



POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)

URL: <http://dergipark.org.tr/politeknik>



Derin öğrenme tabanlı model ile bir olayın sonraki olma zamanının tahmini

Prediction of the next time of an event with deep learning based model

Yazar(lar) (Author(s)): Anıl UTKU¹, M. Ali AKCAYOL²

ORCID¹: 0000-0002-7240-8713

ORCID²: 0000-0002-6615-1237

Bu makaleye şu şekilde atıfta bulunabilirsiniz(To cite to this article): Utku A., Akcayol M. A., “Derin öğrenme tabanlı model ile bir olayın sonraki olma zamanının tahmini”, *Politeknik Dergisi*, 24(1):1-15, (2021).

Erişim linki (To link to this article): <http://dergipark.org.tr/politeknik/archive>

DOI: 10.2339/politeknik.620613

Derin Öğrenme Tabanlı Model ile Bir Olayın Sonraki Olma Zamanının Tahmini

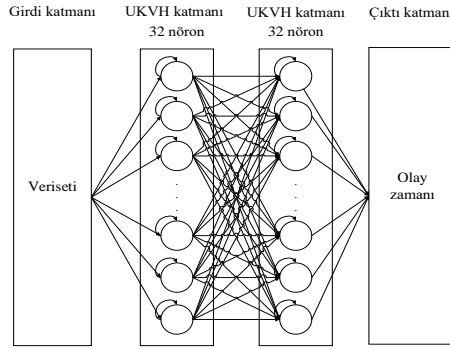
Prediction of the Next Time of an Event with Deep Learning Based Model

Önemli noktalar (Highlights)

- ❖ Derin öğrenme tabanlı tahmin modeli / Deep learning based prediction model.
- ❖ Suç ve acil çağrı analizine yönelik derin öğrenme tabanlı bir tahmin modeli geliştirilmiştir / Deep learning-based a prediction model has been developed for crime and emergency call analysis.
- ❖ Geliştirilen modelin deneysel sonuçları ARIMA ile kapsamlı olarak karşılaştırılmıştır / The experimental results of proposed model have been comprehensively compared with ARIMA.

Grafik Özet (Graphical Abstract)

Bu çalışmada, suç ve acil çağrı analizine yönelik derin öğrenme tabanlı bir tahmin modeli geliştirilmiştir./ In this study, a deep learning-based prediction model has been developed for crime and emergency call analysis.



Şekil 1. Geliştirilen UKVH tabanlı model / Figure 1. Developed LSTM based model

Amaç (Aim)

Bu çalışmanın temel amacı, geliştirilen hibrit derin öğrenme modelini kullanarak olay gerçekleşmeden önce bir sonraki olma zamanını tahmin etmektir. / The main purpose of this study is to predict the next event time before it occurs using the developed hybrid deep learning model.

Tasarım ve Yöntem (Design & Methodology)

Bu çalışmada, geliştirilen LSTM tabanlı derin öğrenme modeli ve ARIMA kapsamlı olarak karşılaştırılmıştır./In this study, the developed LSTM-based deep learning model and ARIMA are compared comprehensively.

Özgünlük (Originality)

Bu çalışma, olay zamanı tahminine yönelik yapılmış ilk çalışmadır./ This study is the first study on emergency call and crime time prediction.

Bulgular (Findings)

Deneysel sonuçlar, geliştirilen derin öğrenme tabanlı modelin ARIMA'dan daha başarılı olduğunu göstermiştir./ The experimental results showed that the developed deep learning-based model is more successful than ARIMA.

Sonuç (Conclusion)

Geliştirilen model ile geçmiş bilgilere dayanarak bir olayın bir sonraki olma zamanı başarılı bir şekilde tahmin edilmektedir./With the developed model, the next occurrence time of an event is successfully predicted based on past information.

Etik Standartların Beyanı (Declaration of Ethical Standards)

Bu makalenin yazarları çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler. / The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Derin Öğrenme Tabanlı Model ile Bir Olayın Sonraki Olma Zamanının Tahmini

Araştırma Makalesi / Research Article

Anıl UTKU*, M. Ali AKCAYOL

Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Gazi Üniversitesi, Türkiye

(Geliş/Received : 16.09.2019 ; Kabul/Accepted : 28.01.2020)

ÖZ

Bazı olayların olmadan önce olma zamanının tahmin edilebilmesine yönelik çalışmalar uzun yıllardır devam etmektedir. Böylelikle olay ortaya çıktığında meydana gelecek zararı minimuma indirmek veya elde edilecek faydayı maksimum yapmak amaçlanır. Bir müşterinin sonraki davranışının tahmini, doğal afetlerin sonraki olma zamanının tahmini, belirli bir zaman aralığında gelecek talep sayısının tahmini gibi çok farklı alanlarda sonraki olayların tahminine yönelik çalışmalar giderek artmaktadır. Ancak, literatürde suçların ve acil çağrılarının meydana gelmeden önce, sonraki olma zamanının ve olay türünün tahminine yönelik başarılı sonuç veren bir çalışma bulunmamaktadır. Suç analizi, tehdit altındaki bölgelerin güvenliğini sağlamayı, suç işleme oranını azaltmayı ve kolluk kuvveti gücünden tasarruf etmeyi amaçlayan bir araştırma alanıdır. Suçun işlendikten kısa süre sonra çözümlenmesine yönelik çalışmalarda güncel teknolojilerin kullanımıyla yüksek başarı elde edilmektedir. Benzer şekilde acil çağrı analizi ile çağrılara yanıt süresi kısaltılmakta ve kaynak kullanımı optimize edilmektedir. Bu çalışmada, suç ve acil çağrı analizine yönelik derin öğrenme tabanlı bir tahmin modeli geliştirilmiştir. Geliştirilen model ile suç işlenmeden önce ve acil çağrı gelmeden önce bir sonraki olma zamanı tahmin edilmektedir. Geliştirilen model ile elde edilen sonuçlar, istatistiksel zaman serisi tahminleme yöntemlerinden olan ARIMA ile kapsamlı bir şekilde karşılaştırılmıştır. Deneysel sonuçlar, geliştirilen derin öğrenme tabanlı modelin ileriye dönük zaman tahmininde ARIMA'dan daha başarılı olduğunu göstermiştir.

Anahtar Kelimeler: Suç analizi, acil çağrı analizi, zaman serisi analizi, derin öğrenme, ARIMA.

Prediction of the Next Time of an Event with Deep Learning Based Model

ABSTRACT

Studies have been going on for many years to predict the time before some events happen. Thus, it is aimed to minimize the damage that occurs when the event occurs or to maximize the benefit to be obtained. Studies on the prediction of subsequent events in many different areas, such as the prediction of the subsequent behavior of a customer, the prediction of the subsequent occurrence of natural disasters, the estimate of the number of future demands in a given time interval, are gradually increasing. However, in the literature, there is no successful study for predicting the time and type of event before the occurrence of crimes and emergency calls. Crime analysis is a field of research aimed at securing the threatened areas, reducing the rate of crime and saving law enforcement. High success is achieved with the use of up-to-date technologies in the efforts to resolve the crime shortly after it is committed. Similarly, emergency call analysis reduces response time and optimizes resource usage. In this study, a deep learning based prediction model for crime and emergency call analysis has been developed. With the developed model, the time of the next crime and the time of the next emergency call are predicted. The results obtained with the developed model has been compared with ARIMA which is one of the statistical time series prediction methods. Experimental results have shown that the developed deep learning-based model is more successful than ARIMA in forward-looking event time prediction.

Keywords: Crime analysis, emergency call analysis, time series analysis, deep learning, ARIMA.

1. GİRİŞ (INTRODUCTION)

Suç, yaşam kalitesini ve ekonomik büyümeyi etkileyen sosyoekonomik bir problemdir. Suçun özellikleri ve nasıl işlendiği, toplumun türüne bağlı olarak değişmektedir. Suç analizi konusunda yapılan araştırmalar eğitim, yoksulluk, işsizlik ve iklim gibi faktörlerin suç oranını etkilediğini göstermiştir [1].

Dinamik bir süreç olan suçun önlenmesi, başarılı bir toplumun kilit unsurlarından biridir. Ancak bu süreç insan faktöründen çok daha fazla parametre içerdiği için,

genel bir suç önleme mekanizması oluşturmak zordur [2]. Dolayısıyla, bir suçun işlenip işlenmediğini ve nerede meydana geleceğini tahmin etmek, kolluk kuvvetlerinin öne çıkan görevlerinden biridir. Bu tür bir tahmin, olası şüphelileri aramak zorunda kalmayacakları ve olay mahallinde yakalayabilecekleri için işlerini daha kolay bir hale getirecektir. Ayrıca, suçun gerçekleşmesini bile önlemek mümkün olabilecektir [3].

Suç verileri, uzun süreler boyunca, resmi kayıtlar ve yargılama süreçleri için kayıt altına alınmaktadır. Ancak buradaki verilerin odak noktası verilerin kendisinden ziyade insanların suç işleme eylemleri ve zamanıdır [4]. Bu sebeple uzun zaman boyunca suç eylemleri benzersiz ve rastgele olarak tanımlanmıştır. Bu nedenle suç tahmini

*Sorumlu Yazar (Corresponding Author)
e-posta : anilutku@gazi.edu.tr

rastgele olayların önceden tahmin edilmesinin imkânsız olması sebebiyle hiçbir anlam ifade etmemiştir. Bu ifade, farklı suçlar arasında genel bir örüntü veya bir bağlantı bulunmasının mümkün olmadığı anlamına gelmektedir [5]. Ancak, suçun eylemleri hakkında zaman, mekân, suç türü, suçluya ilişkin demografik bilgiler gibi daha fazla verinin toplanması ve analiz edilmesi suçların her durumda benzersiz ve rastgele olmadığını, verilerde örüntüler olduğunu göstermiştir [1].

Suç analizi, ortaya çıkması en muhtemel suçların tanımlanmasında, veri ve istatistiksel analiz yöntemlerini kullanan bir tekniktir [6]. Suç olayları ile ilgili toplanan veri miktarındaki artış nedeniyle bu verileri analiz etmek için yeni teknolojilere ihtiyaç duyulmaktadır. Günümüzde veri miktarı arttıkça, makine destekli bir yardıma olan ihtiyaç da artmıştır. Tahmine dayalı analitikler, genel olarak geçmiş verileri kullanarak belirli olayların gelecekteki sonuçlarını tahmin etmek için kullanılmaktadır.

911 acil çağrı hattı, Amerika'da acil durumlarda başvuru bir yardım merkezi numarasıdır. 911 acil çağrı hattı ekipleri, günlük olarak çok fazla miktarda yardım talebine cevap üretmektedir. Ancak hastalık ya da doğal afet gibi durumlarda, cevap verilmesi gereken çağrı sayısı katlanarak artmaktadır. Bu gibi durumlar, daha iyi kaynak kullanımı ve çağrılara yanıt süresinin kısaltılması açısından geleneksel 911 acil durum merkezlerinin daha yeni teknolojilerle güçlendirilmesini gerekli kılmaktadır. Geleneksel 911 merkezleri yalnızca sesli aramaları işlemektedir. Makine öğrenmesi ve veri madenciliği yöntemleri kullanılarak metin, video, ses ve görüntü gibi farklı türlerdeki veriler analiz edilebilirler. Bu gibi farklı veri türleri kullanılarak yapılan çıkarımlar, sahadaki acil müdahale ekiplerine aktarılabilir.

Bu çalışma kapsamında, zaman serisi analizi konusunda yapılmış literatürdeki çalışmalar incelenmiştir.

