO KAM
5	Sivil Katılım	ID SKO SPU SDU
6	Alt Yapı Hizmetlerine Erişim	ID IAS SEO HEO BHM
7	Sosyal Yaşam	ID STS AMO SIM SHM
8	Tüm Değişkenleri İçeren Model	ID IO KDT OUG TIK HDM SMO KSM OOE TEOG YGS FYM EHM CO TKO GGO KAM SKO SPU SDU IAS SEO HEO BHM STS AMO SIM SHM

Tablo 2. Veri Setinin İstatistik Değerleri

Değişkenler	Min	Max	Ortalama	Standart Sapma
IO	4,2	23,4	8,801	3,878

ID	1	81	41	23,527
KDT	616	18131	4342,033	2821,418
OUG	16,27	58,91	34,382	8,633
TIK	32,78	74,95	50,95	10,19
HDM	2763	8067	5834,383	1245,116
SMO	59,15	80,76	71,997	4,448
KSM	54,55	89,13	77,468	7,305
OOE	23,53	53,16	35,275	6,247
TEOG	215,33	337,99	295,936	27,672
YGS	178,6	207,95	197,64	5,512
FYM	8,56	22,65	13,118	2,348
EHM	48,18	88,89	74,089	8,46
CO	4,47	69,34	25,499	11,118
TKO	0,71	4,59	2,436	0,776
GGO	45,1	87,23	67,598	9,242
KAM	58,88	94,86	84,236	7,161
SKO	77,1	93,1	88,157	3,194
SPU	12,44	34,73	21,243	3,935
SDU	3,54	22,08	6,735	2,582
IAS	2,16	17,66	8,68	3,866
SEO	31,11	100	74,37	15,827
HEO	0	9874,8	669,867	1571,486
BHM	23,46	78,81	58,394	11,581
STS	0,29	147,44	45,353	34,005
AMO	0	284,01	69,54	68,29
SIM	78,23	96,16	88,831	3,793
SHM	21,5	80,88	54,265	11,355

3.2 Yöntem

3.2.1. Genel Regresyon Sinir Ağı

Örnek sayısı arttıkça optimal regresyon yüzeyine yakınsama prensibi ile çalışan GRNN yönetimi, hızlı öğrenmeye dayalı bir algoritmadır. Seyrek veriye sahip bir set üzerinde çalışıldığında GRNN yöntemi avantajlıdır. Örnek olarak, veri hiper yüzeyi, her yeni numune noktasının eklenmesiyle artarak, karmaşıklaşır. GRNN yöntemi, verileri düşük ve yüksek olarak ikiye bölmek için, hiper uzay aralarında geçişe izin verir (Dtreg, 2021a).

GRNN ağı dört katmanlı bir yapıdan oluşmaktadır. Giriş katmanı, her tahmin değişkeni için giriş katmanında bir nöron bulunmaktadır. N (kategori sayısı), Kategorik değişkenler ile çalışırken N-1 nöronları kullanılır. “Giriş nöronları, medyanı çıkararak ve çeyrekler arası aralığa bölerek değerlerin aralığını standartlaştırır. Sonrasında bu değerleri gizli katmandaki nöronların her birine iletir” (Dtreg, 2021a). Gizli katman, eğitim veri kümesindeki her durum için bir nörona sahiptir. Nöron, hedef değerle birlikte durum için öngörücü değişkenlerin değerlerini depolamaktadır. Gizli bir nöron, girdi katmanından gelen girdi değerlerinin x vektörü ile sunulduğunda, test senaryosu, nöronun merkez noktasından Öklid mesafesini hesaplar ve ardından sigma değerlerini kullanarak RBF çekirdek işlevini uygular. Elde edilen sonuç değeri, desen katmanındaki nöronlara iletilir. Toplama katmanı, GRNN ağları için, toplama katmanında yalnızca iki nöron vardır; “Bir nöron payda toplama birimi, diğeri pay toplama birimidir. Payda toplama birimi, gizli nöronların her birinden gelen ağırlık değerlerini toplar. Pay toplama birimi, her bir gizli nöron için gerçek hedef değerle çarpılan ağırlık değerlerini toplar.” Karar katmanı, pay toplama biriminde biriken değeri payda toplama birimindeki değere böler ve sonucu tahmin edilen hedef değer olarak kullanır (Dtreg, 2021a).

