

Ekonomik Verilerden Makine Öğrenmesi Yöntemiyle Suç Oranı Tahminlemesi: Analitik Araştırma

Machine Learning-Based Prediction of Crime Rates from Economic Data: Analytical Research

 Abdullah GENÇAY^a,  Recep ERYİĞİT^b

^aT.C. İçişleri Bakanlığı Emniyet Genel Müdürlüğü, Bilgi Teknolojileri ve Haberleşme Daire Başkanlığı, Ankara, Türkiye

^bAnkara Üniversitesi Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Ankara, Türkiye

ÖZET Amaç: Suçla ilişkili göstergelerin tespit edilmesi, suçun nedenlerinin ortaya çıkarılması üzerine yapılan kriminolojik analizler suçla mücadele sürecinin önemli destek noktalarıdır. Suçun gelecekte hangi düzeyde gerçekleşeceği, hangi suç tipinin daha yoğun meydana geleceği üzerine birçok araştırma yapılmaktadır. Bu çalışmada ekonomik göstergelerle suç arasındaki ilişkinin yapay zekanın alt dallarından makine öğrenmesi yöntemleri ile ortaya konulması amaçlanmıştır. **Gereç ve Yöntemler:** Çalışmada; işsizlik, gayrisafi yurt içi hasıla, nüfus, 15 ayrı suç tipine ait hükümlü sayıları kullanılarak K-En Yakın Komşu, Rastgele Orman, Naive Bayes isimli makine öğrenmesi yöntemleri kullanılarak suç oranları tahmin edilmiştir. Ayrıca veri bölgesel açıdan Türkiye geneli ve İstatistik Bölge Birimleri Sınıflandırması-2 (26 Bölge) coğrafi bölümlere seviyelerini içermektedir. **Bulgular:** Çalışmada suç oranı tahminiyle ilgili farklı eğitim ve test oranlarına ilişkin sonuçlar, aynı yıl, 1 yıl sonraki ve 2 yıl sonraki suç oranına ilişkin sonuçlar ve ayrıca her suç türünün hangi yöntemle daha başarılı tahmin edildiğine ilişkin sonuçlar olmak üzere birçok bulgu edinilmiştir. Hırsızlık, Sahtecilik, Yaralama ve Uyuşturucu İmal ve Ticareti suçlarına ilişkin suç oranlarının tüm yöntemlerle başarılı bir şekilde tahmin edilebildiği diğer suçlarda ise başarılı olan yöntemin değiştiği hesaplanmıştır. **Sonuç:** Sosyal bilimlerde yapılan çalışmalarla ortaya konulmuş olan ekonomik göstergelerle suç arasındaki ilişkinin makine öğrenmesi algoritmalarının katkısı da eklenerek ileri boyutlara taşınabilmesi suçla etkin mücadeleye önemli katkılar sağlayacaktır. Türkiye'deki bölgelerin ayrı ayrı suç oranının ilerleyen yıllar için doğru tahmin edilebilmesi; suçla ilgili öncesinde önlem alınabilmesi, kolluk birimlerinin karar mekanizmalarını güçlendirmesi, kamu kaynaklarının verimli kullanılabilmesi gibi birçok açıdan güvenlik hizmetlerini destekleyecektir.

ABSTRACT Objective: Criminological analyses that identify crime-related indicators and reveal the root causes of criminal activity are crucial instruments in combating against crime. Numerous studies have been conducted regarding the future crime rate and the kind of crimes that will occur more frequently. In this study, it is aimed to reveal the relationship between economic indicators and crime using machine learning methods, one of the sub-branches of artificial intelligence. **Material and Methods:** In this study, crime rates were predicted with three machine learning methods [K-Nearest Neighbor (KNN), Random Forest (RF), Naive Bayes (NB)] by using unemployment, gross domestic product, population, number of convicted criminals of 15 different crime types data. Data is available both at the Turkey country level and at the Nomenclature of Territorial Units for Statistics-2 (NUTS-2) geographical grouping level (26 distinct regions). **Results:** In the study, many findings were obtained on crime rate prediction, including results for different training and test rates, results for crime rates in the same year, 1 year later and 2 years later, as well as results on which method predicts each type of crime more successfully. As a result of the prediction study; theft, counterfeiting, injury, production and trade of narcotics and psychotropic substances crime rates predicted more successfully with all three methods. **Conclusion:** Which has been demonstrated by studies in the field of social sciences, the ability to carry the relationship between economic indicators and crime, to advanced dimensions by adding the contribution of machine learning algorithms will make significant contributions to the effective fight against crime. The ability to accurately predict the crime rates of individual regions in Turkey for the coming years will support security services in many ways, such as taking precautions against crime in advance, strengthening the decision-making mechanisms of law enforcement units, and using public resources efficiently.

Anahtar Kelimeler: Suç tahminleme; kriminolojik analiz; makine öğrenmesi; sınıflandırma

Keywords: Crime prediction; criminological analysis; machine learning; classification

KAYNAK GÖSTERMEK İÇİN:

Gençay A, Eryiğit R. Ekonomik verilerden makine öğrenmesi yöntemiyle suç oranı tahminlemesi: Analitik araştırma. Türkiye Klinikleri J Foren Sci Leg Med. 2024;21(1):21-35.

Correspondence: Abdullah GENÇAY

T.C. İçişleri Bakanlığı Emniyet Genel Müdürlüğü, Bilgi Teknolojileri ve Haberleşme Daire Başkanlığı, Ankara, Türkiye

E-mail: abdullahgençay@gmail.com



Peer review under responsibility of Türkiye Klinikleri Journal of Forensic Medicine and Forensic Sciences.

Received: 23 Nov 2023

Received in revised form: 20 Feb 2024

Accepted: 26 Feb 2024

Available online: 29 Feb 2024

2619-9459 / Copyright © 2024 by Türkiye Klinikleri. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Suç kavramı, insanlık tarihinde var olan ve zaman içinde daha fazla ilgi çeken bir konu olmuştur. Sıklıkla gözlemlenmesi, suçun nedenlerini anlama ve araştırma ihtiyacını doğurmuştur. Suç, toplumun kabul ettiği ahlaki normlara, törelere veya yasalara aykırı davranışları ifade eder. Suç sadece hukuki bir konu değil, aynı zamanda psikolojik, sosyolojik ve ekonomik faktörlerin bir kombinasyonudur. Bu nedenle, suçla mücadelede yalnızca hukuki önlemler değil, aynı zamanda sosyal bilimlerin farklı alanlarından gelen bilgilerin birleştirilmesi önemlidir.^{1,2} Rutin Aktivite Teorisi suçun nedenini açıklarken kişisel özelliklerin suç işleme sebebi olduğunu, Rasyonel Teori ise suçun çevresel etkilere karşı verilen tepkilerden kaynaklandığını ileri sürmektedir.³ Meşru faaliyetlerden elde edilecek gelirle, gayrimeşru faaliyetlerden elde edilecek gelir kıyaslanarak fayda maliyet analizi ile suça yönelme olduğunu gösteren araştırmalar vardır.⁴

Suçla bulaşma nedenlerinin belirlenmesi ve bu nedenlerin suçla ilişkisinin ortaya konulması suçla mücadeleye önemli katkı sağlayabilir. Suçun gerekçeleri, suça etki eden faktörlerin belirlenmesi suç tiplerinin bölgesel ve zamansal olarak tahmin edilmesinde önemlidir.⁵

ABD, Kanada ve Türkiye'deki suçla ilgili veriler üzerinde yapılan incelemede; hayata karşı suçlar, vücut dokunulmazlığına karşı suçlar, cinsel dokunulmazlığa karşı suçlar, hürriyete karşı suçlar, mal varlığına karşı suçlar, kamunun sağlığına karşı suçların diğer suçlara kıyasla daha fazla olduğu sonucuna varılmıştır. Bu durumun bir sonucu olarak, kolluk kuvvetlerinin bu tipteki suçların önlenmesi ve sonuçlarının aydınlatılması için artırılması önerilmiştir.⁶⁻⁸

İşsizlik, enflasyon, gelir durumu gibi ekonomik göstergelerin suçla ilişkisini ortaya koyan çalışmalar dikkate alındığında belirtilen ekonomik göstergelerdeki iyileşmenin suçun ortaya çıkmasını engelleyebilecek önemli bir getiri sağlayacağını göstermektedir.^{5,9} Bir bölgede doğruluğu kanıtlanmış bir suç tahmini yapısının başka bölgelere uygulanabilir olması, benzer başarı oranlarının diğer bölgelerde de sağlanabilmesi yapıyı güçlü kılacak önemli bir husustur.^{10,11} Güçlü bir suç tahminleme altyapısının

oluşturulması sonrasında icra edilecek kolluk devriye faaliyetleri doğru bölgelere kanalize edebilecek ve böylece önlenmiş olan suçla yargı makamlarının iş gücü ortadan kalkmış olacaktır.^{12,13}

Sosyal ve fen bilimleri alanlarında suç tahminlemesi için istatistik ve yapay zekâ yöntemleri kullanılarak çok sayıda çalışma yapılmıştır. Suç tahmini yapılan çalışmalar **Tablo 1** ve **Tablo 2'**de özetlenmiştir. Bu çalışmada, benzer çalışmalara kıyasla aynı yıldaki meydana gelen suçun tahmin edilmesine ilave olarak 1 yıl sonrası ve 2 yıl sonrasında meydana gelecek olan suçlar tahmin edilmeye çalışılmıştır. Ayrıca yine benzer çalışmalara kıyasla daha küçük suç gruplarının tahmin edilmesi ya da tüm suçları dâhil edecek şekilde tek bir verinin tahmin edilmesine kıyasla 15 ayrı suçun her biri için ayrı ayrı toplamda 27 farklı bölge türüne göre tahmin yapılmıştır.