Cesario ve ark. tarafından 2016 yılında yapılan çalışmada [7], kentsel alanlardaki suç eğiliminin tahmin edilmesine yönelik bütünlük otoregresif hareketli ortalamalar (AutoRegressive Integrated Moving Average-ARIMA) modeli tabanlı bir yöntem geliştirilmiştir. Çalışmada veriseti olarak Chicago'da meydana gelmiş suç olaylarının tür, konum ve zaman gibi bilgilerini içeren, 2001 yılından beri haftalık olarak güncellenen veriler kullanılmıştır. Model parametrelerinin belirlenmesi için akaike ölçütü (Akaike Information Criterion-AIC) kullanılarak yapılan deneysel çalışmalar sonucunda ARIMA(1,1,1) modeli seçilmiştir. Deneysel çalışmalar, geliştirilen modelin suç tahmininde literatürdeki çalışmalara göre daha başarılı sonuçlar verdiğini göstermiştir.

Fu ve ark. tarafından 2016 yılında yapılan çalışmada [8], ARMA (AutoRegressive Moving Average) ve ARIMA gibi doğrusal modellerin, araç trafik akışının rastsal ve doğrusal olmayan doğasına uyum sağlayamaması sebebiyle derin öğrenme tabanlı bir sistem geliştirilmiştir. Geliştirilen sistemde, kısa süreli trafik akışını öngörmek için Uzun-Kısa Vadeli Hafıza (UKVH)

ve kapı tekrarlı birimler (Gated Recurrent Units-GRU) sinir ağı yöntemleri kullanılmıştır. Çalışma kapsamında, Kaliforniya'da konuşlandırılmış 15.000'den fazla sensörün 30 saniyede bir gönderdiği veriden oluşan PeMS veriseti kullanılmıştır. Deneysel çalışmalar, derin öğrenme tabanlı UKVH ve GRU yöntemlerinin, ARIMA modelinden daha iyi performans gösterdiğini ortaya koymuştur.

Kang ve Kang tarafından 2017 yılında yapılan çalışmada [9], Derin Sinir Ağı (Deep Neural Network-DNN) tabanlı bir sistem geliştirilmiştir. Çalışma kapsamında kullanılan veriseti, Chicago'daki suç istatistikleri, demografik ve meteorolojik bilgiler ile görüntü veritabanlarından toplanan verilerden oluşmaktadır. Çalışmada eğitim verileri oluşturmadan önce, istatistiksel analizler yapılarak suçla ilgili veriler seçilmiştir. DNN modeli mekân, zaman, çevresel faktörler ve ortak özellikler kullanılarak eğitilmiştir. Çevresel bağlam bilgisinin çıkarılması için görüntü verileri kullanılmıştır. Deneysel sonuçlar, DNN modelinin suç oluşumunu öngörmeye diğer tahmin modellerinden daha doğru sonuçlar verdiğini göstermiştir.

Zheng ve ark. tarafından 2017 yılında yapılan çalışmada [10], akıllı şebekelerdeki elektrik yükü tahmininde, zaman serilerinin doğrusal ve durağan olmayan yapıları nedeniyle geleceğe dönük öngörüler yapmanın zor olması sebebiyle UKVH tabanlı yeni bir sistem geliştirilmiştir. UKVH, daha doğru tahminler elde etmek için zaman serilerindeki uzun vadeli bağımlılıkları kullanmaktadır. Yapılan çalışmada destek vektör makinesi (Support Vector Machine-SVM), otoregresif sinir ağı (Neural Network AutoRegression-NNETAR), UKVH, doğrusal olmayan otoregresif eksojen (Nonlinear autoregressive exogenous-NARX) ve ARIMA modeli kullanılmıştır. Yapılan deneysel çalışmalar UKVH'nın daha başarılı sonuçlar verdiğini ortaya koymuştur.

Kouziokas tarafından 2017 yılında yapılan çalışmada [11], yüksek suç riski taşıyan toplu ulaşım alanları tahmin etmek için mekânsal kümelenme yöntemleri ve yapay sinir ağı modelleri kullanılmıştır. Mekânsal analizler için coğrafi bilgi sistemleri kullanılmıştır. Artan sayıda suç olayına sahip alanları bulmak için komşuluk analizi kullanılmıştır. Deneysel sonuçlar geliştirilen sistemin klasik kümeleme yöntemlerine göre daha başarılı sonuçlar verdiğini göstermektedir.

Lin ve ark. Tarafından 2017 yılında yapılan çalışmada [12], sinir ağı tabanlı hibrit bir öğrenme yaklaşımı olan TreNet modeli sunulmuştur. TreNet modeli, Evrimsel Sinir Ağları (ESA) ve UKVH modellerinin hibrit bir uygulamasıdır. TreNet özellik çıkarma aşamasında ESA'nı, sıralı bağımlılıkları belirlemek için ise UKVH'yı kullanır. TreNet modeli ESA, UKVH ve Gizli Markov Modeli yaklaşımları ile gerçek verisetleri üzerinde karşılaştırmalı olarak test edilmiştir. Test sonuçları geliştirilen hibrit yöntemin daha başarılı sonuçlar verdiğini göstermiştir.

McNally ve ark. tarafından 2018 yılında yapılan çalışmada [13], Amerikan doları cinsinden Bitcoin fiyatının öngörülmesine yönelik ARIMA, Tekrarlı Sinir Ağları (TSA) ve UKVH yöntemlerinin karşılaştırmalı bir uygulaması sunulmuştur. Geliştirilen sistemde veriseti olarak Bitcoin fiyat endeksleri kullanılmıştır. Deneysel çalışmalar, grafik işlemci üzerindeki eğitim süresi ve yapılan öngörülerin doğruluğu açılarından karşılaştırmalı olarak yapılmıştır. Deneysel sonuçlar, UKVH ve TSA'nın ARIMA'dan daha yüksek tahmin oranına sahip olduğunu göstermiştir.

Tian ve ark. tarafından 2018 yılında yapılan çalışmada [14], tahmin doğruluğunu iyileştirmek için ESA ve UKVH'yı birleştiren yeni bir derin sinir ağı modeli önerilmiştir. Veriseti olarak Transsoency Platformunun sağladığı İtalya-Kuzey Bölgesi'ndeki elektrik yükü verileri kullanılmıştır. Geliştirilen model, Karar Ağacı, Rastgele Orman, Derin Enerji modeli, UKVH ve ESA ile karşılaştırılmalı olarak incelenmiştir. Deneysel sonuçlar ortalama mutlak hata (Mean Absolute Error-MAE), ortalama mutlak yüzde hata (Mean Absolute Percentage Error-MAPE) ve kare ortalamalarının karekökü (Root Mean Square Error-RMSE) metrikleri açısından geliştirilen modelin daha başarılı sonuçlar verdiğini göstermiştir.

Stec ve Klabjan tarafından 2018 yılında yapılan çalışmada [15], suç tahminine yönelik derin öğrenme tabanlı bir sistem geliştirilmiştir. Suç verilerine ek olarak hava durumu, nüfus verileri ve toplu taşıma verilerinin bulunduğu Chicago ve Portland suç verileri, veriseti olarak kullanılmıştır. Suç verileri 10 parçaya ayrılarak her bir bölge için günlük seviyede tahminler gerçekleştirilmiştir. Veriseti, suç tahmin probleminin mekansal ve zamansal yönleri ele alınarak ileri beslemeli ağlar (Feed Forward Network-FFN), ESA ve TSA kullanılarak karşılaştırmalı olarak analiz edilmiştir. Deneysel çalışmalar TSA+ESA hibrit modelinin Chicago ve Portland verisetlerinin her biri için ESA ve TSA'na göre daha başarılı sonuçlar verdiğini göstermiştir.

İncelenen literatürdeki çalışmalarda, zaman serisi analizlerinin elektrik tüketimi tahmini, rüzgâr hızı tahmini, trafik yükü tahmini, elektrikli araçların şarj süresi tahmini, motorlardaki yağ tüketim verilerine göre performans tahmini, borsa tahmini, akıllı evlerdeki cihazların kullanım verilerine göre kullanıcı davranış tahmini ve kentsel alanlardaki suç eğiliminin tahmini gibi uygulama alanlarında kullanıldığı görülmüştür. İncelenen çalışmalarda ARMA ve ARIMA gibi istatistiksel yöntemlerin dışında makine öğrenmesi ve derin öğrenme modellerinin kullanıldığı görülmüştür. İncelenen literatürdeki çalışmalarda, olayların meydana gelme zamanlarını tahmin etmeye yönelik geliştirilmiş bir uygulama olmadığı görülmüştür. Olay zamanı tahmini, Web tabanlı saldırılar, belirli bölgelerde yaşanan adli durumlar ya da e-ticaret kullanıcı davranış analizleri gibi uygulama alanlarında kritik bir öneme sahip olmaktadır. Bu sayede olaylar meydana gelmeden önce olma zamanları tahmin edilerek oluşabilecek riskler azaltılabilir. Diğer bir taraftan e-ticaret uygulamaları

açısından bakıldığında ise kullanıcılara zamanında ve kişiselleştirilmiş tavsiyeler sunularak elde edilecek fayda maksimize edilebilir.

Bu çalışma kapsamında, olay zamanlarını tahmin etmeye yönelik bir model geliştirilmiştir. Olay zamanı tahmini ile olayların ortaya çıktığında oluşturacağı zararın minimuma indirilmesini, elde edilecek faydanın maksimuma çıkarılması amaçlanmaktadır. Zaman serisi analizi literatürde kullanıcı davranış analizi, talep tahmini, Web tabanlı saldırı tahmini gibi farklı uygulama alanları için kullanılmaktadır. Ancak olay zamanı tahmini konusunda yapılmış bir çalışma literatürde bulunmamaktadır. Bu çalışmada, suçların ve acil çağruların meydana gelmeden önce, sonraki olma zamanının ve olay türünün tahmin edilmesi amaçlanmıştır. Tahmin doğruluğunu ve etkinliğini arttırmak için UKVH tabanlı bir derin öğrenme modeli geliştirilmiştir. Geliştirilen model ile suç işlenmeden önce ve acil çağrı gelmeden önce bu olayların bir sonraki olma zamanı tahmin edilmektedir. Geliştirilen model ile elde edilen sonuçlar, istatistiksel zaman serisi tahminleme yöntemlerinden olan ARIMA ile kapsamlı bir şekilde karşılaştırılmıştır. Deneysel sonuçlar, geliştirilen derin öğrenme tabanlı modelin ileriye dönük zaman tahmininde ARIMA'dan daha başarılı olduğunu göstermiştir.

2. SUÇ VE ACİL ÇAĞRI VERİSİ ANALİZİ (CRIME AND EMERGENCY CALL DATA ANALYSIS)

Günümüzde, dünya tarihindeki en büyük kentsel büyüme dalgası yaşanmaktadır. Birleşmiş Milletler raporuna göre, kentsel nüfusun 2030 yılında 4,98 milyara çıkması beklenmektedir [16]. Bu sayı, şehirlerde yaşayan mevcut nüfusun kabaca % 60'ına karşılık gelmektedir. Giderek artan bu kentleşme durumu, büyük oranda sosyal, ekonomik ve çevresel dönüşümleri zorunlu kılmaktadır. Aynı zamanda su ve elektrik gibi kaynak yönetimi, trafik, hava ve su kalitesi, kamu politikaları ve kamu güvenliği gibi hizmetler açısından şehir yönetimini zorlaştırmaktadır.

Şehirler büyüdükçe, kamu güvenliği sorunları da ön plana çıkmaktadır. Kolluk kuvvetlerinin karşılaştığı en büyük zorluk, suç sayısının artması ve suçla ilgili öngörülebilir bulunabilecek benzer veri kaynaklarının çeşitliliğinin doğru ve verimli bir şekilde analiz edilmesidir [17]. Bu durum, özellikle kolluk kuvvetlerinin büyük suç verilerinden faydalı örüntüleri belirleyecek araç ve teknolojilerin bulunmadığı büyük kentsel alanlarda söz konusudur.

Benzer şekilde giderek artan insan nüfusuna paralel olarak duyulan acil yardım taleplerinde de artış yaşanmaktadır. Trafik kazaları, yaralanmalar, kalp krizi ya da yangın gibi durumlarda acil yardım ihtiyacı ortaya çıkmaktadır. Bu gibi durumlarda sunulacak acil yardım hizmetinin kalitesi ve hızı ön plana çıkmaktadır. Kaynak yönetimi ve zaman planlaması can ve mal güvenliği açısından hayati önem taşımaktadır.