3.2.2. Destek Vektör Makineleri

Temel olarak, SVM yönteminin arkasındaki ana fikir, doğrusal olarak ayrılabilir modeller için sınıflandırma için kullanılacak bir optimal hiper düzlemin oluşturulmasıdır. Optimal hiper düzlem; “hiper düzlemin marjını, yani hiper düzlemden her modelin en yakın noktasına olan mesafeyi maksimize eden kalıpları sınıflandırmak için hiper düzlemler kümesinden seçilen bir hiper düzlemi” şeklinde tanımlanmıştır. SVM yönteminin temel amacı, marjı maksimize etmektir, böylece verilen desenleri doğru bir şekilde sınıflandırabilir, yani marj boyutu büyüdükçe desenleri daha doğru sınıflandırabilir. Hiper düzlem, $aX + bY = C$ ile formülize edilir (Pradhan, 2012).

SVM tabanlı modeller üç farklı hattan oluşmaktadır. $W.x - b = 0$ çizgisi, ayırma marjı veya marjinal çizgi olarak bilinir. $W.x - b = 1$ ve $w.x - b = -1$ hatları, kenar boşluğu çizgisinin her iki tarafındaki çizgilerdir. Bu üç çizgi birlikte, verilen desenleri ayıran hiper düzlemi oluşturur ve hiper düzlemin kenarlarında yatan desene destek vektörleri denir. Kenar boşluğu çizgisi ile hiper düzlemin kenarları arasındaki dikey mesafe, kenar boşluğu olarak bilinir. Doğru sınıflandırma için SVM'nin amaçlarından biri, daha iyi sınıflandırma için bu marjı maksimize etmektir. Kenar boşluğunun değeri veya dikey mesafe ne kadar büyükse, sınıflandırma süreci o kadar iyidir ve dolayısıyla hata oluşumunu en aza indirir (Pradhan, 2012).

SVM tabanlı modeller, klasik çok katmanlı algılayıcı sinir ağlarına benzerdir. Bir çekirdek işlevi kullanan SVM yöntemi, polinom, radyal temelli işlev ve çok katmanlı algılayıcı sınıflandırıcılar gibi ağırlıklarının konveks olmayan bir denklem yerine doğrusal kısıtlamalarla ikinci dereceden bir programlama problemini çözen alternatif bir eğitim yöntemi kullanır (Dtreg, 2021b).

3.2.3 Gen İfadesi Programlama

GEP, ifade-mutasyon problemine zarif ve etkili bir çözüm olarak, 1999 yılında Cândida Ferreira tarafından önerilmiştir. Ferreira, ifadeleri kodlamak için, çok çeşitli mutasyon ve melezleme tekniklerinin hızlı uygulanmasına izin verirken ortaya çıkan ifadenin her zaman sözdizimsel olarak geçerli olacağını garanti eden bir sistem tasarlamış ve bu yaklaşıma GEP adı verilmiştir (Dtreg, 2021c).

GEP yöntemi biyolojik evrimi taklit eden bir bilgisayar algoritmasıdır. Karar ağaçları, sinir ağları ve polinom yapılar dahil olmak üzere birçok farklı model türü oluşturmak için kullanılabilir. DTREG programında uygulanan GEP türü, sembolik regresyondur. Sembolik matematiksel veya

mantıksal bir işlev oluşturduğu için bu şekilde adlandırılmaktadır. Girdi verileri fonksiyonlara ve değişkenlere ek olarak, ifadeler sabitler içerebilir (Dtreg, 2021c). DTREG tarafından GEP şu şekilde tarif edilmektedir; “Sabit uzunluktaki doğrusal kromozomlarda kodlanmış bilgisayar programlarını geliştiren tam teşekküllü bir genotip / fenotip sistemidir”

3.3 Tahmin Duyarlılık Analizi

Bir tahmin yönetimin duyarlılığı, tahmin edilen değer ile gerçek değerler arasındaki sapmaların analizi sonucu ortaya çıkmaktadır. Tahmin duyarlılık analizinde, hata tahmin ölçüleri ve bunları hesaplamaya yarayan yöntemlerden yararlanılmaktadır (Kobu, 2017). Hata değeri, çalışmadan elde edilen sonuçlar ile gerçek sonuçlar arasındaki farkın alınmasıyla hesaplanır. En düşük hataya sahip tahmini yapan yöntem, en iyi sonuçları veren yöntemdir. Bu çalışmada, tahmin sonuçlarını değerlendirmek için duyarlılık analizinin hata tahmin ölçülerinden faydalanılmıştır.