GEREÇ VE YÖNTEMLER

Çalışmada Türkiye İstatistik Kurumunun (TÜİK) yıl bazında raporladığı suçun işlendiği yere göre ceza infaz kurumuna giren hükümlü sayısı, nüfus, işsizlik oranı, kişi başı gayrisafi yurt içi hasıla (GSYİH), Gini, 15 yaş üstü okuma yazma bilmeyen nüfus verileri kullanılmıştır. Çalışma bölgesi olarak Türkiye geneli ve Türkiye'deki 26 bölge olarak sınıflandırılmış alanlara ait veriler kullanılmıştır. TÜİK bölge sınıflandırmasına ait şehir bilgileri **Tablo 3'**te verilmiştir.

Veri ön işleme kapsamında hükümlü sayısı, 100.000 kişideki hükümlü sayısına çevrilerek bölgeler arasındaki nüfus farkından kaynaklanan farklar ortadan kaldırılmıştır. Kullanılan verilerin bir kısmının ortalama değerleri **Tablo 4'**te özetlenmiştir. Tabloda cinayet hükümlüsü ortalama değeri örnek olarak verilmiş olup tüm suçları içerecek şekilde verilerin tamamı çalışma ekinde ayrı bir belge olarak sunulmuştur.

Tahmin çalışmasından önce suçlu sayısına ilişkin veri 0-1 arasında ölçeklenmiştir. Sonrasında suçun yoğunluk açısından grubunun tahmin edilebilmesi için 3 ayrı gruba sınıflama yapılmıştır. Normal suç (0-0,25), Fazla suç (0,26-0,40), Çok fazla suç (0,41-1,00) olacak şekilde yapılan bu sınıflama işlemi sonrasında suçlu sayısının sınıfı tahmin edilmeye çalışılmıştır.

TABLO 1: Sosyal bilimler alanında yapılan çalışmalar.

Çalışma yılı/ SN	Kaynakça	Bulgu	Bölge, Dönem	Veri
1	2022/14	Kaçakçılık suçu ile bazı illerimizde gelir düzeyi ve işsizlik arasında anlamlı bir ilişki olduğu belirtilmiştir.	Türkiye, 2004	Suç, İşsizlik, GSYİH
2	2022/15	İşsizlikle suç düzeyi arasında pozitif bir ilişki olduğu tespit edilmiştir.	Türkiye, 2004-2018	İşsizlik, Suç
3	2018/16	Suç ile işsizlik arasında pozitif bir ilişkinin var olduğu belirtilmiştir.	ABD, 1978-2013	Suç, İşsizlik, Savunma harcamaları
4	2022/17	Uzun vadede suç, işsizlik ve büyümenin birbirlerini etkilediği belirtilmiştir.	Türkiye, 1980-2018	Büyüme, İşsizlik, Suç
5	2021/18	İşsizlik artışlarının suç oranlarında artışa neden olduğu tespit edilmiştir.	Fransa, 1975-2013, İngiltere, 1983-2018	Suç, İşsizlik, GSYİH, Polis oranı, Nüfus artış oranı
6	2017/19	Suç ile gelir düzeyi arasında negatif bir ilişki olduğu sonucuna ulaşılmıştır.	İsveç, 1998-2010	Suç, İşsizlik
7	2015/20	Suç ile işsizlik ilişkisinin uzun vadede anlamlı olduğu, GSYİH ile suç arasındaki ilişkinin ise kısa vadede anlamlı olduğu belirtilmiştir.	Kanada, 1981-2009	Suç, İşsizlik, GSYİH
8	2014/21	İşsizlik oranında meydana gelen değişimlere suçun pozitif yönde tepki verdiği tespit edilmiştir.	15 AB ülkesi, Norveç, Kanada ve ABD 1970-2020	İşsizlik, Suç, GSYİH
9	2014/22	Suçla işsizlik arasında pozitif bir ilişki olduğu tespit edilmiştir.	Almanya, 2005-2009	İşsizlik, Mal varlığına karşı işlenen suçlar
10	2014/23	Tüm suç türleriyle işsizlik arasında pozitif bir ilişki olduğu tespit edilmiştir.	İtalya, 2000-2005	Genç işsizlik, Suç
11	2013/24	Mal varlığına karşı suçlarla işsizlik arasında pozitif bir anlamlı ilişki olduğu tespit edilmiştir.	Finlandiya, 2001-2006	İşsizlik, Suç
12	2013/25	Mal varlığına karşı suçların genç işsizliği temel belirleyici olduğu tespit edilmiştir.	Almanya, 2009-2010	İşsizlik, Suç

GSYİH: Gayrisafi yurt içi hasıla.

TABLO 2: Yapay zekâ yöntemleri kullanılarak suç tahmini yapılan çalışmalar.

SN	Veri	Yöntem(ler)	Bölge, Dönem, Veri	Çalışma yılı/Kaynakça
1	Tüm suçlar, Lokasyon, Zaman, Acil çağrı (Montgomery)	LSTM, ARIMA	Baltimore 2011, Montgomery 2015	2022/26
2	Evden hırsızlık, Gasp, Otdan hırsızlık, Oto hırsızlığı, Zaman, Lokasyon	Genetic-Fuzzy	Tahran 2014-2016	2021/27
3	Suç endeksi, İşsizlik, Nüfus, Kişi başı gayrisafi yurt içi hasıla, Göç oranı, Mülteciler oranı, Gelişmişlik sıralaması, Enflasyon oranı	Lojistik regresyon, Artificial Neural Network, MOORA	106 ülke 2018	2020/28
4	Suç (mala karşı suçlar), Lokasyon, Zaman	LSTM, KNN , Random Forest , SVM, Naive Bayes , CNN	Çin (Güney Doğu) 2015-2018	2020/10
5	Suç, Foursquare ilgi alanı, Foursquare Check-in	Lineer regresyon	Şikago 2018 Los Angeles 2018 New York 2018	2019/29
6	Tüm suçlar, Lokasyon, Zaman, Hava durumu	ST-ResNet, ARIMA, KNN , Historical Average	Los Angeles 2015 (6 ay)	2019/9
7	Tüm suç tipleri, Lokasyon, Zaman	CCRBoost-ST-ResNet, Decision Tree, Naive Bayes , Random Forest , SVM, KNN , MLP, Logic Boost	Philadelphia 2006-2017, Seattle 1996-2016, Minneapolis 2010-2016, DC Metro 2008-2017, San Francisco 2003-2015	2021/30
8	35 farklı suç, Lokasyon, Zaman	Lojistik regresyon, Decision Tree, Random Forest , MLP, Naive Bayes , SVM, XGBoost, KNN , LSTM, ARIMA	Şikago 2001-2021 ve Los Angeles 2010-2018,	2021/11
9	Tüm suç tipleri	Historical Average (HA), KNN , ARIMA, DNN, LSTM	Los Angeles, Chicago 2014-2015	2018/31
10	Nüfus, Şiddet suçları, Demografik veriler, Polis sayısı	Naive Bayes , Multilayer Perceptron, Super Vector Machine, Logistic, Multi Class Classifier, Random Forest , Decision Stump, Random Tree	ABD 1995 (Şiddet suçları), 1990 (Nüfus, Sosyoekonomik veriler ve Kolluk kuvveti (Polis verileri)	2017/32

TABLO 3: İstatistiki Bölge Birimleri Sınıflandırması-2 bölge kod ve isimleri.

TR10 (İstanbul)	TR21 (Tekirdağ, Edirne, Kırklareli)
TR22 (Balıkesir, Çanakkale)	TR31 (İzmir)
TR32 (Aydın, Denizli, Muğla)	TR33 (Manisa, Afyon, Kütahya, Uşak)
TR41 (Bursa, Eskişehir, Bilecik)	TR42 (Kocaeli, Sakarya, Düzce, Bolu, Yalova)
TR51 (Ankara)	TR52 (Konya, Karaman)
TR61 (Antalya, Isparta, Burdur)	TR62 (Adana, Mersin)
TR63 (Hatay, Kahramanmaraş, Osmaniye)	TR71 (Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir)
TR72 (Kayseri, Sivas, Yozgat)	TR81 (Zonguldak, Karabük, Bartın)
TR82 (Kastamonu, Çankırı, Sinop)	TR83 (Samsun, Tokat, Çorum, Amasya)
TR90 (Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane)	TRA1 (Erzurum, Erzincan, Bayburt)
TRA2 (Ağrı, Kars, Iğdır, Ardahan)	TRB1 (Malatya, Elazığ, Bingöl, Tunceli)
TRB2 (Van, Muş, Bitlis, Hakkâri)	TRC1 (Gaziantep, Adıyaman, Kilis)
TRC2 (Şanlıurfa, Diyarbakır)	TRC3 (Mardin, Batman, Şırnak, Siirt)

TABLO 4: Veri setindeki bazı göstergelere ilişkin ortalama değerler (2004-2020 yılları arası).