Suç verisi analizinde, kolluk kuvvetleri için en önemli konulardan biri uzun ve kısa vadeli suç tahminlerine doğru bir şekilde sahip olmaktır [18]. Özellikle, devriye bölgeleri gibi küçük coğrafi bölgelerdeki alanlarda kısa vadeli suç tahminleri etkili olacaktır. Benzer şekilde, uzun vadeli suç eğilimlerinin tahmini ise şehirlerdeki güvenlik durumunun anlaşılmasında ve güvenlik politikalarının düzenlenmesinde etkili olmaktadır [19].

Suç olayları sistematik ya da rastgele bir yapıya sahip değildir. Suç analizleri, tehdit altındaki bölgelerin güven altına alınması, suç eylemlerinin gerçekleşmeden belirlenmesi ve kolluk kuvvetleri açısından iş yükünün azaltılmasını sağlar [2]. Bu sebeple suç olaylarının daha iyi anlaşılması için mekânsal ve zamansal örüntülerin belirlenmesi ve ilişkilerin analiz edilmesi önemlidir [3]. Bilgisayarlara insan müdahalesi olmadan karar yetisi kazandırma bilimi olarak tanımlanabilen makine öğrenmesi ve derin öğrenme suç analizlerinde yaygın olarak kullanılmaktadır [4]. Derin öğrenme yöntemleri veri toplama, sınıflandırma, örüntü tanıma, tahmin ve görselleştirme gibi uygulama alanlarında suç analizi için kullanılmaktadır [9].

Suç verisi miktarındaki artış, özellik sayısının da büyük oranda artmasını sağlamıştır. Bu sayede, çok sayıda özellik ele alınarak özellikler arasındaki karmaşık ilişkilerin belirlenmesi daha uygun olacaktır. Derin öğrenme, suç analizi problemlerinin zamansal ve mekânsal bileşenlerinin yanı sıra, suç verisetlerine ek olarak ilgili kamu verisetleri kullanılarak oluşturulabilecek geniş bir özellik setinin ele alınması için de uygundur [3]. Suç analizi probleminin doğal karmaşıklığı, bu problemi derin öğrenme için uygun hale getirmektedir [4].

Acil çağrı verilerini analiz ederek verilerdeki örüntülerin çıkarılması, acil durum müdahale ekiplerinin olaylarla başa çıkma konusunda daha donanımlı olmasını sağlayacaktır. Trafik kazaları ya da yangın gibi durumlar göz önüne alındığında, bu olaylardaki artış aynı zamanda ambulans hizmetlerinde olan taleplerde de artışa neden olacaktır [20]. Örneğin trafik kazaları, yol koşullarının iyileştirilmesi ile engellenebilir. Benzer şekilde solunum problemleri neticesinde ortaya çıkabilecek acil durumlar, bölgenin havasındaki kirlilikten kaynaklanıyor olabilir. Acil çağrı verileri analiz edilerek olaylar hakkında çeşitli çıkarımlar ve öngörüler yapılabilir.

Suç analizi ve acil çağrı verisi analizinde derin öğrenme yöntemlerinin kullanılması, özellik mühendisliğine duyulan ihtiyacı azaltarak büyük veri kümeleri üzerinde eğitim işleminin gerçekleştirilmesini sağlar. Geleneksel bir sinir ağında, tüm girdi ve çıktılar birbirinden bağımsız olduğu varsayımı geçerlidir ancak suç analizi için durum böyle değildir [9]. Dün olan suçlar bugün olan suçları etkileyecektir. Suç tahmininin zamansal yönlerini ele almak için TSA ya da UKVH kullanılmaktadır. Örneğin, bir bölgedeki hırsızlıklarda bir artış varsa, o bölgedeki hırsızlık oranlarının belirli bir süre boyunca normalden daha yüksek olması mümkündür.

3. ZAMAN SERİSİ ANALİZİ (TIME SERIES ANALYSIS)

Zaman serileri, ardışık zaman aralıklarında ölçülen veri noktası dizileridir. Matematiksel olarak, t geçen süreyi ifade etmek üzere y_t , $t = 0, 1, 2, \dots, T$ vektörlerinin kümesi olarak tanımlanmaktadır. Bir zaman serisi, her bir y değişkeninin, t zamanındaki değerini temsil eden y_t değerleri tarafından oluşturulmaktadır. Bu nedenle öngörü işlemi, $y_{t+1} + (y_{t+1})$ değerinin belirlenmesi ve $y_{t+1} - (y_{t+1})$ işlevi olarak temsil edilen hata oranını en aza indirmeyi hedeflemektedir. y_t değişkeni, rasgele değişken olarak kabul edilmektedir. Bir zaman serisinde, bir olay sırasında alınan ölçümler, belirli bir zamansal düzene uygun olarak sıralanmıştır [20].

Zaman serisi analizi, uygun bir model kullanılarak belirli bir zaman serisinin bilinen veri değerlerinden, karşılık gelen gelecekteki değerlerin belirlenmesidir [21]. Zaman serisi analizi, temel olarak verinin doğasını anlamaya çalışan ve geleceğe dönük öngörüler ve simülasyonlar için kullanılabilir yöntemleri içermektedir [22].

Zaman serisi analizi ve dinamik modelleme, işletme, ekonomi, finans ve bilgisayar bilimleri gibi birçok uygulama alanı bulunan bir araştırma konusudur [23]. Zaman serisi analizinin amacı, zaman serilerindeki gözlemlenen veri değerlerini incelemek ve tanımlanan veri yapılarıyla zaman serilerinin gelecekteki değerlerini tahmin etmek için bir model oluşturmaktır [24]. Zaman serilerinde tahmin doğruluğunun iyileştirilmesi, temel olarak oluşturulan modelin etkinliğine bağlıdır [25].

Zaman serisi analizlerinde yaygın olarak istatistik tabanlı ARIMA modelleri ve makine öğrenmesi tabanlı yöntemler kullanılmaktadır. ARIMA modelleri ekonomi ve finans verilerinin modellenmesinde yaygın olarak kullanılıyor olsa da, bazı önemli kısıtlamaları vardır [26]. Örneğin, ARIMA modelinde, değişkenler arasındaki doğrusal olmayan ilişkileri modellemek zordur. Ayrıca, ARIMA modelinde hataların sabit bir standart sapmaya sahip olduğu varsayılmaktadır. ARIMA gibi doğrusal modellerle ilgili zorlukları aşabilmek için zaman serisi analizinde derin öğrenme tabanlı yeni teknikler geliştirilmiştir [27].

3.1. Doğrusal Yöntemler (Linear Methods)

Zaman serileri, doğası gereği durağan olmadığı için gelecekte ne olacağı konusunda kesin bir tahminde bulunulamamaktadır [28]. Genel olarak $\{y_t, t = 0, 1, 2, \dots\}$ şeklindeki zaman serilerinin rastgele

bir y_t değişkeninin birleşik olasılık dağılımını izlediği varsayılmaktadır [29]. Bir zaman serisinin olasılık yapısını tanımlayan matematiksel ifade, olasılıksal bir süreç olarak adlandırılmaktadır [30]. Dolayısıyla, serinin gözlem dizisi aslında onu üreten olasılıksal sürecin bir örneklemdir [31]. Zaman serilerindeki genel varsayım, serideki y_t değişkenlerinin, normal dağılımı takiben

bağımsız ve özdeş olarak dağıtılmış olmasıdır. Ancak zaman serileri, uzun vadede daha düzenli bir örüntü izlerler [32]. Örneğin, belirli bir şehrin bugünkü sıcaklığı çok yüksekse, yarınki sıcaklığının da muhtemelen yüksek olacağı tahmin edilebilir. Bu durum zaman serilerinde durağanlık kavramı ile ilişkilidir [33].

Zaman serilerinde, serinin alt yapısını yansıttığı için uygun modelin seçilmesi, geleceğe yönelik yapılacak öngörülerin doğruluğunu birinci dereceden etkilemektedir [34]. Genel olarak, zaman serilerinde kullanılacak modeller farklı olasılıksal süreçleri içerebilmektedir [35]. Literatürde yaygın olarak kullanılan doğrusal zaman serisi modelleri AR ve MA modelleridir. Ayrıca bu iki modeli birleştiren ARMA VE ARIMA modelleri kullanılmaktadır [36].

ARMA Model

Tek değişkenli zaman serilerinin modellenmesi için kullanılan ARMA(p , q) modeli, AR(p) ve MA(q) modellerinin bir birleşimidir [26]. AR(p) modelinde, bir değişkenin gelecekteki değerinin, geçmiş p adet gözlemin doğrusal bir birleşimi olduğu ve sabit bir hata terimine sahip olduğu varsayılmaktadır. Matematiksel olarak AR(p) modeli Eş.1 kullanılarak ifade edilmektedir [39].

$$y_t = c + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (1)$$

Burada y_t ve ε_t sırasıyla t zamandaki gerçek değeri ve rastgele hatayı ifade etmektedir. φ_i ($i = 1, 2, \dots, p$) model parametrelerini ve c ise model sabitini ifade etmektedir. Tamsayı sabiti p , modelin sırası olarak bilinir ve ihmal edilebilmektedir.

MA(q) modeli ise Eş.2'de görüldüğü gibi açıklayıcı değişken olarak geçmiş hataları kullanmaktadır [13].

$$y_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Burada μ zaman serisinin ortalamasını, θ_j ($j = 1, 2, \dots, q$) ise model parametrelerini ifade etmektedir. Rastgele hatanın, 0 ortalamaya sahip rastgele dağıtılmış ve normal dağılıma sahip değişkenler dizisi olduğu varsayılmaktadır. AR ve MA modelleri, ARMA modeli olarak bilinen genel ve kullanışlı bir zaman serisi modeli oluşturmak için etkin bir şekilde birleştirilebilir. Matematiksel olarak bir ARMA (p , q) modeli, Eş.3 kullanılarak ifade edilmektedir [37].

$$y_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (3)$$

Modeldeki p , q parametrelerinden p parametresi, otoregresifliği ve q parametresi ise hareketli ortalamayı ifade etmektedir.

ARMA modelleri yalnızca sabit zaman serisi verileri için kullanılabilen modellerdir. Ancak e-ticaret uygulamaları

gibi gerçek dünya uygulamalarındaki zaman serileri durağan olmayan davranış göstermektedir. Bu sebeple ARMA modelleri, uygulamalarda sıklıkla karşılaşılan durağan olmayan zaman serilerini doğru bir şekilde tanımlamakta yetersiz kalmaktadır. ARIMA modeli, ARMA modelinin durağan olmayan durumları da içerecek şekilde güncellenmiş bir halidir [13]. ARIMA modeli, ARMA modelini kullanarak eşit aralıklı tek değişkenli zaman serisi verilerini analiz eden doğrusal bir modeldir. ARIMA modeli, zaman serisindeki geçmiş değerlerin ve geçmiş hataların doğrusal bir birleşimine göre öngörü sürecini işletir [32]. Durağan bir zaman serisi için ARIMA modeli, bağımlı değişkenin gecikme sürelerinden ve tahmin hatalarının gecikmesinden oluşan doğrusal bir modeldir. Modelde tahmin edilecek y değeri, y değişkeninin son değerleri ile hataların ağırlıklı bir toplamıdır [13].

ARIMA Model

ARIMA, otoregresif AR(p) ve hareketli ortalama MA(q) süreçlerini birleştirerek zaman serilerinde birleşik süreç oluşturan bir modeldir. ARIMA modeli p , d ve q parametrelerine sahiptir. ARIMA modelindeki otoregresif AR(p), bir gözlem ile bir dizi gecikmeli gözlem arasındaki bağımlılıkları kullanan bir regresyon modelidir [25]. Entegrasyon, farklı zamanlardaki gözlem farklılıklarını ölçerek zaman serilerini durağan hale getirmeyi ifade etmektedir (d). ARIMA modelindeki hareketli ortalama MA(q) ise gözlemler arasındaki bağımlılığı ve kalan hata terimlerini dikkate alan bir yaklaşımdır (q).