3.3.1 Kolerasyon Katsayısının Karesi (R^2)

Kolerasyon katsayısı, “iki sayısal ölçüm arasında doğrusal bir ilişki olup olmadığını, varsa bu ilişkinin yönünün ve değerinin ne olduğunu belirlemek için kullanılan bir istatistiksel yöntem” olarak tanımlanmaktadır. Korelasyon katsayısı negatif ise iki değişken arasında ters ilişki vardır, pozitif ise değişkenlerden biri artarken diğeri de artmaktadır R^2 Formülü (1)'de gösterilmiştir (Kobu, 2017).

$$R^2 = 1 - \frac{\sum(Y - Y')^2}{\sum(Y - \bar{Y})^2} \quad (1)$$

Bulgularda R^2 değeri yorumlanırken; %80 < R^2 çok iyi, %60 < R^2 < %80 iyi, %40 < R^2 < %60 orta, %20 < R^2 < %40 kabul edilebilir, R^2 < %20 hatalı olarak değerlendirilmiştir.

3.3.2. Hata kareler ortalamasının karekökü (Root Mean Square Error - RMSE)

RMSE, kareler toplamının bölünmüş tahmin sayısının kare kökü olarak tanımlanır. Gerçekte gözlenen değerler ile tahmin modeli tarafından öngörülen değerler arasındaki farkı ölçmektedir. Dolayısıyla düşük RMSE değeri, modelin daha doğru tahmin ürettiğini göstermektedir. RMSE Formülü (2)'de gösterilmiştir (Kobu, 2017).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y - Y')^2} \quad (2)$$

Bulgularda RMSE değeri yorumlanırken düşük RMSE değerine sahip (hata değeri daha düşük olan) modelin diğer modelden daha doğru (daha iyi) tahmin gerçekleştirdiği düşünülmüştür.

Denklemler 1'de, Y gerçek IO, Y' tahmin edilen IO, \bar{Y} gerçek IO değerlerinin ortalamasını, n ise test edilen alt küme örneklerinin sayısını göstermektedir.

3.3.3. Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error - MAPE)

MAPE, genellikle hata değerlerinin birim değerleri farklılık gösteriyorsa tercih edilmektedir. MAPE ile farklı birim değerlerin karşılaştırılmasında ortaya çıkabilecek dezavantajlar elimine edilmektedir. MAPE Formülü (3)'de gösterilmiştir (Kobu, 2017).

$$MAPE = 100 * \frac{\sum_{i=1}^n \frac{|Y-Y'|}{Y}}{n} \quad (3)$$

Denklemlerde Y gerçek IO, Y' tahmin edilen IO, \bar{Y} gerçek IO değerlerinin ortalamasını, n ise test edilen alt küme örneklerinin sayısını göstermektedir.

Bulgularda MAPE değeri yorumlanırken; $MAPE < \%10$ çok iyi, $\%10 < MAPE < \%20$ iyi, $\%20 < MAPE < \%50$ kabul edilebilir, $\%50 < MAPE$ hatalı olarak değerlendirilmiştir.

Bu çalışmada kullanılan makine öğrenme yöntemleri DTREG programı yardımıyla çalıştırılmıştır. GRNN, SVM ve GEP algoritmaları 10 katlı çapraz doğrulama yapılarak tahmin modellerine uygulanmıştır. Tablo 3’de makine öğrenme yöntemlerinde kullanılan parametreler ve değerleri verilmiştir. Bu çalışmada veri setinin hem girdi hem çıktı değerleri ağa tanıtılmış önce eğitim sonra test sonuçları hesaplatılmıştır. Elde edilen sonuçlar Tablo 4’de gösterilmiştir. Sonuçlar incelendiğinde sınıflandırma sonuçlarına göre düzenlenmiş 8 farklı tahmin modeli görülmektedir. Bu tahmin modellerinin GRNN, SVM ve GEP yöntemlerinden elde edilen RMSE değerleri bulunmaktadır.