Bölge kodu (İller)	Nüfus ortalama değer	İşsizlik oranı ortalama değer	Gayrisafi yurt içi hasıla ortalama değer	Cinayetten hükümlü sayısı 100.000 nüfusa oranı ortalama değer
TR Genel	76.726.490	11,06	9.969,06	7,06
TR10 (İstanbul)	14.007.274	12,59	16.575,61	5,67
TR21 (Tekirdağ, Edirne, Kırklareli)	1.638.594	8,73	11.806,06	7,52
TR22 (Balıkesir, Çanakkale)	1.680.802	6,63	9.053,89	9,89
TR31 (İzmir)	4.057.158	14,22	11.611,06	8,73
TR32 (Aydın, Denizli, Muğla)	2.872.690	8,74	8.894,89	9,66
TR33 (Manisa, Afyon, Kütahya, Uşak)	3.014.128	6,78	8.202,56	9,50
TR41 (Bursa, Eskişehir, Bilecik)	3.763.336	9,04	11.431,17	6,07
TR42 (Kocaeli, Sakarya, Düzce, Bolu, Yalova)	3.493.000	11,32	12.901,06	7,39
TR51 (Ankara)	5.040.733	11,97	14.214,11	5,34
TR52 (Konya, Karaman)	2.329.918	7,73	7.689,17	6,94
TR61 (Antalya, Isparta, Burdur)	2.838.072	9,84	10.693,78	9,01
TR62 (Adana, Mersin)	3.847.579	13,57	7.569,89	9,95
TR63 (Hatay, Kahramanmaraş, Osmaniye)	3.087.890	14,39	6.135,67	6,22
TR71 (Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir)	1.532.201	10,43	6.886,22	8,13
TR72 (Kayseri, Sivas, Yozgat)	2.375.697	11,25	7.622,44	6,93
TR81 (Zonguldak, Karabük, Bartın)	1.027.632	8,16	7.130,94	6,39
TR82 (Kastamonu, Çankırı, Sinop)	759.066	6,38	7.336,89	8,36
TR83 (Samsun, Tokat, Çorum, Amasya)	2.756.595	7,08	6.416,72	8,45
TR90 (Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane)	2.576.643	6,57	6.751,83	6,31
TRA1 (Erzurum, Erzincan, Bayburt)	1.072.955	7,01	6.392,44	7,13
TRA2 (Ağrı, Kars, Iğdır, Ardahan)	1.130.540	7,48	4.191,22	9,00
TRB1 (Malatya, Elazığ, Bingöl, Tunceli)	1.677.302	10,77	5.897,61	7,13
TRB2 (Van, Muş, Bitlis, Hakkâri)	2.070.168	14,63	3.862,61	4,34
TRC1 (Gaziantep, Adıyaman, Kilis)	2.555.522	13,07	6.102,50	7,86
TRC2 (Şanlıurfa, Diyarbakır)	3.407.279	15,01	4.185,00	5,81
TRC3 (Mardin, Batman, Şırnak, Siirt)	2.113.718	21,54	4.708,78	3,98

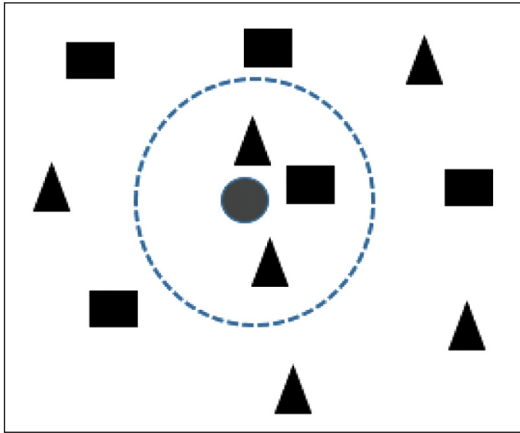
Problemin girdileri belirlenirken yukarıda sayılan nedenlerle en uzun yıl aralığına sahip ekonomik veriler olan GSYİH (kişi başı) ve işsizlik verileri kul-

lanılmıştır. Yöntem olarak tahmin çalışması için en sık kullanılan makine öğrenmesi sınıflandırma yöntemlerinden Naive Bayes (NB) Sınıflandırıcı, K-En

Yakın Komşu [K-nearest neighbors (KNN)] ve Rastgele Orman [Random forest (RF)] metotlarıyla çalışma yürütülmüştür.

NB Sınıflandırıcısı: Olasılık teoremi içerisinde incelenen Bayes teoreminin sınıflandırma problemlerine uygulanmasıyla kullanılır. Sınıflandırılacak yeni bireyin tüm sınıflara ait olup olmadığı olasılık olarak ayrı ayrı hesaplayarak olasılık değeri en yüksek olan sınıfa ait olduğuna dair tahmin yapılır. Gerçek zamanlı tahmin, metin sınıflandırması, “spam” filtreleme, çok sınıflı tahmin gibi problemlere uygulanabilmektedir. Bu çalışmada NB yöntemine ait olan sürekli veri için kullanılan versiyonu olan “Gaussian NB” yöntemi kullanılmıştır.³³ Hiperparametre olarak varsayılan hiperparametreler kullanılmıştır.

KNN algoritması sınıflandırma problemleri için kullanılan algoritmalarından biridir. Öncelikle sistem eğitim sürecinden geçirilir sonrasında yeni bireyin hangi sınıfa ait olduğu tahmin edilmeye çalışılır. Tahmin çalışması sırasında tahmin edilecek yeni bağımsız değişkenin mevcut sınıflar dikkate alınarak k tane komşuya yakınlığına bakılmakta ve k tane komşunun ait oldukları sınıf çoğunluğuna göre sınıfı tahmin edilmeye çalışılmaktadır. Kullanımı ve problemlere uygulanması basit olup birçok alanda kullanılabilir. ³⁴ En yakın 3 komşuya bakılarak sınıfına karar verilmek istenen yeni üyenin (daire) örnekte en yakın 3 komşudan 2 tanesinin üçgen olması nedeniyle üçgen olduğu tahmini yapılacaktır (Şekil 1). Çalışmada KNN yöntemi kullanılırken hiperparametreler olarak; komşuluk (n_neighbor) 5, ağırlık



ŞEKİL 1: Örnek K-En Yakın Komşu tahmin gösterimi.

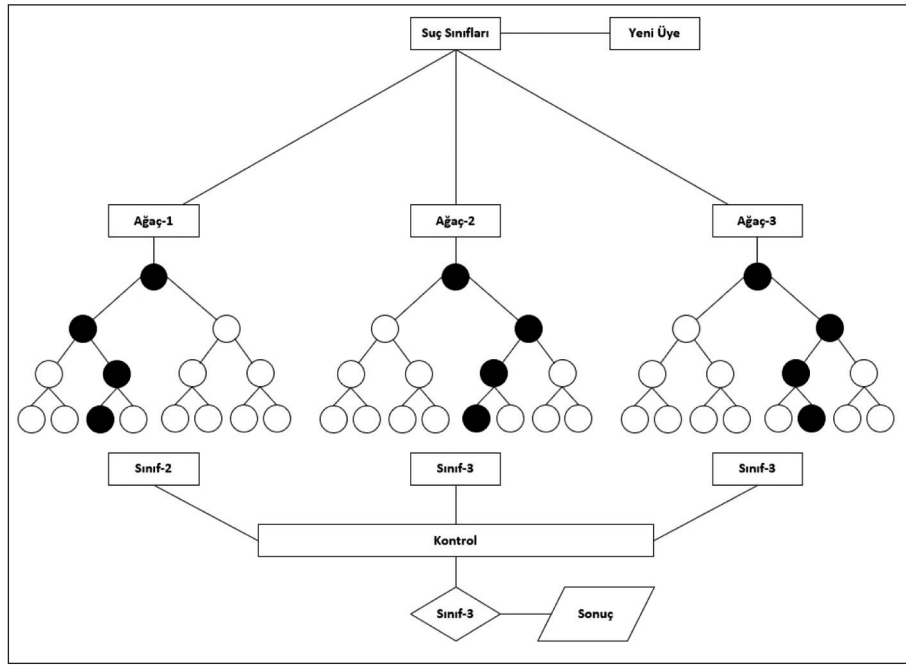
(weights) “uniform”, yaprak boyutu (leaf size) 30, “metric minkowski” olarak belirlenmiştir.

RF algoritması hem regresyon hem sınıflandırma problemlerine uygulanabilen yapay zekâ makine öğrenmesi regresyon yöntemlerinden biridir. Karar ağacının (decision tree) algoritmasının bir uzantısıdır, önce eğitim verileriyle çok sayıda karar ağacı oluşturulur, ardından yeni veriler oluşturulan ağaçlarda denenerek en iyi sonuç bulunmaya çalışılır (Şekil 2).³⁵ Yöntem kullanılırken hiperparametreler olarak; derinlik 5, ağaç sayısı 25 olacak şekilde yapılmış olup, verinin %80’i eğitim verisi, kalan %20’si ise test verisi olacak şekilde uygulama yapılmıştır.