ARIMA modeli, Box ve Jenkins tarafından 1976 yılında tanımlanan bir metodolojiyi izlemektedir [29]. Box ve Jenkins metodolojisine göre ARIMA, modelin tanımlanması, parametrelerin tahmini, modelin doğrulanması ve öngörü aşamalarından oluşmaktadır.

Modelin tanımlanması, zaman serisinin durağan olup olmadığının yani olasılıksal bir sürecin ortalama ve standart sapmasının zamanla değişip değişmediğinin belirlenmesidir. Zaman serileri bu kıstası karşılamıyorsa, dönüşüm uygulanarak serinin durağanlığa ulaşana kadar farklılaştırılması gerekmektedir. Serinin farklılaştırılması gereken zaman sayısı d ile gösterilir ve ARIMA modellerinde belirlenecek parametrelerden biridir [33].

Parametre tahmini, p ve q parametrelerinin en küçük kareler minimizasyonu ya da maksimum olasılık algoritmaları kullanılarak tahmin edilmesini ifade etmektedir. En küçük kareler minimizasyonu eğitim seti tahminindeki karesel hatayı minimuma indirirken maksimum olasılık algoritması ise benzerlik fonksiyonunu en üst düzeye çıkarmayı hedeflemektedir [13].

Model doğrulanması, oluşturulan modelin uygun olup olmadığının belirlenmesidir. Veri bütünüyle rastsal ise beyaz gürültüye sahip olduğu kabul edilerek model onaylanmaktadır [25]. Farklı modellerin performanslarını değerlendirmek için AIC, MAE veya RMSE gibi metrikler kullanılabilir. Öngörü

aşaması ise parametreleri doğru bir şekilde belirlenmiş ve onaylanmış modelin tahminleri yapmaya hazır olduğunu ifade etmektedir. ARIMA modeli ile tahminler Eş.4 kullanılarak hesaplanmaktadır.

$$y_t = \mu + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (4)$$

Burada \hat{y}_t öngörü değerini, e_t t zamanındaki hatayı, φ_i ve θ_j ise katsayıları ifade etmektedir. μ , otoregresif AR(p) ve hareketli ortalama MA(q) polinomlarının sayısı olan beyaz gürültüyü ifade etmektedir. ARIMA modelinin AR bölümü, y_t değişkenin önceki değerlerinden elde edildiğini, MA bölümü ise regresyon hatasının önceki hataların doğrusal bir birleşimi olduğunu göstermektedir. I bölümü, normal veri değerlerinin, veriler ve önceki değerler arasındaki farkla değiştirildiğini gösterir [13].

3.2. Derin Öğrenme Tabanlı Yöntemler (Deep Learning Based Methods)

Zaman serileri, borsadaki fiyat değişimleri, hastalıkların yayılma süreçleri ya da ses sinyallerindeki değişimlerin izlenmesi gibi birçok gerçek dünya uygulaması kullanılarak elde edilebilmektedir. Farklı uygulama alanları için kullanılan uygulamalar verilerin niteliğine ve amacına bağlı olarak değişmektedir [38]. Örneğin, doktorlar hastanın uyku düzenindeki anormallikleri belirlemekle ilgilenirken, ekonomistler faiz oranının gelecek değerini belirlemekle ilgilenmektedirler. Bu tür problemler sınıflandırma, tahmin ya da anomali tespiti problemleri olarak incelenmektedir [34].

Yapay Sinir Ağları (YSA), biyolojik sinir ağlarının yapı ve işlevlerine dayanan bir hesaplama modelidir [35]. YSA, girdiler ve çıktılar arasındaki karmaşık ilişkilerin modellendiği ya da örüntülerin bulunduğu doğrusal olmayan istatistiksel bir veri modelleme aracı olarak kabul edilmektedir. YSA, günümüzde robotik işleme, nesne tanıma, konuşma ve el yazısı tanıma gibi çeşitli uygulama alanlarında kullanılmaktadır [36].

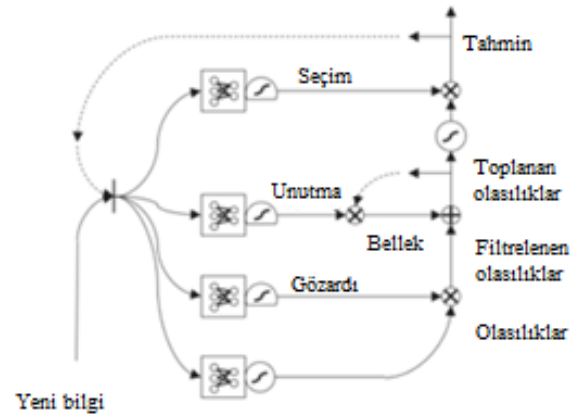
UKVH

Yapay sinir ağları, temel olarak giriş verileri kullanılarak çıkış verilerinin hesaplanmasını sağlamaktadır. Elde edilen çıkış değerlerinin, gerçek veriler ile karşılaştırılması sonucunda geriye yayılım yöntemi ile ağırlık değerleri üzerinde güncellemeler yapılmaktadır [37]. Ağ yapısının çok karmaşık olması durumunda ağırlık değerleri üzerinde güncellemeler yapılmaz ve ağırlık eğitimi sonlanmış olur. Bu problem, ağırdaki önceki katmanların parametrelerini öğrenmeyi ve ayarlamayı zorlaştırmaktadır [33]. UKVH, tekrarlı sinir ağı yapılarındaki zaman içinde geriye bağlılık sorunlarına çözüm olarak geliştirilmiş bir sinir ağı mimarisidir [37].

UKVH, zaman içinde geriye yayılım kullanılarak eğitilen ve yok olan eğim problemini (vanishing gradient problem) ortadan kaldıran tekrarlı bir sinir ağıdır. Yok

olan eğim problemi, yapay sinir ağlarının geriye yayılım temelli yöntemlerde eğitim aşamasında yaşanan bir zorluktur [38]. UKVH, nöronlar yerine katmanlara bağlı bellek bloklarına sahiptir. Bellek bloklarının bileşenleri, blokları klasik bir nörondan daha akıllı hale getiren, bloğun durumunu ve çıkışını yöneten kapılar içerir. Bir blok, bir giriş dizisi üzerinde çalışır ve blok içindeki her bir kapı, tetiklenip tetiklenmediklerini kontrol etmek için sigmoid aktivasyon ünitelerini kullanır. Sigmoid aktivasyon üniteleri de durumun değişmesini ve blok koşulundan akan bilginin eklenmesini sağlar [34].

UKVH, veri dizisini öğrenmek için ek özelliklere sahip bir tür TSA'dır. Verilerin daha önceki eğiliminin öğrenilmesi, Şekil 1'de görüldüğü gibi UKVH'ya dâhil edilen bir bellek hattı ve kapılar aracılığıyla mümkündür.



Şekil 1. Temel UKVH yapısı (Basic UKVH structure)

UKVH, veri akışlarının elde edildiği ve saklandığı bir dizi hücre veya sistem modülüdür. Hücreler, bir modülden diğerine verileri ileten ve toplayan bir nakil hattına benzemektedir. Her hücrede bazı kapıların kullanılması nedeniyle hücrelerdeki veriler, bir sonraki hücreler için atılabilir, filtrelenebilir veya eklenebilir. Bu nedenle, sigmoidal sinir ağı katmanına dayanan kapılar, hücrelerin isteğine bağlı olarak verilerin geçmesini veya engellenmesini sağlar [37]. Her bir sigmoid katmanı, her hücrede geçirilmesine izin verilen veri bölümünün miktarını gösteren, 0 ve 1 sayıları çıkarır. 0 değeri, hiçbir şeyin geçmesine izin verilmemesi, 1 değeri ise her şeyin geçmesine izin verilmesi anlamına gelmektedir [31].

Sigmoid aktivasyon ünitelerinde her bir hücrenin durumunu kontrol etmek amacıyla giriş kapısı, çıkış kapısı ve unutma kapısı olmak üzere üç tip kapı bulunmaktadır. Giriş kapısı, girişteki hangi değerlerin bellek durumunu güncelleyeceğine koşullu olarak karar verir. Çıkış kapısı, giriş ve bloğun belleğine bağlı olarak neyin çıkış yapılacağına koşullu olarak karar verir. Unutma kapısı, şartlı olarak hangi bilgilerin bloktan atılacağına karar verir. Burada her bir ünite, ünite kapılarının eğitim sırasında öğrenilen ağırlıklara sahip olduğu bir durum makinesi gibidir [33].

UKVH temel olarak girdi, çıktı, unutma kapıları, hafıza hücreleri ve odaklanma mekanizmasından oluşmaktadır.

Unutma kapıları, modelin ne zaman hatırlayıp ne zaman unutacağına öğrenmesini sağlayan yapılardır. Hafıza hücreleri ise yeni bir örüntünün öğrenilmeye ya da depolanmaya değer olup olmadığına karar veren yapılardır. Bu sayede yeni veriler geldikçe model ihtiyacı olan verileri saklayacak, ihtiyaç duymadıklarını ise uzun vadeli hafızasından silecektir. Odaklanma mekanizması ise uzun vadeli hafızanın kullanım durumu ve işlem hafızası üzerinde karar vermektir sorumludur.

Temel olarak t zamanında alınan bir x_t girdisi için uzun vadeli hafıza C_{t-1} ve işlem hafızası h_{t-1} kullanılarak bir önceki zaman adımından t zamanına güncelleme yapılmak istenmektedir. 0 ile 1 arasında olan n adet hatırlatma geçidi vasıtasıyla n adet ögenin hafızada ne kadar uzun vadeli olarak saklanacağına karar verilir. 0 değeri unutmayı, 1 ise öğeyi saklamayı ifade etmektedir. UKVH'nın kapı ve hücrelerine ait hesaplama fonksiyonları Eş.5-Eş.9'da görülmektedir.

$$i_t = \sigma(w_{xi} x_t + w_{hi} h_{t-1} + w_{ci} c_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma(w_{xf} x_t + w_{hf} h_{t-1} + w_{cf} c_{t-1} + b_f) \quad (6)$$

$$c_t = f_t \square c_{t-1} + i_t \square \tanh(w_{xc} x_t + w_{hc} h_{t-1} + b_c) \quad (7)$$

$$o_t = \sigma(w_{xo} x_t + w_{ho} h_{t-1} + w_{co} c_{t-1} + b_o) \quad (8)$$

$$h_t = o_t \square h(c_t) \quad (9)$$

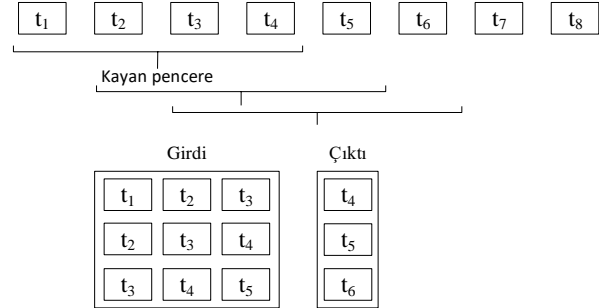
Burada i giriş kapısını, o çıkış kapısını, f unutma kapısını, c hücre aktivasyon vektörünü, W ağırlık matrisini ve b ise ön yargı vektörünü ifade etmektedir. \odot ise, iki vektörün skaler çarpımını göstermektedir. x_t ve h_t giriş-çıkış dizilerini ifade etmektedir. g_t , giriş dizisini ve önceki durum ile birlikte şimdiki duruma dönüştürmeyi sağlar. c_t ise hücre durumunun güncellenmesini sağlamaktadır.