Şekil 1’de ise GRNN tabanlı modeller ile elde edilen sonuçların SVM ve GEP tabanlı modeller ile elde edilen sonuçlara göre RMSE değerlerinin yüzde iyileştirme oranlarının grafiği verilmiştir.

Tablo 5’de GRNN yöntemine göre veri gruplarının RMSE oranları sıralanmıştır.

Tablo 6’de ise GRNN yöntemine göre veri gruplarının etkinliği sunulmuştur. Analiz yapılırken RMSE oranının yanı sıra R^2 , MAPE ve programın çalışma süresi hesaplanmıştır.

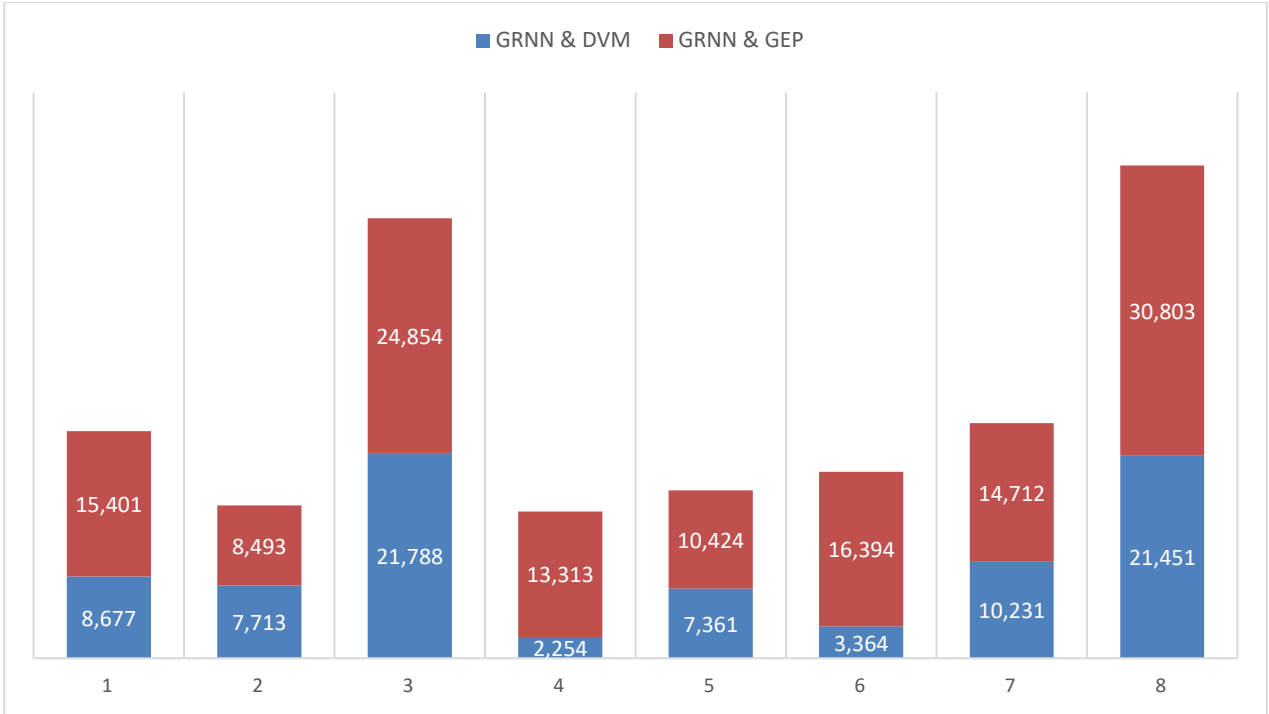
Tablo 3. Yöntemlerin Parametreleri

Yöntemler	Parametreler	Değerler
GRNN	Sigma	[0,0001 - 10]
	Arama Sayısı	20
	Çekirdek Fonksiyonu	Gaussian
SVM	Cost (C)	[0,1 - 5000]
	Gamma (γ)	[0.001 - 50]
	Çekirdek Fonksiyonu	RBF
GEP	Populasyon Değeri	[10-60]
	En Yüksek Gen Sayısı	2000
	Kromozom Başına Gen Sayısı	4

Tablo 4. Yöntemlerin Karşılaştırmalı Sonuçları

No	Veri Grubu	GRNN		DVM		GEP	
		Eğitim	Test	Eğitim	Test	Eğitim	Test
		RMSE	RMSE	RMSE	RMSE	RMSE	RMSE