Hükümlü sayısı tahmini yapılırken her bir suç türü için 3 ayrı tahmin yapılmıştır. Öncelikle aynı yılın girdileriyle aynı yılın hükümlü sayısı tahmin edilmiş, sonrasında 1 yıl sonrasının ve son olarak da 2 yıl sonrasının hükümlü sayısı tahmin edilmiştir.

Yapılan çalışma Intel Corporation-Amerika Birleşik Devletleri tarafından üretilen Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz işlemciye ve 8 GB hafızaya sahip bir bilgisayarda yapılmıştır. Her suç türü, her yöntem, her eğitim-test grubu ve her yıl aralığı için ayrı ayrı, 1980’lerde Hollanda’da tasarlanmış ve farklı yıllarda yeni versiyonlarına güncellenmiş olan, Python programlama dilinde yazılan kodlar çalıştırılarak sonuçlar elde edilmiştir. Kodların Python programlama dilinde yazılması sürecinde Amerika Birleşik Devletlerinde bulunan Anaconda Inc. isimli şirket tarafından geliştirilmiş olan Anaconda isimli entegre geliştirme ortamı kullanılmıştır. Herhangi bir suç türünün herhangi bir yıl tahmini için 3 ayrı algoritmanın Python kodlarıyla çalışma süresi hesaplanmıştır. Buna göre NB algoritmasıyla 668 milisaniye, KNN algoritmasıyla 732 milisaniye, RF algoritmasıyla 1026 milisaniye sürede tahmin sonucu elde edilebilmiştir.

GSYİH ve işsizlik verisi aynı anda kullanılarak; aynı yılın suç sayısı, 1 yıl sonrasının suç sayısı, 2 yıl sonrasının suç sayısı ayrı ayrı her bir suç türü için tahmin edilmiştir. Tahminler, doğruluk (accuracy) metriğiyle yorumlanmıştır. Doğruluk metriği 0-1 arasında sonuç alınabilmekte olup ve 1’e yakın olması doğruluğun arttığını, 0’a yakın olması doğruluğun azaldığını göstermektedir. Örnek olarak 1 yıl



ŞEKİL 2: Örnek Rastgele Orman tahmin gösterimi.

sonraki suçun tahmini için 2006 yılının işsizlik ve GSYİH verisiyle 2007 yılındaki suç tahmin edilmeye çalışılmıştır. Ülke geneli ortalama ve 26 bölgenin sonuçlarının ayrı ayrı suç tipine göre gösterilmiştir (Tablo 5).

Yöntemlerin, farklı eğitim-test oranlarının ve farklı yıl tahminlerinin daha rahat yorumlanabilmesi için çalışmanın bundan sonraki bölümünde Türkiye geneli tahminler de yeni bir bölge gibi kabul edilmiştir. Ayrıca çalışmanın tamamında yer alan 26 bölgeye bu yeni bölge de eklenerek toplam 27 bölgenin ortalamalarının alındığı tablolar kullanılmıştır.

BULGULAR

Bu bölümde yapılan tahminler sonucunda oluşturulan tablolardaki sonuçlar belirtilmiştir. Öncelikli olarak %70 eğitim verisi %30 test verisi tahminleriyle %80 eğitim verisi %20 test verisi tahminlerinin hepsi tüm yöntemler ve tüm yıllar bir arada değerlendirilerek incelenmiştir. Bu inceleme sonucunda, %80 eğitim %20 test verisi sonuçlarının daha iyi olduğu tespit edilmiştir (Tablo 5).

%80 eğitim %20 test verisine göre yapılan tahminlerin daha başarılı olması nedeniyle bu oranda

doğrultusunda tüm yöntemlerin Aynı yıl sonrası suç oranı tahmini, 1 yıl sonrası suç oranı tahmini, 2 yıl sonrası suç oranı tahmini sonuçları birlikte gösterilmiştir. Tabloda da görüleceği üzere bazı yöntemler bazı suç türlerini daha başarılı olarak tahmin edebilmiştir (Tablo 6). Örnek olarak cinayet suçu oranı RF yöntemiyle diğer suçlara göre daha başarılı tahmin edilmiştir. Ayrıca yine aynı tabloda aynı yıldaki suç oranının tahmin sonuçlarıyla 1 yıl sonrası ve 2 yıl sonrası suç oranının tahmin sonuçları kıyaslandığında, 1 yıl sonrası ve 2 yıl sonrası tahminlerin suçların büyük bölümünde daha başarılı olduğu tespit edilmiştir.

Her bir yöntemle ayrı ayrı 1 yıl sonrası ve 2 yıl sonrası suç oranı tahminlerinin daha iyi yorumlanabilmesi için 27 bölgenin tamamına ilişkin tahmin grafikleri sunulmuştur (Şekil 3, Şekil 4, Şekil 5). Devamında her üç yöntemin her suç türü için tahmin başarımının kıyaslanabilmesi için birleştirilmiş tahmin grafiği sunulmuştur (Şekil 6).

Sonuçların yorumlanmasına katkısının olması amacıyla her 3 yöntemin karışıklık matrisleri gösterilmiştir. Şekiller her suçun aynı yıl, 1 yıl sonrası ve 2 yıl sonrası sonuçlarının toplamını göstermektedir (Şekil 7, Şekil 8, Şekil 9).

TABLO 5: Türkiye geneli ve 26 bölge için suç oranı tahminleme sonuçları.

Suç adı	Türkiye geneli sonuçlar																	
	Aynı yıl						1 yıl sonrası						2 yıl sonrası					
	RF		NB		KNN		RF		NB		KNN		RF		NB		KNN	
	%20	%30	%20	%30	%20	%30	%20	%30	%20	%30	%20	%30	%20	%30	%20	%30	%20	%30
Cinayet	0,67	0,80	0,50	0,67	0,50	0,67	1,00	1,00	0,25	0,40	0,75	0,80	1,00	0,80	1,00	0,80	1,00	0,80
Dolandırıcılık	0,67	0,60	0,33	0,40	0,67	0,80	0,67	0,60	0,67	0,60	0,67	0,60	0,67	0,80	0,67	0,40	0,33	0,60
Hakaret	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Hırsızlık	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Kaçakçılık	0,67	0,80	0,67	0,60	0,67	0,60	0,67	0,80	0,67	0,60	1,00	1,00	0,67	0,80	0,67	0,60	1,00	1,00
Kötü muamele	1,00	0,80	0,67	0,80	1,00	1,00	1,00	0,60	0,67	0,80	1,00	1,00	1,00	0,80	0,67	0,80	1,00	1,00
Rüşvet	0,67	0,40	0,67	0,40	0,33	0,40	0,67	0,60	0,67	0,60	0,33	0,60	0,67	0,80	0,67	0,80	0,33	0,60
Sahtecilik	1,00	0,80	1,00	0,80	1,00	0,80	1,00	0,80	1,00	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,80
Silah-Bıçak	0,33	0,60	0,33	0,40	0,33	0,40	0,67	0,80	0,67	0,80	0,33	0,60	1,00	0,60	1,00	1,00	0,67	0,60
Trafik suçları	0,33	0,60	0,67	0,60	0,67	0,80	0,33	0,40	0,67	0,60	0,67	0,80	0,33	0,40	0,67	0,60	0,67	0,80
Uyuşturucu ticareti	1,00	0,80	1,00	0,80	1,00	0,80	1,00	0,80	1,00	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,80
Uyuşturucu kullanma, satın alma	0,33	0,80	0,33	0,40	0,67	0,80	0,67	0,60	0,33	0,40	0,67	0,80	0,67	0,60	0,33	0,40	1,00	1,00
Yağma	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Yaralama	1,00	0,80	1,00	0,80	1,00	0,80	1,00	0,80	1,00	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,80
Zimmet	0,67	0,40	0,67	0,80	0,67	0,60	0,67	0,40	1,00	0,60	1,00	1,00	0,67	0,60	1,00	1,00	0,67	0,60
Ortalama	0,76	0,75	0,72	0,70	0,77	0,76	0,82	0,72	0,77	0,72	0,83	0,88	0,84	0,81	0,84	0,83	0,84	0,83
Suç adı	26 bölge sonuçları																	
Cinayet	0,82	0,88	0,53	0,66	0,52	0,68	0,96	0,92	0,33	0,45	0,57	0,65	0,99	0,84	0,85	0,72	0,87	0,75
Dolandırıcılık	0,62	0,69	0,65	0,52	0,62	0,74	0,67	0,55	0,71	0,56	0,81	0,64	0,63	0,59	0,69	0,59	0,71	0,78
Hakaret	0,91	0,88	0,82	0,83	0,83	0,83	0,92	0,79	0,87	0,84	0,81	0,87	0,92	0,90	0,88	0,84	0,83	0,83
Hırsızlık	0,99	0,98	0,99	0,98	0,99	0,98	0,99	0,84	0,99	0,98	0,99	0,99	0,99	0,99	0,99	0,98	0,99	0,98
Kaçakçılık	0,63	0,63	0,62	0,59	0,62	0,64	0,74	0,75	0,63	0,59	0,82	0,88	0,73	0,75	0,56	0,59	0,76	0,85
Kötü muamele	0,68	0,57	0,62	0,65	0,73	0,72	0,69	0,51	0,67	0,68	0,73	0,76	0,69	0,58	0,63	0,71	0,73	0,75
Rüşvet	0,38	0,42	0,54	0,55	0,51	0,54	0,37	0,48	0,56	0,55	0,53	0,57	0,37	0,50	0,58	0,58	0,51	0,55
Sahtecilik	1,00	0,75	0,96	0,75	0,96	0,75	0,99	0,78	0,96	0,78	0,97	0,97	0,99	0,97	0,96	0,86	0,96	0,79
Silah-Bıçak	0,37	0,40	0,33	0,39	0,44	0,43	0,44	0,48	0,51	0,54	0,47	0,62	0,50	0,55	0,59	0,58	0,55	0,48
Trafik suçları	0,37	0,58	0,68	0,61	0,60	0,76	0,50	0,52	0,67	0,61	0,76	0,83	0,49	0,48	0,68	0,61	0,69	0,79
Uyuşturucu ticareti	0,95	0,89	0,94	0,88	0,94	0,88	0,97	0,86	0,94	0,88	0,92	0,95	0,97	0,97	0,91	0,88	0,94	0,89
Uyuşturucu kullanma, satın alma	0,42	0,64	0,63	0,62	0,59	0,74	0,73	0,69	0,72	0,62	0,90	0,93	0,69	0,66	0,63	0,58	0,85	0,89
Yağma	0,97	0,88	0,97	0,74	0,97	0,87	0,99	0,92	0,85	0,79	0,99	0,99	0,99	0,98	0,83	0,83	0,97	0,95
Yaralama	0,97	0,84	0,92	0,83	0,91	0,81	0,97	0,81	0,95	0,84	0,97	0,98	0,97	0,98	0,96	0,94	0,94	0,82
Zimmet	0,42	0,35	0,46	0,48	0,47	0,42	0,53	0,44	0,47	0,47	0,40	0,47	0,45	0,43	0,40	0,45	0,36	0,36
Ortalama	0,70	0,69	0,71	0,67	0,71	0,72	0,76	0,69	0,72	0,68	0,78	0,81	0,76	0,75	0,74	0,72	0,78	0,76