4. GELİŞTİRİLEN DERİN ÖĞRENME MODELİ (DEVELOPED DEEP LEARNING MODEL)

Zaman serisi analizlerinde yaygın olarak istatistik tabanlı ARIMA modelleri ve makine öğrenmesi tabanlı yöntemler kullanılmaktadır. ARIMA modelleri ekonomi ve finans verilerinin modellenmesinde yaygın olarak kullanılıyor olsa da, bazı önemli kısıtlamaları vardır. Örneğin, ARIMA modelinde, değişkenler arasındaki doğrusal olmayan ilişkileri modellemek zordur. Ayrıca, ARIMA modelinde hataların sabit bir standart sapmaya sahip olduğu varsayılmaktadır. ARIMA gibi doğrusal modellerle ilgili zorlukları aşabilmek için zaman serisi analizinde derin öğrenme tabanlı yeni teknikler geliştirilmiştir.

Zaman serileri, bir zaman dizini tarafından sıralanan sayı dizilerini ifade etmektedir. Denetimli öğrenme

problemleri ise bir algoritmanın, girdi değişkenlerinden (x) çıktı değişkenlerinin (y) nasıl tahmin edileceğini öğrenebileceği problemlerdir. $y=f(x)$ şeklinde kullanılan bir fonksiyon ile giriş verilerinden çıkış verilerinin tahmin edilmesi amaçlanmaktadır. Zaman serisi verileri, Şekil 2'de görüldüğü gibi giriş değişkeni olarak belirlenen zaman adımlarından bir sonraki zaman adımındaki çıkış değişkeninin belirlenmesi şeklinde denetimli öğrenme problemine dönüştürülebilir.



Şekil 2. Zaman serisi verilerinin denetimli öğrenme verilerine dönüştürülmesi (Converting the time series data to supervised learning data)

Örneğin, Şekil 2'de görülmekte olan zaman serisi verileri, bir sonraki zaman adımındaki değeri tahmin etmek için önceki zaman adımındaki değerler kullanılarak denetimli öğrenme problemi şeklinde yapılandırılabilir. Bir önceki zaman adımının girdisinin (x), denetimli öğrenme problemindeki bir sonraki zaman adımının çıktısı (y) olduğu görülmektedir.

Kayan pencere yöntemi, bir sonraki zaman adımını tahmin etmek için önceki zaman adımlarının kullanımına dayalı olan bir yöntemdir. Kayan pencerenin boyutu, önceki zaman adımlarının sayısına göre belirlenmektedir. Bu çalışma kapsamında kayan pencere boyu 3 olarak seçilmiştir.

Geliştirilen sistemde öncelikle her bir veriseti için veri ön işleme işlemleri gerçekleştirilmiştir. Verisetlerindeki eksik ve hatalı veriler kontrol edilerek temizlenmiştir. Verisetlerindeki dağılımlarının büyüklüğü sebebiyle çete saldırısı ve trafik kazası olay olarak seçilmiştir. Olay seçiminden sonra belirlenen olayların meydana gelme zamanları arasındaki saniye cinsinden zaman farkları hesaplanmıştır. Zaman serisi verilerinin denetimli öğrenme verilerine dönüştürülebilmesi için boyutu 3 olan kayan pencere kullanılmıştır. Bu sayede t_1 , t_2 ve t_3 zamanındaki veriler girdi, t_4 zamanındaki veri ise çıktı olacak şekilde yapılandırılmıştır.

Daha sonra olaylar arasındaki zaman farkları normalize edilmiştir. Normalleştirme, makine öğrenmesi için genellikle veri hazırlamanın bir parçası olarak uygulanan bir tekniktir. Normalleştirme amacı, veri kümesindeki sayısal sütunların değerlerini, değerler aralığındaki farklılıkları bozmadan ortak bir ölçüğe değiştirmektir. Normalizasyon geliştirilen modelin performansını ve eğitimin kararlılığını etkilemektedir. Veriler, 0-1 aralığında yeniden ölçeklendirilmiştir. Normalizasyon

için verideki en maksimum ve minimum değerler belirlenerek $y = (x - \min) / (\max - \min)$ eşitliği ile x verisi için normalize edilmiş değer olan y elde edilmiştir. Normalizasyon adımından sonra veriler %67 eğitim, %33 test olacak şekilde ayrılmıştır. Parametre optimizasyonu ile en düşük MSE değerine sahip parametreler belirlenerek model oluşturulmuş ve olay zamanı tahmin değerleri elde edilmiştir.

Suç veriseti

Bu çalışma kapsamında, bir olayın bir sonraki olma zamanının belirlenmesi hedeflenmiştir. Geliştirilen model, Baltimor şehrinin suç verileri kullanılarak ARIMA modeli ile karşılaştırmalı olarak analiz edilmiştir [40]. Kullanılan veriseti 285808 satır ve 12 sütundan oluşmaktadır. Suçun işlenme tarihi, saati, kodu, mekânı, açıklaması, kapalı mekân/açık alan, silah türü, posta kodu, semt, yakın komşuluk, konum ve olay sayısı sütunları bulunmaktadır. Bu çalışma kapsamında verisetindeki suçun işlenme tarihi, saati ve kodu alanları kullanılmıştır.

Çizelge 1’de ise suç tarihi, saati ve suç koduna göre düzenlenmiş Baltimor şehrindeki suç verisetinin ilk 5 satırını görmektedir.

Çizelge 1. Baltimor şehrindeki suç verileri (Crime data in Baltimore)

Suç tarihi	Suç saati	Suç kodu
2011-01-01	00:00:00	6E
2011-01-01	00:00:00	6G
2011-01-01	00:01:00	6J
2011-01-01	00:01:00	6G
2011-01-01	00:05:00	4B

Verisetindeki eksik ve tekrarlı veriler atıldıktan sonra 256688 satır olay verisi elde edilmiştir. Suç koduna göre yapılan inceleme sonucunda 41967 adet bulunan 4E kodlu çete saldırısı olay olarak seçilmiştir.

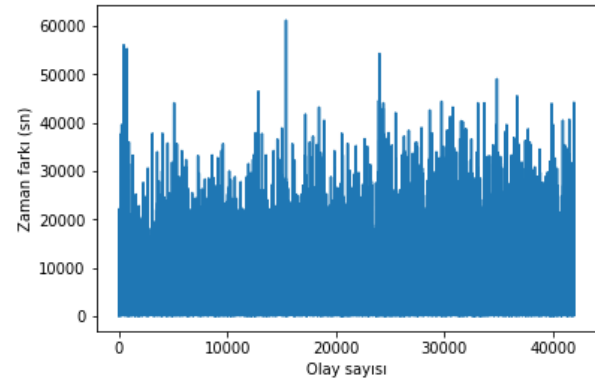
Bir olayın, bir sonraki olma zamanının belirlenmesi için verisetindeki suç olayların olma zamanları arasındaki saniye cinsinden fark hesaplanmıştır. Baltimor

şehrindeki suç veriseti için 4E kodlu çete saldırısı olaylarının meydana gelme zamanları arasındaki saniye cinsinden farklara göre oluşturulan verinin ilk 5 satırını Çizelge 2’de görmektedir.

Çizelge 2. Olaylarının meydana gelme zamanları arasındaki zaman farkı (Time difference between the occurrence of events)

Olay	Zaman farkı (s)
1	0
2	540
3	900
4	60
5	720

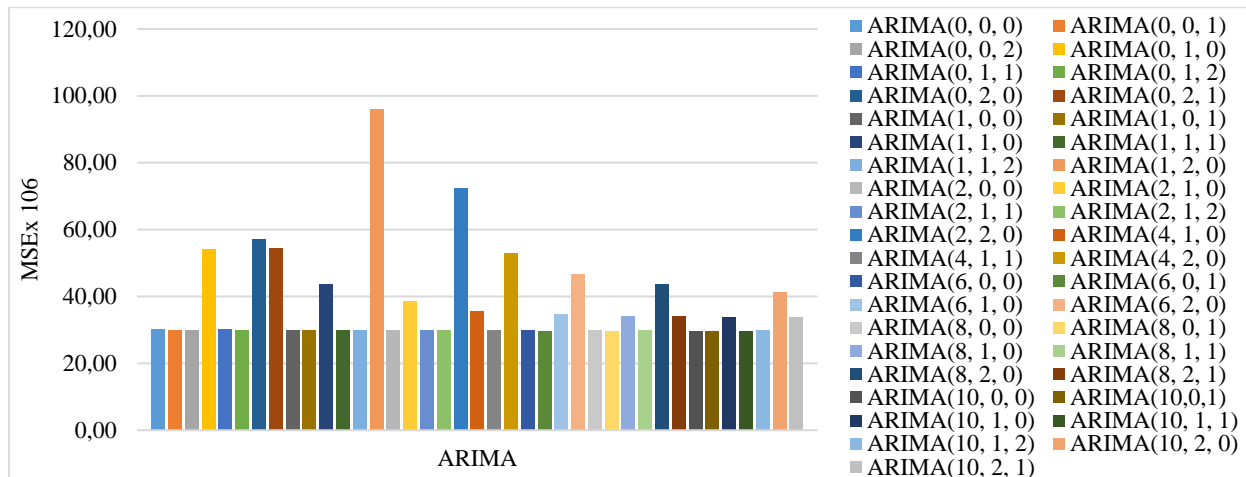
Çete saldırısı olayları arasındaki saniye cinsinden zaman farkları Şekil 3’te görmektedir.



Şekil 3. Olaylar arasındaki zaman farkları (Time differences between events)

Şekil 3’te görüldüğü gibi kullanılan veriseti 41967 adet 4E kodlu çete saldırısı olayından oluşmaktadır. Verilerin %67’si eğitim, %33’ü test olarak ayrılmıştır. Ayırma sonucunda, toplam 13850 elemandan oluşan bir test verisi elde edilmiştir.

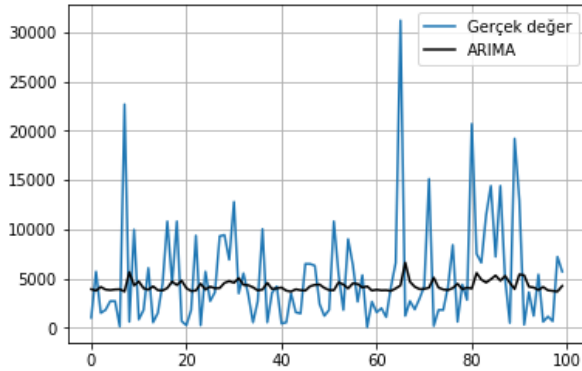
ARIMA modelinin alacağı p , d ve q parametrelerinin belirlenmesi için Şekil 4’te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.



Şekil 4. ARIMA parametre seçimi (ARIMA parameter selection)

Şekil 4'te görüldüğü gibi analiz sonuçları ARIMA(10,1,1) modelinin en düşük MSE değerine sahip olduğunu gösterdiği için, ARIMA(10,1,1) modeli deneysel çalışmalarda kullanılmak üzere seçilmiştir.

Şekil 5'te test verisinin ilk 100 satırlık örnekleme için ARIMA'nın tahmin sonuçları görülmektedir.



Şekil 5. ARIMA tahmin sonuçları (ARIMA prediction results)

Şekil 5'te verisetinin tamamı çok büyük olduğu için tahmin adımlarının net bir şekilde görülebilmesi amacıyla test verisinin ilk 100 satırlık örnekleme ve ARIMA'nın tahmin sonuçları görülmektedir. Şekil 5'te görüldüğü gibi ARIMA, veride yaşanan dalgalanmaları ve ani değişimleri etkin bir şekilde yakalayamamıştır. ARIMA istatistiksel bir model olduğu için geçmiş gözlemlerin ortalamasına ve standart sapmasına dayalı olarak olasılıksal bir hesaplama yapmaktadır. Bu sebeple veride yaşanan ani artış ve azalışları tahmin etmekte zorlanmaktadır.