1	Gelir ve Servet	2,927	3,143	2,770	3,442	3,178	3,715
2	Sağlık	1,423	3,438	3,668	3,725	3,436	3,757
3	Eğitim	1,485	2,614	2,401	3,342	3,482	3,478
4	Güvenlik	2,397	3,123	3,122	3,195	3,215	3,602
5	Sivil Katılım	2,634	3,315	2,954	3,578	3,554	3,700
6	Alt Yapı Hizmetlerine Erişim	1,186	2,831	1,740	2,929	2,958	3,386
7	Sosyal Yaşam	1,303	2,845	2,550	3,169	3,019	3,336
8	Tüm Değişkenleri İçeren Model	0,257	2,112	1,832	2,688	2,340	3,052



Şekil 1. Yöntemlerin Karşılaştırmalı Sonuçları

Tablo 5. GRNN Yöntemine Göre Veri Gruplarının Etkinlik Sıralaması

Model No	Veri Grubu	Değişkenler	Eğitim RMSE	Test RMSE
8	Tüm Değişkenleri İçeren Model	ID IO KDT OUG TIK HDM SMO KSM OOE TEOG YGS FYM EHM CO TKO GGO KAM SKO SPU SDU IAS SEO HEO BHM STS AMO SIM SHM	0,257	2,112
3	Eğitim	ID OOE TEOG YGS FYM EHM	1,485	2,614

6	Alt Yapı Hizmetlerine Erişim	ID IAS SEO HEO BHM	1,186	2,831
7	Sosyal Yaşam	ID STS AMO SIM SHM	1,303	2,845
4	Güvenlik	ID CO TKO GGO KAM	2,397	3,123
1	Gelir ve Servet	ID IO KDT OUG TIK	2,927	3,143
5	Sivil Katılım	ID SKO SPU SDU	2,634	3,315
2	Sağlık	ID HDM SMO KSM	1,423	3,438

Tablo 6. GRNN Tabanlı Modellerin Veri Grupları Bazında Detaylı Analizi

GRNN Yöntemi	Eğitim			Test			
	R ²	RMSE	MAPE	R ²	RMSE	MAPE	ZAMAN
Gelir ve Servet	0,4234	2,927	23,841	0,335	3,143	24,801	00:01:36
Sağlık	0,86364	1,423	9,344	0,2044	3,438	24,671	00:01:07
Eğitim	0,85151	1,485	11,698	0,5401	2,614	20,126	00:01:53
Güvenlik	0,61324	2,397	18,459	0,34356	3,123	22,241	00:01:37
Sivil Katılım	0,5331	2,634	22,007	0,26038	3,315	28,211	00:01:27
Alt Yapı Hizmetlerine Erişim	0,90532	1,186	9,044	0,46061	2,831	18,848	00:01:09
Sosyal Yaşam	0,88568	1,303	10,142	0,4551	2,845	22,548	00:01:13
Tüm Değişkenleri İçeren Model	0,99555	0,257	15,281	0,69984	2,112	16,932	00:04:16
Ortalama Sonuçlar	0,75893	1,701471	14,97687	0,412374	2,927368	22,2971	00:01:47

4. BULGULAR

Tablo 4, 5, 6 ve Şekil 1 ile elde edilen bulgular şu şekildedir;

- Tablo 4 incelendiğinde, IO tahmininde tüm modeller için GRNN yönteminin en iyi (en düşük RMSE hata değerli) sonuçları verdiği ve SVM yönteminin ise GEP yönteminden daha iyi sonuçlar verdiği gözlemlenmiştir.
- Tablo 4 incelendiğinde, GRNN tabanlı modellerin ortalama RMSE değerleri SVM tabanlı modellere göre %10,355 & GEP tabanlı modellere göre %16,800 daha düşüktür.
- Tablo 4 incelendiğinde, Tüm yöntemler için en iyi tahmin modelinin tüm değişkenleri içeren 8.model olduğu gözlemlenmektedir