%20: Toplam verinin %80'i eğitim %20'si test verisi; %30: Toplam verinin %70'i eğitim için %30'si test verisi; RF: Rastgele Orman; NB: Naive Bayes; KNN: K-En Yakın Komşu.

TARTIŞMA

Bu bölümde yapılan yorumlar, 27 bölgenin [İstatistiki Bölge Birimleri Sınıflandırması-2 bölgeleri (26 bölge) ve Türkiye geneli] ortalaması dikkate alınarak yapılmıştır. Hırsızlık suçunun tüm yöntemler ve yıl aralıkları açısından en başarılı olarak tahmin edilen suç olduğu tespit edilmiştir. Sahtecilik, yaralama, uyuşturucu imal ve ticareti suçlarıyla ilgili olarak tüm

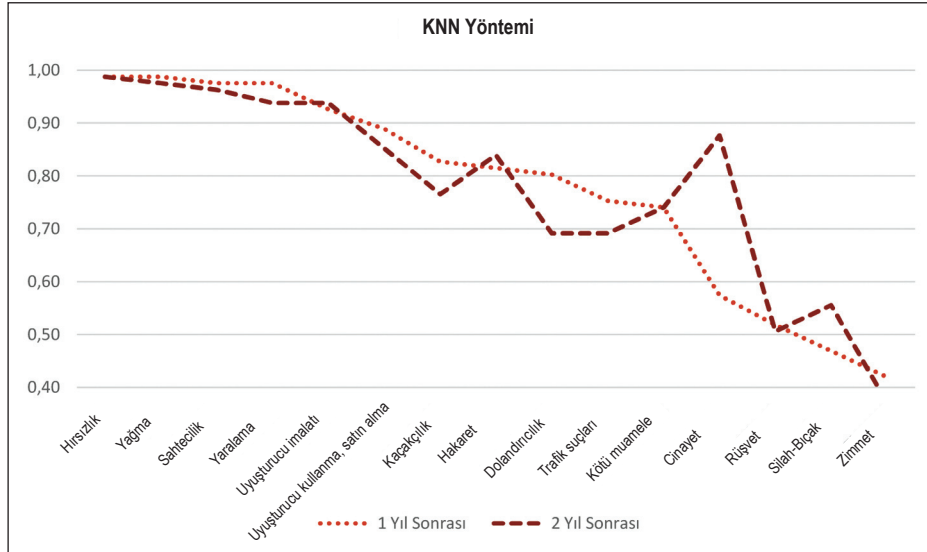
yöntemlerle yapılan tahminlerde de 0,9'un üzerinde bir tahmin başarısı elde edilmiştir.

Yağma suçunun RF ve KNN yöntemleriyle yapılan tahminlerin NB yöntemiyle yapılan tahminlerden daha başarılı olduğu, cinayet ve hakaret suçlarının RF yöntemiyle yapılan tahminlerin diğer yöntemlerden daha başarılı olduğu, kötü muamele, dolandırıcılık, kaçakçılık ve uyuşturucu kullanma ve kullanmak için satın alma suçlarının KNN yönte-

TABLO 6: Tüm bölgeler ve Türkiye geneli (27 bölge) için yapılan tahminler.

Suç adı	Aynı yıl tahmini			1 yıl sonrası tahmini			2 yıl sonrası tahmini		
	KNN	NB	RF	KNN	NB	RF	KNN	NB	RF
Cinayet	0,52	0,53	0,81	0,57	0,32	0,96	0,88	0,85	0,99
Dolandırıcılık	0,62	0,64	0,62	0,80	0,70	0,67	0,69	0,69	0,63
Hakaret	0,84	0,83	0,91	0,81	0,88	0,93	0,84	0,89	0,93
Hırsızlık	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99
Kaçakçılık	0,62	0,62	0,63	0,83	0,63	0,74	0,77	0,57	0,73
Kötü muamele	0,74	0,62	0,69	0,74	0,67	0,70	0,74	0,63	0,70
Rüşvet	0,51	0,54	0,40	0,52	0,57	0,38	0,51	0,58	0,38
Sahtecilik	0,96	0,96	1,00	0,98	0,96	0,99	0,96	0,96	0,99
Silah-Bıçak	0,43	0,33	0,37	0,47	0,52	0,44	0,56	0,60	0,52
Trafik suçları	0,60	0,68	0,37	0,75	0,67	0,49	0,69	0,68	0,48
Uyuşturucu ticareti	0,94	0,94	0,95	0,93	0,94	0,98	0,94	0,91	0,98
Uyuşturucu kullanma, satın alma	0,59	0,62	0,42	0,89	0,70	0,73	0,85	0,62	0,69
Yağma	0,98	0,98	0,98	0,99	0,85	0,99	0,98	0,84	0,99
Yaralama	0,91	0,93	0,98	0,98	0,95	0,98	0,94	0,96	0,98
Zimmet	0,48	0,47	0,43	0,42	0,49	0,53	0,37	0,42	0,46
Genel toplam	0,72	0,71	0,70	0,78	0,72	0,77	0,78	0,75	0,76

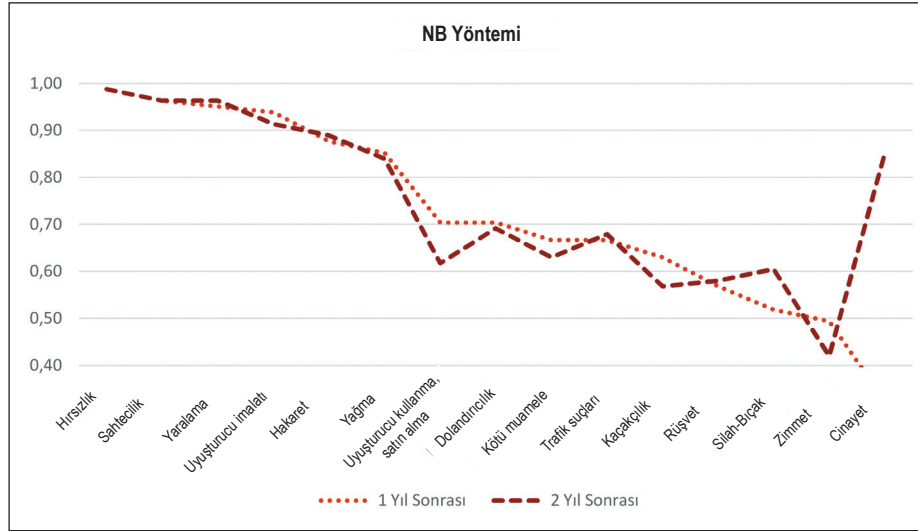
KNN: K-En Yakın Komşu; NB: Naive Bayes; RF: Rastgele Orman.