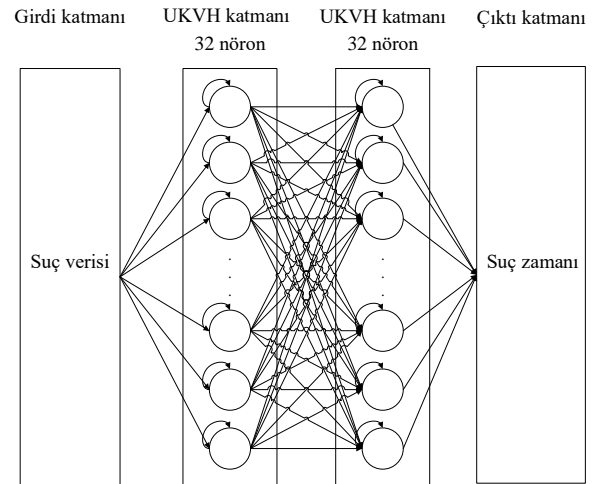
Geliştirilen modelin yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 3'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 3. Geliştirilen model için parametre seçimi (Parameter selection for developed model)

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
256	100	4	31352312,79
512	100	4	31459216,75
256	100	8	31492248,87
64	100	32	31608559,41
256	100	32	*31306043,78
512	100	32	31422236,24
1024	100	32	31489282,86
256	100	64	31547119,56
256	200	32	31485571,40

* Seçilen parametreler

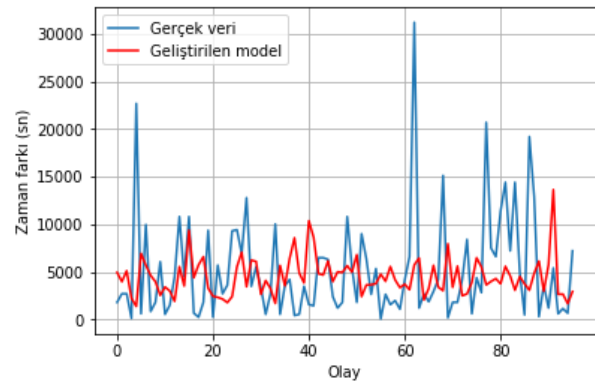
Analiz sonuçları geliştirilen modelin en düşük MSE değerine yığın boyutu 256, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir. Şekil 6'da geliştirilen UKVH tabanlı model görülmektedir.



Şekil 6. Geliştirilen UKVH tabanlı model (Developed LSTM based model)

Geliştirilen UKVH tabanlı model, Şekil 6'da görüldüğü gibi 2 katmana sahiptir. Her bir katmanda 32 UKVH nöron bulunmaktadır.

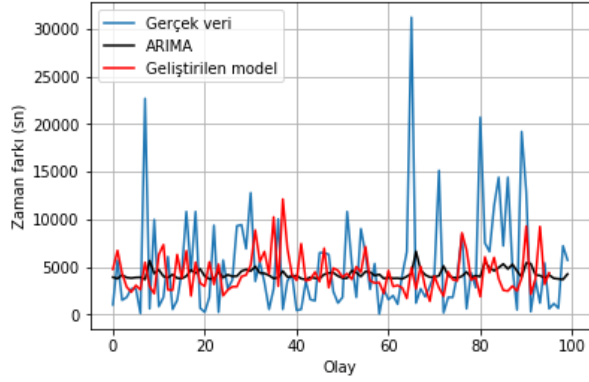
Şekil 7'de test verisinin ilk 100 satırlık örnekleme için geliştirilen modelin tahmin sonuçları görülmektedir.



Şekil 7. Geliştirilen modelin tahmin sonuçları (Prediction results of developed model)

Şekil 7'de verisetinin tamamı çok büyük olduğu için tahmin adımlarının net bir şekilde görülebilmesi amacıyla test verisinin ilk 100 satırlık örnekleme ve geliştirilen modelin tahmin sonuçları görülmektedir. UKVH, geçmiş gözlemlerden öğrenerek verideki gelecek değerleri tahmin etmektedir. Şekil 7'de görüldüğü gibi verideki ani iniş ve çıkışlar UKVH tarafından başarılı bir şekilde tahmin edilmiştir.

Şekil 8'de test verisinin ilk 100 elemanı için ARIMA ve geliştirilen modelin tahmin sonuçları görülmektedir.



Şekil 8. Test verisinin ilk 100 satırı için ARIMA ve geliştirilen modelin tahmin sonuçları (Prediction results of the ARIMA and developed model for the first 100 rows of test data)

Şekil 8’de test verisinin ilk 100 satırlık örnekleme için ARIMA’nın ve geliştirilen modelin tahmin sonuçları görülmektedir. Geliştirilen modelin tahmin sonuçları kırmızı ile ARIMA’nın tahmin sonuçları ise siyah çizgi ile çizilmiştir. Geliştirilen modelin, verisetindeki değişimleri ARIMA’ya göre daha başarılı bir şekilde tahmin edebildiği görülmektedir. Geliştirilen model, her bir adımda veriden öğrenerek tahmin modelini güncellemektedir. Bu sayede ARIMA’ya kıyasla daha başarılı bir tahmin performansı göstermektedir.

Sistem 10 defa çalıştırılarak elde edilen hata sonuçlarının ortalaması alınmıştır. MSE, RMSE ve MAE hata metriklerine göre ARIMA ve geliştirilen model için elde edilen deneysel sonuçlar Çizelge 4, Çizelge 5 ve Çizelge 6’da görülmektedir.

Çizelge 4. Suç veriseti için MSE metriğine göre deneysel sonuçlar (Experimental results according to MSE metric for crime dataset)

Test	ARIMA	Geliştirilen model
1	31709828,72	31306043,78
2	31709828,72	31308170,63
3	31709828,72	31341762,95
4	31709828,72	31345630,36
5	31709828,72	31350648,50
6	31709828,72	31346412,56
7	31709828,72	31342939,61
8	31709828,72	31341493,07
9	31709828,72	31344663,28
10	31709828,72	31343726,13
Ortalama	31709828,72	31337149,08

Çizelge 4’de, ARIMA ve geliştirilen model için MSE metriğine göre yapılan deneysel çalışmalar sonucunda, geliştirilen modelin en iyi MSE değerinin 31306043,78, en kötü MSE değerinin ise 31343726,13 olduğu görülmektedir. ARIMA ise deterministik bir model olduğu için yapılan 10 farklı deney sonucunda da

31709828,72 MSE değerine sahip olmuştur. Geliştirilen model daha düşük MSE değerine sahip olduğu için ARIMA’dan daha başarılıdır.

Çizelge 5. Suç veriseti için RMSE metriğine göre deneysel sonuçlar (Experimental results according to RMSE metric for crime dataset)

Test	ARIMA	Geliştirilen model
1	5631,14	5595,18
2	5631,14	5595,37
3	5631,14	4817,27
4	5631,14	5598,72
5	5631,14	5599,16
6	5631,14	5598,79
7	5631,14	5598,48
8	5631,14	5598,35
9	5631,14	5598,63
10	5631,14	5598,55
Ortalama	5631,14	5519,85

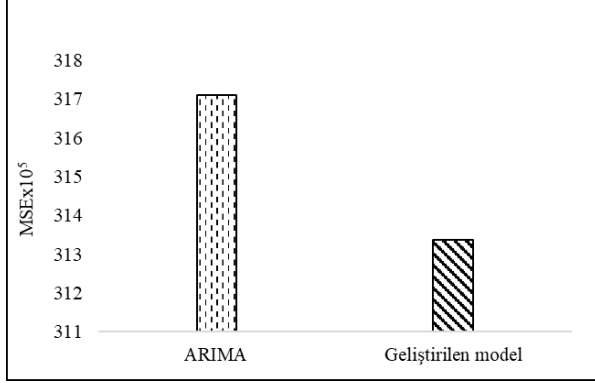
Çizelge 5’te, ARIMA ve geliştirilen model için RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, geliştirilen modelin en iyi RMSE değerinin 5595,18, en kötü RMSE değerinin ise 5598,55 olduğu görülmektedir. ARIMA ise deterministik bir model olduğu için yapılan 10 farklı deney sonucunda da 5631,14 RMSE değerine sahip olmuştur. Geliştirilen model daha düşük RMSE değerine sahip olduğu için ARIMA’dan daha başarılıdır.

Çizelge 6. Suç veriseti için MAE metriğine göre deneysel sonuçlar (Experimental results according to MAE metric for crime dataset)

Test	ARIMA	Geliştirilen model
1	4108,62	3891,01
2	4108,62	3890,87
3	4108,62	3796,91
4	4108,62	3795,82
5	4108,62	3795,77
6	4108,62	3795,99
7	4108,62	3796,24
8	4108,62	3796,91
9	4108,62	3796,04
10	4108,62	3796,53
Ortalama	4108,62	3815,20

Çizelge 6’da, ARIMA ve geliştirilen model için MAE metriğine göre yapılan deneysel çalışmalar sonucunda, geliştirilen modelin en iyi MAE değerinin 3891,01, en kötü MAE değerinin ise 3796,53 olduğu görülmektedir. ARIMA ise deterministik bir model olduğu için yapılan 10 farklı deney sonucunda da 4108,62 MAE değerine sahip olmuştur. Geliştirilen model daha düşük MAE değerine sahip olduğu için ARIMA’dan daha başarılıdır.

Çizelge 4, Çizelge 5, Çizelge 6 ve Şekil 9'da geliştirilen modelin ARIMA'ya göre daha düşük MSE, RMSE ve MAE değerlerine sahip olduğu görülmektedir. Suç olaylarının olma zamanlarının belirlenmesinde geliştirilen yöntemin ARIMA'dan daha başarılı sonuçlar verdiği görülmektedir.



Şekil 9. Suç veriseti için MSE metriğine göre deneysel sonuçlar (Experimental results according to MSE metric for crime dataset)

$\frac{\text{ARIMA-Geliştirilen model}}{\text{ARIMA}} \times 100$ eşitliğine göre geliştirilen

modelin ortalama değerinin ARIMA'ya göre iyileşme oranı hesaplanmıştır. Geliştirilen modelin ARIMA'ya göre MSE metriği için %1,17, RMSE metriği için %1,97 ve MAE metriği için %7,14 iyileşme sağladığı görülmektedir.

Acil çağrı veriseti

Geliştirilen model, Amerika'daki Montgomery şehrinin acil çağrı verileri kullanılarak ARIMA modeli ile karşılaştırmalı olarak analiz edilmiştir [41]. Kullanılan veriseti 423909 satır ve 8 sütundan oluşmaktadır. Verisetinde acil çağrı yapılan bölgenin enlem ve boylam bilgileri, acil durum hakkındaki açıklama bilgisi, posta kodu, acil çağrının dâhil olduğu kategori, tarih ve saat bilgisi, kasaba ve adres sütunları bulunmaktadır. Bu çalışma kapsamında verisetindeki acil çağrının alındığı tarih, saat ve kategori alanları kullanılmıştır.

Çizelge 7'de ise acil çağrının alındığı tarih, saat ve kategori alanlarına göre düzenlenmiş acil çağrı verisetinin ilk 5 satırını görülmektedir.

Çizelge 7. Montgomery şehrinin acil çağrı verileri (Emergency call data of Montgomery city)

Tarih	Saat	Kategori
10-12-2015	14:39:21	Yangın
10-12-2015	16:17:05	Kafa travması
10-12-2015	16:32:10	Trafik kazası
10-12-2015	16:46:48	Bulantı/Kusma

10-12-2015	16:47:36	Kalp krizi
------------	----------	------------

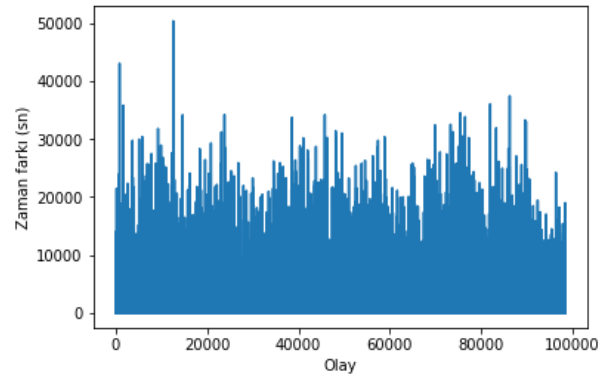
Acil çağrı veriseti Çizelge 7'de görüldüğü gibi çeşitli kategorilerdeki acil durumlardan oluşmaktadır. Bu çalışma kapsamında verisetinde diğer olaylara göre daha yaygın olarak görülen trafik kazası olayı olarak seçilmiştir. Verisetinde 98400 adet trafik kazası olayı bulunmaktadır.