- Tablo 4 incelendiğinde, En iyi model olan 8.model (en düşük hata değerleri sahip) ile en fazla hata değerine sahip 2.model sağlık veri grubu arasındaki RMSE değerlerinin farkı (iyileşme oranı) 1,022'dir.
- Tablo 5 incelendiğinde GRNN yöntemine göre en iyi tahmin modelinin en düşük test RMSE değerine sahip olan 8 model olduğu gözlemlenmiştir.
- Tablo 5 incelendiğinde GRNN yöntemine göre en iyi tahmin model sıralamasının 8,3,6,7,4,1,5,2 olduğu gözlemlenmektedir. Buna göre IO tahmininde veri gruplarının etkinliği şu şekilde sıralanabilir; öncelikle eğitim, alt yapı hizmetleri, sosyal yaşam, güvenlik, gelir ve servet, sivil katılım ve son olarak sağlık veri grubu.
- Şekil 1'den elde edilen verilere göre en fazla 8.model ve 3.modelde en az 4.model ve 2.modelde iyileşme gözlemlenmiştir.
- Tablo 6 incelendiğinde, R^2 değerinin ortalama %41 olarak görülmektedir. Bu oran orta düzeyde uyumluluk olarak yorumlanmıştır. En iyi sonucu veren tüm değişkenleri içeren modelin %69 R^2 değeri ile iyi düzeyde uyumluluğa sahip olduğu söylenebilir.
- Tablo 6 incelendiğinde ortalama MAPE değerinin %22 olması iyiye yakın kabul edilebilir hata oranının olduğunu. Tüm değişkenleri içeren modelin ise %16 hata değerine sahip olduğu yani iyi bir tahmin modeli olduğu ortaya konulmuştur.
- Tablo 6 incelendiğinde tahmin modellerinin GRNN yönteminde çalışma zamanlarının ortalama 1 dakika 47 saniye olduğu görülmektedir. Bu durum çok kısa sürede kabul edilebilir oranda sonuçlar elde edildiğini göstermektedir.

5. SONUÇ

Veri madenciliği, veri tabanında bulunan verilerin bilgiye dönüştürülmesi kullanılmasının yanı sıra, tahmin için uygun verilerin tespitinde de kullanılmaktadır. Bu çalışmada veri madenciliği yöntemlerinin sınıflandırma (TÜİK'in özelliklerine göre veri grupları), ilişkilendirme ve regresyon (makine öğrenme algoritmaları) fonksiyonlarından yararlanılmıştır.

Bu çalışmada, TÜİK'in İYE araştırmasının veri seti ve veri grupları kullanılarak 8 farklı tahmin modelleri geliştirilmiştir. 8 tahmin modeli GRNN, DVM ve GEP makine öğrenme yöntemleriyle analiz edilmiştir. Elde edilen sonuçların duyarlılığı RMSE oranı kullanılarak karşılaştırılmıştır. Tüm modeller için kullanılan yöntemden bağımsız olarak en iyi sonuçlar tüm değişkenleri içeren 8.modelden elde edilmiştir ve en düşük performans ise sağlık verilerini içeren 2.modelden elde edilmiştir. 8 tahmin modelinde de tutarlı olarak en iyi sonuçlar GRNN yönteminden elde edilmiştir. Ayrıca SVM yöntemi ile GEP yönteminden daha iyi sonuçlar elde edilmiştir. GRNN yönteminin performansı incelendiğinde tüm tahmin modellerinde orta derece kolerasyon ve kabul edilebilir MAPE doğruluk oranı olduğu gözlemlenmiştir. GRNN yöntemi ile 8.modelden elde edilen sonuçlarda ise iyi derecede kolerasyon ve doğruluk oranı gözlemlenmiştir.

Gelecek çalışmalarda, IO tahmininde kullanılan yapay öğrenme yöntemleri farklılaştırılabilir. Derin öğrenme yöntemleri ile sonuçlar analiz edilebilir. Ayrıca girdi değişkenlerinin sayısı artırılarak hata oranları düşürülebilir.