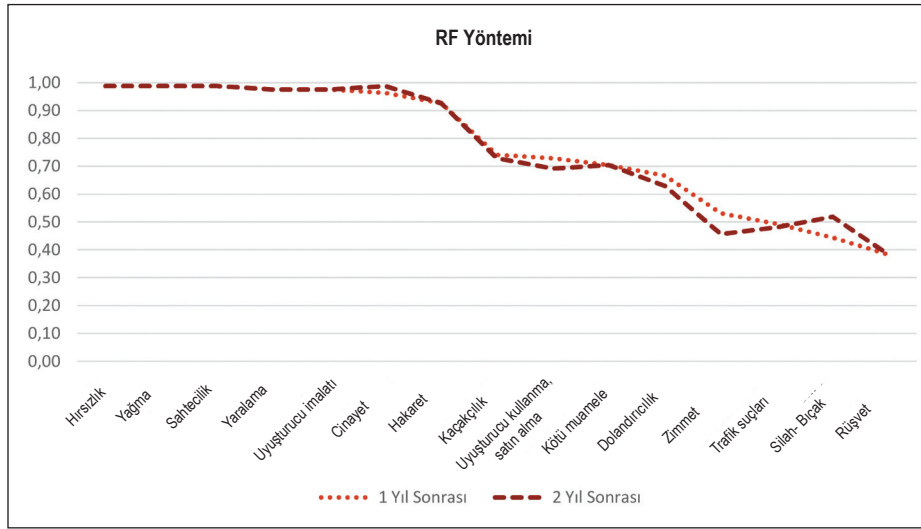
ŞEKİL 3: 27 bölge (Türkiye geneli ve 26 bölge) için KNN yöntemiyle 1 yıl sonrası ve 2 yıl sonrası suçlu oranı tahmin grafiği.
KNN: K-En Yakın Komşu.

miyle yapılan diğer yöntemlerden daha başarılı olduğu, trafik suçları, ateşli silahlar ve bıçaklar kanununa muhalefet, rüşvet ve zimmet suçlarının tüm yöntemlerle yapılan tahminlerin başarılı sayılabile-

cek seviyenin (0,7) altında kaldığı tespit edilmiştir. Başarı oranı düşük sayılan suçlar, bazı bölgelerde başarılı olmuş olsa da 27 bölgenin ortalamasına göre değerlendirmeler yapılmıştır.



ŞEKİL 4: 27 bölge (Türkiye geneli ve 26 bölge) için NB yöntemiyle 1 yıl sonrası ve 2 yıl sonrası suçlu oranı tahmin grafiği. NB: Naive Bayes.

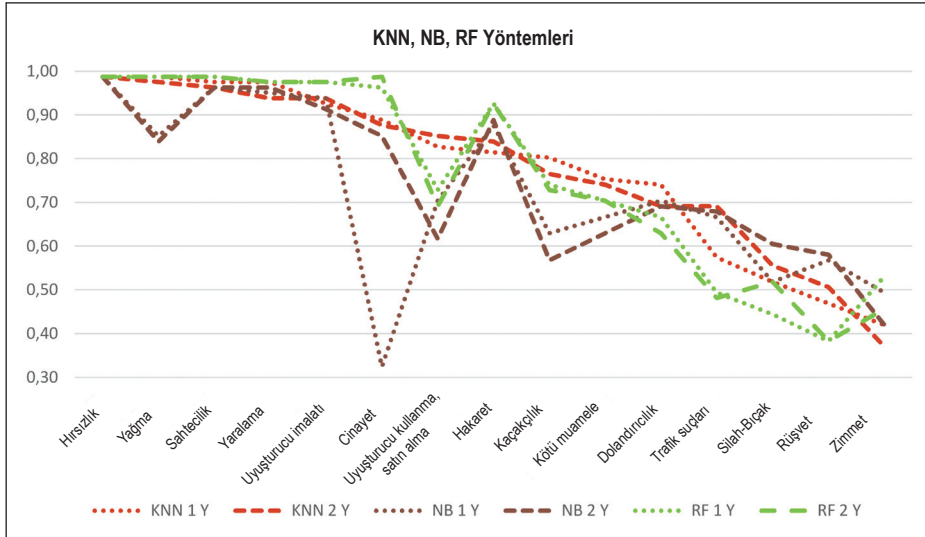


ŞEKİL 5: 27 bölge (Türkiye geneli ve 26 bölge) için RF yöntemiyle 1 yıl sonrası ve 2 yıl sonrası suçlu oranı tahmin grafiği. RF: Rastgele Orman.

Bulunulan ülkenin ekonomik durumuyla suç arasında bağ anılan literatür örneklerinde açıkça görülmektedir. Sonuçlar bu hâliyle ekonomik durumdaki problemlerin suçu etkilediğini göstermektedir.

Bu çalışmada yapılan tahmin sonucunda hırsızlık, yağma, sahtecilik, yaralama, uyuşturucu imal ve ticareti suçlarını büyük oranda ülke genelindeki her bölgede doğru tahmin edilebildiği görülmektedir. İşsizlik ve GSYİH gibi önemli ekonomik göstergelerin

bu suçları daha doğru tahmin edebiliyor olmasının yanında bu 5 suçtan özellikle hırsızlık, yağma ve sahtecilik suçlarının diğer tüm suçlara kıyasla ekonomik problemlerden daha fazla kaynaklandığını söylemek mümkündür. Ancak, diğer suçların tahminlerindeki başarı oranlarının nispeten daha düşük olması ilave parametrelerin de dikkate alınması gerektiğini göstermektedir. Örnek olarak kaçakçılık, uyuşturucu imal ve ticareti suçlarıyla ilgili olarak bulunulan böl-



ŞEKİL 6: 27 bölge (Türkiye geneli ve 26 bölge) için KNN, RF, NB yöntemleriyle 1 yıl sonrası ve 2 yıl sonrası birleştirilmiş suçlu oranı tahmin grafiği. KNN: K-En Yakın Komşu; RF: Rastgele Orman; NB: Naive Bayes.

genin/ülkenin sınırlarında bulunan ülkelerin kaçakçılığı tetikler durumunda olması ilave neden olarak, cinayet, hakaret, ateşli silahlar ve bıçaklar kanununa muhalefet ve kötü muamele suçlarıyla ilgili olarak eğitimsizlik, aşiret yapısının devamı, bireysel silahlanma seviyesi gibi hususlar ilave nedenler olarak, dolandırıcılık suçuyla ilgili olarak halkın farkındalık seviyesinin düşüklüğü ve eğitim kalitesi ilave nedenler olarak, trafik suçlarıyla ilgili olarak eğitim ve denetim yapısının kalitesi ilave nedenler olarak, rüşvet ve zimmet suçlarıyla ilgili olarak bulunulan kamu bürokrasisinin kontrole uygun olmayışı ve çalışan maaşları ilave nedenler olarak belirtmek mümkündür.

Ayrıca literatür örneklerinde daha küçük suç gruplarının tahmin edilmesi ve genelde aynı yıldaki suçların tahmin edilmesi amaçlarıyla çalışmaların yürütüldüğü görülmekteyken bu çalışmada 15 ayrı suçun her biri için ayrı ayrı ve 3 farklı yıl grubunun tahmininin kıyaslanmış olması çalışmanın önemli bir farkını ortaya koymaktadır.

Bu çalışmayla, sosyal bilimlerde suçla ekonomik göstergeler arasında var olduğu belirtilen bağın bu çalışmayla makine öğrenmesiyle de varlığı ortaya konulmuştur.

Çalışmanın önemli bir kısıtlılığı olarak verinin yıl bazında miktarının az olması hususu olsa da TÜİK tarafından mevcut verinin düzenli olarak güncellenmesi benzer çalışmaların doğruluğunu güçlendirecektir.

SONUÇ

Sonuç olarak, makine öğrenimi algoritmalarının suç oranı tahmini için etkili bir araç olarak kullanılabilirliği değerlendirilmiştir. Bu durum sosyal bilimlerde yapılan çalışmalarda ortaya konulmuş olan suç ile ekonomik göstergeler arasındaki ilişkiyi doğrular niteliktedir. Veri arttıkça (yıllar geçtikçe) doğruluk tahmini hâlihazırda yüksek olan bölgelerde ve doğruluk tahmini düşük olduğu bölgelerde tahmin performansının daha detaylı gözlemlenmesi, yöntem uygulanırken kullanılan hiperparametrelerin ihtiyaçlar doğrultusunda güncellenmesinin mümkün olabileceği düşünülmektedir.

Ülkemizdeki yargılama süreleri de dikkate alınarak 1 yıl sonrası ya da 2 yıl sonrası için tahminlerin aynı yıl tahminlerine kıyasla daha başarılı çıkmış olması diğer önemli bir sonuç olarak dikkatle tartışılmalıdır. Ülkemizdeki yargılama sürelerine ilişkin bir projeksiyon sunan bu durum ileride yapılacak aylık tahmin çalışmalarıyla da detaylandırılabilir.