Bir olayın, bir sonraki olma zamanının belirlenmesi için verisetindeki trafik kazası olayların olma zamanları arasındaki saniye cinsinden fark hesaplanmıştır. Trafik kazalarının meydana gelme zamanları arasındaki saniye cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 8'de görülmektedir.

Çizelge 8. Trafik kazası olaylarının meydana gelme zamanları arasındaki fark (Difference between time of occurrence of traffic accident events)

Olay	Zaman farkı (s)
1	0
2	2259
3	178
4	182
5	911

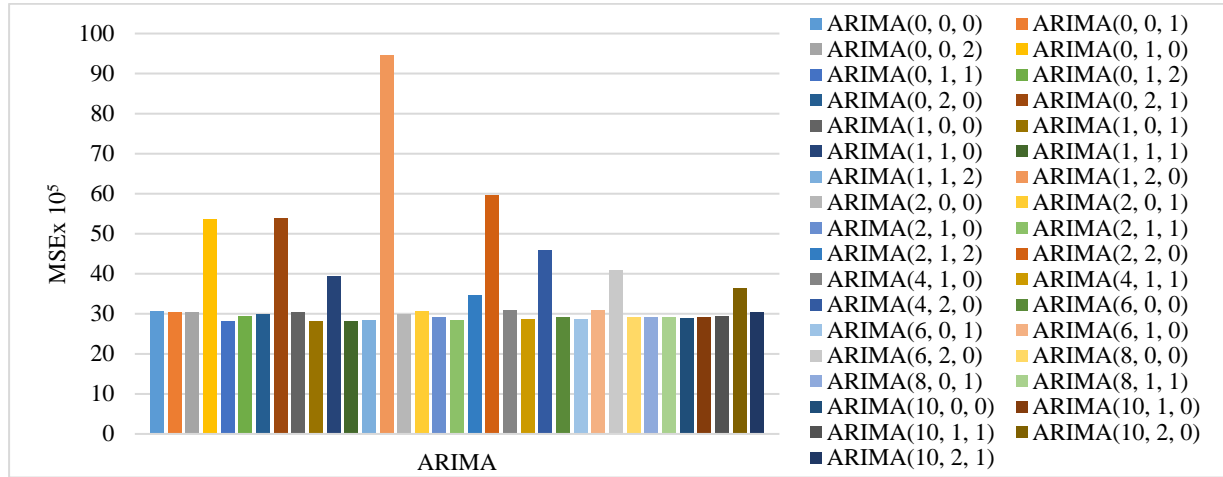
Trafik kazası olayları arasındaki saniye cinsinden zaman farkları Şekil 10'da görülmektedir.



Şekil 10. Trafik kazası olayları arasındaki zaman farkı (Time difference between traffic accident events)

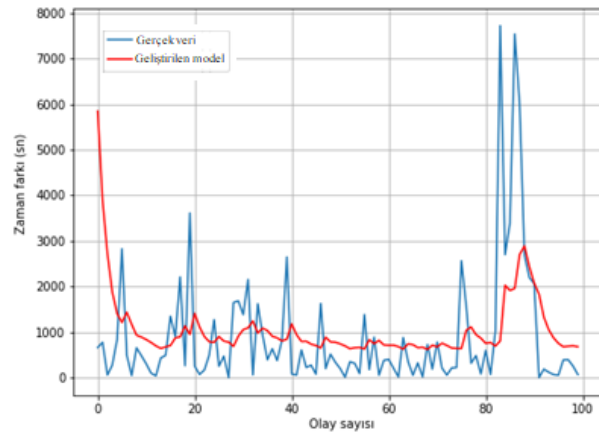
Şekil 10'da görüldüğü gibi kullanılan veriseti 98400 adet trafik kazası olayından oluşmaktadır. Verilerin %67'si eğitim, %33'ü test olarak ayrılmıştır. Ayırma sonucunda, toplam 32472 elemandan oluşan bir test verisi elde edilmiştir.

ARIMA modelinin alacağı p , d ve q parametrelerinin belirlenmesi için Şekil 11'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.



Şekil 11. ARIMA parametre seçimi (ARIMA parameter selection)

Analiz sonuçları ARIMA(1,0,1) modelinin en düşük MSE değerine sahip olduğunu gösterdiği için, ARIMA(1,0,1) modeli deneysel çalışmalarda kullanılmak üzere seçilmiştir. Şekil 12’de test verisinin ilk 100 satırlık örnekleme için ARIMA’nın tahmin sonuçları görülmektedir.



Şekil 12. ARIMA'nın tahmin sonuçları (Prediction results of ARIMA)

Şekil 12’de verisetinin tamamı çok büyük olduğu için tahmin adımlarının net bir şekilde görülebilmesi amacıyla test verisinin ilk 100 satırlık örnekleme ve ARIMA’nın tahmin sonuçları görülmektedir. Şekil 12’de görüldüğü gibi ARIMA, veride yaşanan dalgalanmaları ve ani değişimleri etkin bir şekilde tahmin edememiştir. ARIMA istatistiksel bir model olduğu için geçmiş gözlemlerin ortalamasına ve standart sapmasına dayalı olarak olasılıksal bir hesaplama yapmaktadır. Bu sebeple veride yaşanan ani artış ve azalışları tahmin etmekte zorlanmaktadır.

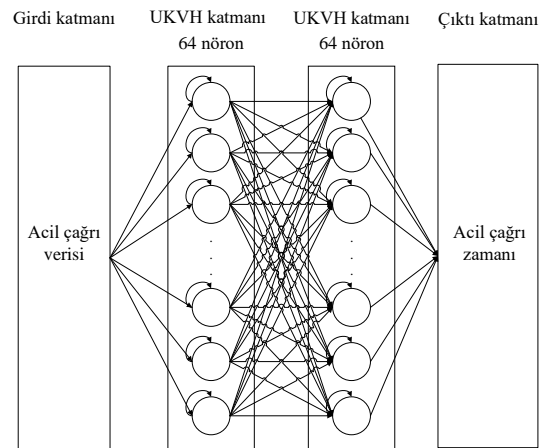
Geliştirilen modelin yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 9’da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 9. Geliştirilen model için parametre seçimi (Parameter selection for developed model)

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
1	100	4	3602169,38
8	100	4	3537409,65
32	100	16	3521867,50
32	100	32	3521145,83
32	50	64	3512457,51
32	100	64	*3492820,03
32	200	64	3498146,73
32	100	128	3506512,29
64	100	64	3506549,02
128	100	64	3525842,21
256	100	64	3513163,79
512	100	64	3518091,12
1024	100	64	3514046,45

* Seçilen parametreler

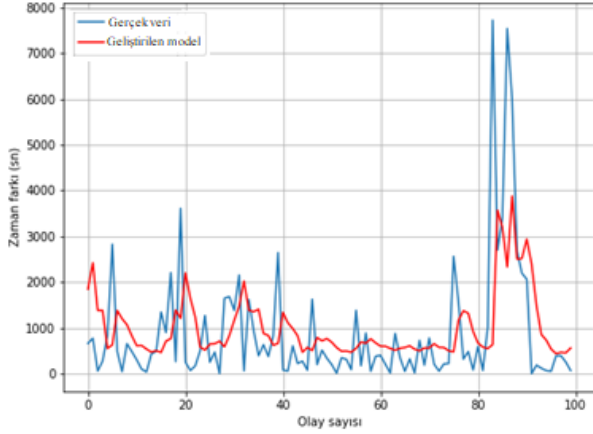
Analiz sonuçları geliştirilen modelin en düşük MSE değerine yığın boyutu 32, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 64 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir. Şekil 13’de geliştirilen UKVH tabanlı model görülmektedir.



Şekil 13. Geliştirilen UKVH tabanlı model (Developed LSTM based model)

Geliştirilen UKVH tabanlı model, Şekil 13’de görüldüğü gibi 2 katmana sahiptir. Her bir katmanda 64 UKVH nöron bulunmaktadır.

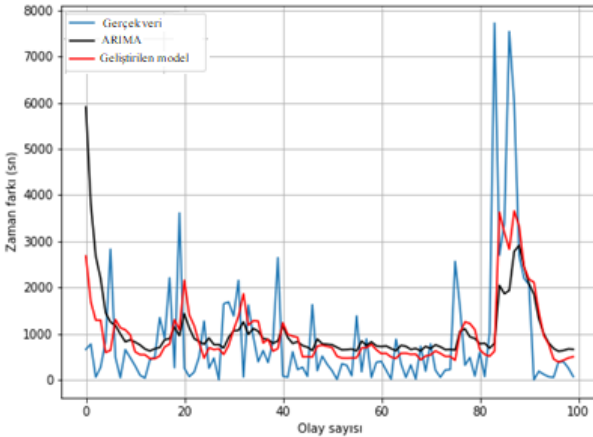
Şekil 14’te test verisinin ilk 100 satırlık örnekleme için geliştirilen modelin tahmin sonuçları görülmektedir.



Şekil 14. Geliştirilen modelin tahmin sonuçları (Prediction results of developed model)

Şekil 14’te verisetinin tamamı çok büyük olduğu için tahmin adımlarının net bir şekilde görülebilmesi amacıyla test verisinin ilk 100 satırlık örnekleme ve geliştirilen modelin tahmin sonuçları görülmektedir. UKVH, geçmiş gözlemlerden öğrenerek verideki gelecek değerleri tahmin etmektedir. Şekil 14’te görüldüğü gibi verideki ani iniş ve çıkışlar UKVH tarafından başarılı bir şekilde tahmin edilmiştir.

Şekil 15’te test verisinin ilk 100 elemanı için ARIMA ve geliştirilen modelin tahmin sonuçları görülmektedir.



Şekil 15. Test verisinin ilk 100 satırı için ARIMA ve geliştirilen modelin tahmin sonuçları (Prediction results of the ARIMA and developed model for the first 100 rows of test data)

Şekil 15’te test verisinin ilk 100 satırlık örnekleme için ARIMA’nın ve geliştirilen modelin tahmin sonuçları görülmektedir. Geliştirilen modelin tahmin sonuçları kırmızı ile ARIMA’nın tahmin sonuçları ise siyah çizgi ile çizilmiştir. Geliştirilen modelin, verisetindeki değişimleri ARIMA’ya göre daha başarılı bir şekilde

tahmin edebildiği görülmektedir. Geliştirilen model, her bir adımda veriden öğrenerek tahmin modelini güncellemektedir. Bu sayede ARIMA’ya kıyasla daha başarılı bir tahmin performansı göstermektedir.

Sistem 10 defa çalıştırılarak elde edilen hata sonuçlarının ortalaması alınmıştır. MSE, metriğine göre ARIMA ve geliştirilen model için elde edilen deneysel sonuçlar Çizelge 10’da görülmektedir.