KAYNAKÇA

Bağcı, B. (2020) "Gri Markov Modeli ile Türkiye'de İşsizlik Oranı Tahmini", Sosyal Güvenlik Dergisi, 10 (2) ss. 259-272

Buckley J.J, (2004) Fuzzy Statistics, Springer, Germany. ISBN 3-540-21084-9

- Carbonell, G.J., Gil, Y. (1987, Haziran) “Learning by Experimentation” Proceedings of the Fourth International Workshop on Machine Learning, 256-266.
- Chen S.Y., Liu X. (2004) “The contribution of data mining to information science” Journal of Information Science, vol 30, no 6. pp. 550–558
- Çiftçi, C. (2016) “Forecasting of Unemployment Rate For Turkey: Markov Chains Approach”, Eurasian Academy of Sciences Eurasian Business & Economics Journal, Vol. S2, 657-665, <http://dx.doi.org/10.17740/eas.econ.2016-MSEMP-140>
- Dener, M., Dörterler M., Orman A. (2009, Şubat)“Açık Kaynak Kodlu Veri Madenciliği Programları: WEKA’da Örnek Uygulama” Akademik Bilişim’09- XI. Akademik Bilişim Konferansı Bildirileri, Harran Üniversitesi, Şanlıurfa, ss. 787-796.
- DTREG, Gene Expression Programming [Çevrimiçi]. <https://www.dtreg.com/methodology/view/gene-expression-programming> Erişim: 2021a
- DTREG, General Regression Neural Networks [Çevrimiçi]. <https://www.dtreg.com/solution/probabilistic-and-general-regression-neural-networks> Erişim: 2021b
- DTREG, Support Vector Machines, [Çevrimiçi]. <https://www.dtreg.com/solution/support-vector-machines> Erişim: 2021c
- Gil-Alina L.A., Ozdemir Z.A., Tansel A. (2019) “Long Memory in Turkish Unemployment Rates”, Emerging Markets Finance and Trade (35) 201-217 (Published online: 27 Jul 2018), <https://doi.org/10.1080/1540496X.2018.1425837>
- Gong, L., Lei, H., & Zhu, L. (2012, May). The application of association rules of data mining in book-lending service. In 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (pp. 761-764). IEEE.
- Günay S., İçen D. (2015) “Türkiye’deki İşsizlik Oranının Bulanık Doğrusal Regresyon Analiziyle Tahmini”, İstatistikçiler Dergisi: İstatistik & Aktüerya, IDIA 8, 10-26
- Kobu, B. (2017) Talep Tahminlerinde Duyarlılık, Üretim Yönetimi, 18.baskı, Beta Basım Yayım Dağıtım, İstanbul, Türkiye, 120-125
- M. Gülenay Chadwick and G. Şengül, (2015) “Nowcasting The Unemployment Rate in Turkey: Let’s Ask Google”, Central Bank Review (15) 15-40, ISSN 1303-0701
- Ngai, E.W.T., Xiu, L., Chau, D.C.K. (2009) “Application of data mining techniques in customer relationship management: A literature review and classification” Expert Systems with Applications, 36, (2). 2592–2602
- Pradhan A., (2012) “Support Vector Machine-A Survey” International Journal of Emerging Technology and Advanced Engineering, 2 (8), 82-85
- Resmi İstatistik, İllerde Yaşam Endeksi, [Çevrimiçi]. Erişim: <http://www.resmiistatistik.gov.tr/detail/subject/illerde-yasam-endeksi/>
- Türkiye İstatistik Kurumu, TÜİK (2016a), Yaşam Memnuniyeti Araştırması amaç, [Çevrimiçi]. Erişim:]https://tuikweb.tuik.gov.tr/MicroVeri/YMA_2016/metaveri/amac/index.html
- Türkiye İstatistik Kurumu, TÜİK (2016b) Yaşam Memnuniyeti Araştırması kapsam [Çevrimiçi]. Erişim: https://tuikweb.tuik.gov.tr/MicroVeri/YMA_2016/metaveri/kapsam/index.html
- Tüzemen A., Yıldız Ç. (2018) “Holt Winters Tahminleme Yöntemlerinin Karşılaştırmalı Analizi: Türkiye İşsizlik Oranları Uygulaması” Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 32, (1), 1-18

- Yıldırım H., Başeğmez H. (2017) “Analysis and forecast of Turkey unemployment rate”, Global Journal of Mathematical Analysis, vol 1, no 1. 11-15,. doi: 10.14419/gjma.v5i1.6841
- Yolcu U., Eren B. (2016) “The Forecasting Of Labour Force Participation And The Unemployment Rate In Poland And Turkey Using Fuzzy Time Series Methods Comparative” Economic Research, 19, (2), doi: 10.1515/er-2016-0010