K-Nearest Neighbor Yöntemiyle Yapılan Tahminler Confusion Matrix Tablosu			
Cinayet Suçlu Oranı Tahmini Confusion Matrix			
Cin	0*	1*	2*
0'	73	18	22
1'	18	1	0
2'	3	5	130
Rüşvet Suçlu Oranı Tahmini Confusion Matrix			
kot	0*	1*	2*
0'	128	7	21
1'	5	10	24
2'	1	5	42
Uyuşturucu İmal ve Ticareti Suçlu Oranı Tahmini Confusion Matrix			
uimal	0*	1*	2*
0'	97	6	5
1'	0	1	3
2'	0	2	129
Dolandırıcılık Suçlu Oranı Tahmini Confusion Matrix			
Dol	0*	1*	2*
0'	45	21	0
1'	11	22	36
2'	0	4	104
Rüşvet Suçlu Oranı Tahmini Confusion Matrix			
rus	0*	1*	2*
0'	69	29	22
1'	17	21	22
2'	22	7	34
Uyuşturucu Madde Kullanma ve Kullanmak İçin Satın Alma Suçlu Oranı Tahmini Confusion Matrix			
ukulsat	0*	1*	2*
0'	116	9	4
1'	5	15	4
2'	15	17	58
Hakaret Suçlu Oranı Tahmini Confusion Matrix			
hak	0*	1*	2*
0'	45	18	9
1'	3	3	6
2'	0	5	154
Sahtecilik Suçlu Oranı Tahmini Confusion Matrix			
sah	0*	1*	2*
0'	75	0	6
1'	0	1	2
2'	0	0	159
Yaralama Suçlu Oranı Tahmini Confusion Matrix			
yar	0*	1*	2*
0'	78	2	0
1'	1	2	4
2'	0	7	149
Hırsızlık Suçlu Oranı Tahmini Confusion Matrix			
hir	0*	1*	2*
0'	81	0	0
1'	0	0	3
2'	0	0	159
Ateşli Silah ve Bıçak Kanununa Muhalefet Suçlu Oranı Tahmini Confusion Matrix			
silbic	0*	1*	2*
0'	58	31	13
1'	32	26	23
2'	7	19	34
Yağma Suçlu Oranı Tahmini Confusion Matrix			
yag	0*	1*	2*
0'	78	3	0
1'	0	1	2
2'	0	0	159
Kaçakçılık Suçlu Oranı Tahmini Confusion Matrix			
kac	0*	1*	2*
0'	79	17	0
1'	5	21	22
2'	4	16	79
Trafik Suçları Suçlu Oranı Tahmini Confusion Matrix			
tra	0*	1*	2*
0'	144	19	2
1'	16	4	28
2'	7	5	18
Zimmet Suçlu Oranı Tahmini Confusion Matrix			
zim	0*	1*	2*
0'	51	20	22
1'	18	20	37
2'	24	19	32
KNN Yöntemi Tüm Suçlar Confusion Matrix			
KNN GENEL	0*	1*	2*
0'	1217	200	126
1'	131	148	216
2'	83	111	1440

0', 1', 2' Gerçek Suç Oranı Sınıfları (Actual Class) 0:Normal Suç (0-0,25), 1:Fazla Suç (0,25-0,40), 2:Çok Fazla Suç (0,40-1,00)

0*, 1*, 2* Tahmin Edilen Suç Oranı Sınıfları Predicted Class) 0:Normal Suç (0-0,25), 1:Fazla Suç (0,25-0,40), 2:Çok Fazla Suç (0,40-1,00)

ŞEKİL 7: KNN yöntemiyle yapılan tahminlere ilişkin karışıklık matrisi.

KNN: K-En Yakın Komşu.

Naive Bayes Yöntemiyle Yapılan Tahminler Confusion Matrix Tablosu

Cinayet Suçlu Oranı Tahmini Confusion Matrix				Rüşvet Suçlu Oranı Tahmini Confusion Matrix				Uyuşturucu İmal ve Ticareti Suçlu Oranı Tahmini Confusion Matrix			
Cin	0*	1*	2*	kot	0*	1*	2*	uimal	0*	1*	2*
0'	77	11	25	0'	143	8	5	0'	74	3	4
1'	15	1	3	1'	30	4	5	1'	2	1	3
2'	9	8	121	2'	37	3	8	2'	5	0	151

Dolandırıcılık Suçlu Oranı Tahmini Confusion Matrix				Rüşvet Suçlu Oranı Tahmini Confusion Matrix				Uyuşturucu Madde Kullanma ve Kullanmak İçin Satın Alma Suçlu Oranı Tahmini Confusion Matrix			
Dol	0*	1*	2*	rus	0*	1*	2*	ukulsat	0*	1*	2*
0'	58	4	4	0'	90	13	17	0'	126	1	2
1'	22	3	44	1'	30	12	18	1'	18	5	1
2'	2	2	104	2'	23	5	35	2'	63	1	26

Hakaret Suçlu Oranı Tahmini Confusion Matrix				Sahtecilik Suçlu Oranı Tahmini Confusion Matrix				Yaralama Suçlu Oranı Tahmini Confusion Matrix			
hak	0*	1*	2*	sah	0*	1*	2*	yar	0*	1*	2*
0'	56	2	14	0'	75	1	5	0'	77	1	3
1'	5	0	7	1'	3	0	0	1'	3	0	3
2'	5	0	154	2'	0	0	159	2'	3	0	153

Hırsızlık Suçlu Oranı Tahmini Confusion Matrix				Ateşli Silah ve Bıçak Kanununa Muhalefet Suçlu Oranı Tahmini Confusion Matrix				Yağma Suçlu Oranı Tahmini Confusion Matrix			
hir	0*	1*	2*	silbic	0*	1*	2*	yag	0*	1*	2*
0'	81	0	0	0'	56	12	34	0'	78	3	0
1'	0	0	3	1'	32	11	38	1'	3	0	0
2'	0	0	159	2'	7	2	51	2'	11	10	138

Kaçakçılık Suçlu Oranı Tahmini Confusion Matrix				Trafik Suçları Suçlu Oranı Tahmini Confusion Matrix				Zimmet Suçlu Oranı Tahmini Confusion Matrix			
kac	0*	1*	2*	tra	0*	1*	2*	zim	0*	1*	2*
0'	85	5	6	0'	164	0	1	0'	53	13	27
1'	33	8	7	1'	48	0	0	1'	17	14	44
2'	38	7	54	2'	30	0	0	2'	18	12	45

NB Yöntemi Tüm Suçlar Confusion Matrix			
NB GENEL	0*	1*	2*
0'	1293	77	147
1'	261	59	176
2'	251	50	1358

0', 1', 2' Gerçek Suç Oranı Sınıfları (Actual Class) 0:Normal Suç (0-0,25), 1:Fazla Suç (0,25-0,40), 2:Çok Fazla Suç (0,40-1,00)

0*, 1*, 2* Tahmin Edilen Suç Oranı Sınıfları Predicted Class) 0:Normal Suç (0-0,25), 1:Fazla Suç (0,25-0,40), 2:Çok Fazla Suç (0,40-1,00)

ŞEKİL 8: NB yöntemiyle yapılan tahminlere ilişkin karışıklık matrisi.

NB: Naive Bayes.

Random Forest Yöntemiyle Yapılan Tahminler Confusion Matrix Tablosu

Cinayet Suçlu Oranı Tahmini Confusion Matrix				Rüşvet Suçlu Oranı Tahmini Confusion Matrix				Uyuşturucu İmal ve Ticareti Suçlu Oranı Tahmini Confusion Matrix			
Cin	0*	1*	2*	kot	0*	1*	2*	uimal	0*	1*	2*
0'	96	10	7	0'	127	7	22	0'	80	0	1
1'	9	8	2	1'	4	3	32	1'	1	0	5
2'	1	2	135	2'	8	0	40	2'	0	1	155
Dolandırıcılık Suçlu Oranı Tahmini Confusion Matrix				Rüşvet Suçlu Oranı Tahmini Confusion Matrix				Uyuşturucu Madde Kullanma ve Kullanmak İçin Satın Alma Suçlu Oranı Tahmini Confusion Matrix			
Dol	0*	1*	2*	rus	0*	1*	2*	ukulsat	0*	1*	2*
0'	56	7	3	0'	53	3	64	0'	76	10	43
1'	33	6	30	1'	25	1	34	1'	7	2	15
2'	13	2	93	2'	23	0	40	2'	18	1	71
Hakaret Suçlu Oranı Tahmini Confusion Matrix				Sahtecilik Suçlu Oranı Tahmini Confusion Matrix				Yaralama Suçlu Oranı Tahmini Confusion Matrix			
hak	0*	1*	2*	sah	0*	1*	2*	yar	0*	1*	2*
0'	69	0	3	0'	81	0	0	0'	81	0	0
1'	9	0	3	1'	0	1	2	1'	0	1	5
2'	3	1	155	2'	0	0	159	2'	0	1	155
Hırsızlık Suçlu Oranı Tahmini Confusion Matrix				Ateşli Silah ve Bıçak Kanununa Muhalefet Suçlu Oranı Tahmini Confusion Matrix				Yağma Suçlu Oranı Tahmini Confusion Matrix			
hir	0*	1*	2*	silbic	0*	1*	2*	yag	0*	1*	2*
0'	81	0	0	0'	56	4	42	0'	81	0	0
1'	3	0	0	1'	30	2	49	1'	1	0	2
2'	0	0	159	2'	10	0	50	2'	1	0	158
Kaçakçılık Suçlu Oranı Tahmini Confusion Matrix				Trafik Suçları Suçlu Oranı Tahmini Confusion Matrix				Zimmet Suçlu Oranı Tahmini Confusion Matrix			
kac	0*	1*	2*	tra	0*	1*	2*	zim	0*	1*	2*
0'	77	3	16	0'	86	17	62	0'	58	5	30
1'	4	1	43	1'	15	2	31	1'	16	2	57
2'	7	0	92	2'	8	1	21	2'	18	2	55
Random Forest Yöntemi Tüm Suçlar Confusion Matrix											
RF GENEL	0*	1*	2*								
0'	1158	66	293								
1'	157	29	310								
2'	110	11	1538								

0', 1', 2' Gerçek Suç Oranı Sınıfları (Actual Class) 0:Normal Suç (0-0,25), 1:Fazla Suç (0,25-0,40), 2:Çok Fazla Suç (0,40-1,00)

0*, 1*, 2* Tahmin Edilen Suç Oranı Sınıfları Predicted Class) 0:Normal Suç (0-0,25), 1:Fazla Suç (0,25-0,40), 2:Çok Fazla Suç (0,40-1,00)

ŞEKİL 9: RF yöntemiyle yapılan tahminlere ilişkin karışıklık matrisi.

RF: Rastgele Orman.

İlerleyen süreçte bu çalışmanın bir sonraki adımı olarak, aynı makine öğrenmesi algoritmalarının performansı daha fazla girdi eklenerek artırılabilir. Farklı makine öğrenmesi algoritmaları eklenerek uygulamalar genişletilebilir. Ayrıca bu çalışmalardan sonra aynı yöntemlerin farklı ülkelerden veri setleri ile kullanılması yöntemin doğruluğunun genelleştirilmesi için önemli bir fırsat olabilir. Ayrıca tahmin bölgelerinin daha küçültülebildiği durumlarda örnek olarak il ya da ilçe seviyesinde hem suç hem GSYİH hem de işsizlik verilerinin yıllık ya da aylık bazda alınabilmesi ileride daha detaylı çalışmalar yapılabilmesini sağlayacaktır.

Doğruluk oranı yüksek ve geliştirilmiş bir tahmin yapısının kurulması suç daha oluşmadan alınacak önlemlerle suçu önleyecek ve böylece hem kolluk kuvvetlerinin yönetimi, yatırım planlamaları, suç oluşmadan önleyebilme kapasitesi ve soruş-

turma-kovuşturma süreçlerinin hiç oluşmayacak olması açılarıyla da önemli katkılar sunacaktır.

Finansal Kaynak

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyebilecek maddi ve/veya manevi herhangi bir destek alınmamıştır.

Çıkar Çatışması

Bu çalışma ile ilgili olarak yazarların ve/veya aile bireylerinin çıkar çatışması potansiyeli olabilecek bilimsel ve tıbbi komite üyeliği veya üyeleri ile ilişkisi, danışmanlık, bilirkişilik, herhangi bir firmada çalışma durumu, hissedarlık ve benzer durumları yoktur.

Yazar Katkıları

Bu çalışma hazırlanırken tüm yazarlar eşit katkı sağlamıştır.

KAYNAKLAR

- Gökulu G. Suç kuramları: biyolojik ve psikolojik yaklaşımlar: eleştirel bir değerlendirme [Criminological theories:biological and psychological approaches: a critical evaluation]. *Social Sciences*. 2019;14(4):1471-88. [Crossref]
- Emsley C. *Crime and Society in England, 1750-1900*. 5th ed. London: Routledge; 2018. [Crossref]
- Nagin DS, Paternoster R. Enduring individual differences and rational choice theories of crime. *Law and Society Review*. 1993;27(3):467-96. [Crossref]
- Raphael S, Winter-Ebmer R. Identifying the effect of unemployment on crime. *The Journal of Law and Economics*. 2001;44(1):259-83. [Crossref]
- Wang H, Kifer D, Graif C, Li Z. Crime rate inference with big data. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1st ed. New York, NY: ACM; 2016. p.635-44. [Crossref]
- Türk Ceza Kanunu, (2004) [cited 2023 22/12/2023]. Available from: [Link]
- Federal Bureau of Investigation Crime Data Explorer [Internet] [Cited: November 22, 2023]. Available from: [Link]
- Government C. Incident-based crime statistics, by detailed violations, Canada, provinces, territories, Census Metropolitan Areas and Canadian Forces Military Police Canada: Canadian Government; 2023. Cited: November 22, 2023. Available from: [Link]
- Wang B, Yin P, Bertozzi AL, Brantingham PJ, Osher SJ, Xin J. Deep learning for real-time crime forecasting and its ternarization. *Chinese Annals of Mathematics, Series B*. 2019;40(6):949-66. [Crossref]
- Zhang X, Liu L, Xiao L, Ji J. Comparison of machine learning algorithms for predicting crime hotspots. *IEEE Access*. 2020;8:181302-10. [Crossref]
- Safat W, Asghar S, Gillani SA. Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE Access*. 2021;9:70080-94. [Crossref]
- Gerber MS. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*. 2014;61:115-25. [Crossref]
- Kang HW, Kang HB. Prediction of crime occurrence from multi-modal data using deep learning. *PLoS One*. 2017;12(4):e0176244. [Crossref] [PubMed] [PMC]
- Özcan SE, Şirin FÖ. Ekonomik suçlar, gelir ve işsizliğin mekânsal ilişkisi: 2014 Türkiye incelemesi [The spatial interactions between economic crimes, per capita income and unemployment: 2014 Turkey case]. *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*. 2021;19(4):79-100. [Crossref]
- Eren M, Özkılbaç S. Türkiye'de genç işsizliği ile suç düzeyi arasındaki ilişki: düzey 1 bölgeleri üzerine bir uygulama [The relationship of youth unemployment and crime level in Turkey: an application on the nuts-1 regions]. *Küresel İktisat ve İşletme Çalışmaları Dergisi*. 2020;9(17):13-20. [Link]
- Costantini M, Meco I, Paradiso A. Do inequality, unemployment and deterrence affect crime over the long run? *Regional Studies*. 2018;52(4):558-71. [Crossref]
- Dikkaya M, Aytakin İ. İktisadi büyüme işsizlik ve suç arasında nedensellik ilişkisi: Türkiye örneği [The causality among economic growth, unemployment and crime: evidence from Turkey]. *Ekonomi İşletme Siyaset ve Uluslararası İlişkiler Dergisi*. 2020;6(2):261-74. [Link]
- Jawadi F, Mallick SK, Cheffou AI, Augustine A. Does higher unemployment lead to greater criminality? Revisiting the debate over the business cycle. *Journal of Economic Behavior & Organization*. 2021;182:448-71. [Crossref]
- Nordin M, Almén D. Long-term unemployment and violent crime. *Empirical Economics*. 2017;52(1):1-29. [Crossref]
- Andresen MA. Unemployment, GDP, and crime: the importance of multiple measurements of the economy. *Canadian Journal of Criminology and Criminal Justice*. 2015;57(1):35-58. [Crossref]

21. Buonanno P, Drago F, Galbiati R. Response of crime to unemployment: an international comparison. *Journal of Contemporary Criminal Justice*. 2014;30(1):29-40. [Crossref]
22. Entorf H, Sieger P. Does the link between unemployment and crime depend on the crime level? A quantile regression approach. *IZA*. 2014:1-32. [Link]
23. Speziale N. Does unemployment increase crime? Evidence from Italian provinces. *Applied Economics Letters*. 2014;21(15):1083-9. [Crossref]
24. Aaltonen M, Macdonald JM, Martikainen P, Kivivuori J. Examining the generality of the unemployment-crime association. *Criminology*. 2013;51(3):561-94. [Crossref]
25. Lastauskas P, Tatsi E. Spatial nexus in crime and unemployment in times of crisis: evidence from Germany. *Cambridge Working Papers in Economics*. 2013:1-50. [Crossref]
26. Utku A, Akçayol MA. Derin öğrenme tabanlı model ile bir olayın sonraki olma zamanının tahmini [Prediction of the next time of an event with deep learning based model]. *Politeknik Dergisi*. 2021;24(1):1-15. [Crossref]
27. Farjami Y, Abdi K. A genetic-fuzzy algorithm for spatio-temporal crime prediction. *Journal of Ambient Intelligence and Humanized Computing*. 2021:1-13. [Crossref]
28. Deniz Başar O, Güneren Genç E. Ülkelerin güvenli olmalarının tahmininde lojistik regresyon, yapay sinir ağları ve moora yöntemlerinin karşılaştırılması [A comparison of logistic regression, artificial neural networks and moora methods in estimation of the safety of countries]. *Journal of Life Economics*. 2020;7(2):123-34. [Crossref]
29. Rumi SK, Luong P, Salim FD. Crime rate prediction with region risk and movement patterns. *arxiv*. 2019. [Link]
30. Stalidis P, Semertzidis T, Daras P. Examining deep learning architectures for crime classification and prediction. *Forecasting*. 2021;3(4):741-62. [Crossref]
31. Wang B, Luo X, Zhang F, Yuan B, Bertozzi AL, Brantingham PJ. Graph-based deep modeling and real time forecasting of sparse spatio-temporal data. *arXiv preprint arXiv*. 2018. [Link]
32. Mohd F, Noor NMM. A comparative study to evaluate filtering methods for crime data feature selection. *Procedia Computer Science*. 2017;116:113-20. [Crossref]
33. Al Bataineh A. A comparative analysis of nonlinear machine learning algorithms for breast cancer detection. *International Journal of Machine Learning and Computing*. 2019;9(3):248-54. [Crossref]
34. Xing W, Bei Y. Medical health big data classification based on KNN classification algorithm. *IEEE Access*. 2019;8:28808-19. [Crossref]
35. Breiman L. Random forests. *Machine Learning*. 2001;45:5-32. [Crossref]