Çizelge 10. Acil çağrı veriseti için MSE metriğine göre deneysel sonuçlar (Experimental results according to MSE metric for emergency call dataset)

Test	ARIMA	Geliştirilen model
1	3652989,22	3492820,03
2	3652989,22	3492751,84
3	3652989,22	3514050,37
4	3652989,22	3514049,89
5	3652989,22	3514041,67
6	3652989,22	3514048,69
7	3652989,22	3514049,08
8	3652989,22	3514042,70
9	3652989,22	3514046,65
10	3652989,22	3514048,47
Ortalama	3652989,22	3509794,93

Çizelge 10’da, ARIMA ve geliştirilen model için MSE metriğine göre yapılan deneysel çalışmalar sonucunda, geliştirilen modelin en iyi MSE değerinin 3492820,03, en kötü MSE değerinin ise 3514048,47 olduğu görülmektedir. ARIMA ise deterministik bir model olduğu için yapılan 10 farklı deney sonucunda da 3652989,22 MSE değerine sahip olmuştur. Geliştirilen model daha düşük MSE değerine sahip olduğu için ARIMA’dan daha başarılıdır.

RMSE metriğine göre ARIMA ve geliştirilen model için elde edilen deneysel sonuçlar Çizelge 11’de görülmektedir.

Çizelge 11. Acil çağrı veriseti için RMSE metriğine göre deneysel sonuçlar (Experimental results according to RMSE metric for emergency call dataset)

Test	ARIMA	Geliştirilen model
1	1911,27	1868,91
2	1911,27	1868,89
3	1911,27	1874,58
4	1911,27	1874,58
5	1911,27	1874,58
6	1911,27	1874,58
7	1911,27	1874,58
8	1911,27	1874,58
9	1911,27	1874,58
10	1911,27	1874,58
Ortalama	1911,27	1873,44

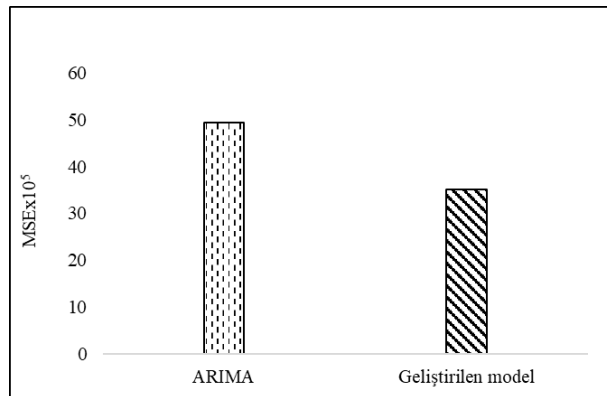
Çizelge 11’de, ARIMA ve geliştirilen model için RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, geliştirilen modelin en iyi RMSE değerinin 1868,91, en

kötü RMSE değerinin ise 1874,58 olduğu görülmektedir. ARIMA ise deterministik bir model olduğu için yapılan 10 farklı deney sonucunda da 1911,27 RMSE değerine sahip olmuştur. Geliştirilen model daha düşük RMSE değerine sahip olduğu için ARIMA'dan daha başarılıdır. MAE metriğine göre ARIMA ve geliştirilen model için elde edilen deneysel sonuçlar Çizelge 12'de görülmektedir.

Çizelge 12. Acil çağrı veriseti için MAE metriğine göre deneysel sonuçlar (Experimental results according to MAE metric for emergency call dataset)

Test	ARIMA	Geliştirilen model
1	892,78	833,18
2	892,78	833,14
3	892,78	860,43
4	892,78	860,43
5	892,78	860,44
6	892,78	860,43
7	892,78	860,43
8	892,78	860,44
9	892,78	860,43
10	892,78	860,43
Ortalama	892,78	854,97

Çizelge 12'de, ARIMA ve geliştirilen model için MAE metriğine göre yapılan deneysel çalışmalar sonucunda, geliştirilen modelin en iyi MAE değerinin 833,18, en kötü MAE değerinin ise 860,43 olduğu görülmektedir. ARIMA ise deterministik bir model olduğu için yapılan 10 farklı deney sonucunda da 892,78 MAE değerine sahip olmuştur. Geliştirilen model daha düşük MAE değerine sahip olduğu için ARIMA'dan daha başarılıdır. Çizelge 10, Çizelge 11, Çizelge 12 ve Şekil 16'da geliştirilen modelin ARIMA'ya göre daha düşük MSE, RMSE ve MAE değerlerine sahip olduğu görülmektedir. Acil çağrı olaylarının olma zamanlarının belirlenmesinde geliştirilen yöntemin ARIMA'dan daha başarılı sonuçlar verdiği görülmektedir.



Şekil 16. Acil çağrı veriseti için MSE metriğine göre deneysel sonuçlar (Experimental results according to MSE metric for emergency call dataset)

ARIMA-Geliştirilen model $\times 100$ eşitliğine göre geliştirilen ARIMA

modelin ortalama değerinin ARIMA'ya göre iyileşme oranı hesaplanmıştır. Geliştirilen modelin ARIMA'ya göre MSE metriği için %3,91, RMSE metriği için %1,97 ve MAE metriği için %4,23 iyileşme sağladığı görülmektedir.

5. SONUÇLAR (CONCLUSIONS)

Bu çalışmada, zaman serisi analizlerinde kullanılan derin öğrenme tabanlı yaklaşımlar ile geleneksel istatistik tabanlı yaklaşımlar karşılaştırmalı olarak analiz edilmiştir. Veriseti olarak Baltimor şehrinde meydana gelen suç verileri ile Montgomery şehrindeki acil çağrı verileri kullanılmıştır. Çalışma kapsamında çete saldırısı ve trafik kazası olayları tahmin edilecek olay olarak belirlenmiştir. Bir olayın bir sonraki olma zamanının belirlenmesi amacıyla olayların olma zamanları arasındaki süre saniye olarak hesaplanmıştır. ARIMA ve geliştirilen model kullanılarak yapılan deneysel çalışmalarda, suç veriseti için geliştirilen modelin 31337149,08, ARIMA'nın ise 31709828,72 MSE değerine sahip olduğu, acil çağrı veriseti için ise geliştirilen modelin 3509794,93, ARIMA'nın ise 3652989,22 MSE değerine sahip olduğu görülmüştür. Suç veriseti için geliştirilen UKVH tabanlı modelin ARIMA'ya göre MSE metriği için %1,17, RMSE metriği için %1,97 ve MAE metriği için %7,14 iyileşme sağladığı görülmüştür. Acil çağrı veriseti için ise geliştirilen modelin ARIMA'ya göre MSE metriği için %3,91, RMSE metriği için %1,97 ve MAE metriği için %4,23 iyileşme sağladığı görülmüştür.

Elde edilen sonuçlar, eldeki mevcut veriler kullanılarak eğitilen derin öğrenme modelinin suçun olma zamanının ve bir sonraki acil çağrı zamanının tahmininde başarılı olduğunu göstermektedir. Geliştirilen model ile geçmiş bilgilere dayanarak bir olayın bir sonraki olma zamanı başarılı bir şekilde tahmin edilmektedir.

ETİK STANDARTLARIN BEYANI (DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazar(lar)ı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

KAYNAKLAR (REFERENCES)

- [1] Adel, H., Salheen, M. and Mahmoud, R. A., "Crime in relation to urban design, Case study: The Greater Cairo Region", *Ain Shams Engineering Journal*, 7(3): 925-938, (2016).
- [2] Yu, C. H., Ward, M. W., Morabito, M. and Ding, W., "Crime forecasting using data mining techniques", *IEEE 11th international conference on data mining workshops*, 779-786, (2011).
- [3] Wang, B., Zhang, D., Zhang, D., Brantingham, P. J. and Bertozzi, A. L., "Deep learning for real time crime forecasting", *arXiv*, (2017).
- [4] Stalidis, P., Semertzidis, T. and Daras, P., "Examining Deep Learning Architectures for Crime Classification and Prediction", *arXiv*, (2018).

- [5] Azeez, J. and Aravindhar, D. J., "Hybrid approach to crime prediction using deep learning", *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1701-1710, (2015).
- [6] Han, J., "*Data mining: concepts and techniques*", Morgan Kaufmann, ABD, (2012).
- [7] Cesario, E., Catlett, C. and Talia, D., "Forecasting crimes using autoregressive models", *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing*, 795-802, (2016).
- [8] Fu, R., Zhang, Z. and Li, L., "Using LSTM and GRU neural network methods for traffic flow prediction", *Chinese Association of Automation (YAC), Youth Academic Annual Conference*, 324-328, (2016).
- [9] Kang, H. W. and Kang, H. B., "Prediction of crime occurrence from multi-modal data using deep learning", *PloS one*, 12(4): (2017).
- [10] Zheng, J., Xu, C., Zhang, Z. and Li, X., "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network", *Information Sciences and Systems (CISS)*, (2017).
- [11] Kouziokas, G. N., "The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment", *Transportation research procedia*, 24: 467-473, (2017).
- [12] Lin, T., Guo, T. and Aberer, K., "Hybrid neural networks for learning the trend in time series", *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2273-2279, (2017).
- [13] McNally, S., Roche, J. and Caton, S., "Predicting the price of Bitcoin using Machine Learning", *2018 26th Euromicro International Conference*, 339-343, (2018).
- [14] Tian, C., Ma, J., Zhang, C. and Zhan, P., "A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network", *Energies*, (2018).
- [15] Stec, A. and Klabjan, D., "Forecasting Crime with Deep Learning", *arXiv*, (2018).
- [16] "*United Nations Settlements Programme, The state of the world's cities 2004/2005: Globalization and urban culture*", Earthscan, ABD, (2004).
- [17] Adhikari, R. and Agrawal, R. K., "An introductory study on time series modeling and forecasting", *arXiv*, (2013).
- [18] Wang, B., Luo, X., Zhang, F., Yuan, B., Bertozzi, A. L. and Brantingham, P. J., "Graph-Based Deep Modeling and Real Time Forecasting of Sparse Spatio-Temporal Data", *arXiv*, (2018).
- [19] Chen, X., Cho, Y. and Jang, S. Y., "Crime prediction using Twitter sentiment and weather", *2015 Systems and Information Engineering Design Symposium*, 63-68, (2015).
- [20] Munjal, K. G., Silverman, R. A., Freese, J., Braun, J. D., Kaufman, B. J., Isaacs, D. and Prezant, D. J., "Utilization of emergency medical services in a large urban area: description of call types and temporal trends", *Prehospital Emergency Care*, 15(3): 371-380, (2011).
- [21] Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G. and Riquelme, J. C., "A survey on data mining techniques applied to electricity-related time series forecasting", *Energies*, 8(11): 13162-13193, (2015).
- [22] Chatfield, C., "The analysis of time series: An introduction", *CRC press*, (2016).
- [23] Granger, C. W. J. and Newbold, P., "Forecasting economic time series", *Academic Press*, (2014).
- [24] Osmanoğlu, B., Sunar, F., Wdowinski, S. and Cabral-Cano, E., "Time series analysis of InSAR data: Methods and trends", *ISPRS Journal of Photogrammetry and Remote Sensing*, 115: 90-102, (2016).
- [25] Weigend, A. S., "Time series prediction: forecasting the future and understanding the past", *Routledge*, (2018).
- [26] Kumar, S. V. and Vanajakshi, L., "Short-term traffic flow prediction using seasonal ARIMA model with limited input data", *European Transport Research Review*, 7(3), (2015).
- [27] Montgomery, D. C., Jennings, C. L. and Kulahci, M., "Introduction to time series analysis and forecasting", *John Wiley & Sons*, (2015).
- [28] Brockwell, P. J. and Davis, R. A., "Introduction to time series and forecasting", *Springer*, (2016).
- [29] Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M., "Time series analysis: forecasting and control", *John Wiley & Sons*, ABD, (2015).
- [30] Valipour, M., Banihabib, M. E. and Behbahani, S. M. R., "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir", *Journal of Hydrology*, 476: 433-441, (2013).
- [31] Siami-Namini, S. and Namin, A. S., "Forecasting Economics and Financial Time Series: ARIMA vs. LSTM", *arXiv*, (2018).
- [32] Wang, W. C., Chau, K. W., Xu, D. M. and Chen, X. Y., "Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition", *Water Resources Management*, 29(8): 2655-2675, (2015).
- [33] Långkvist, M., Karlsson, L. and Loutfi, A., "A review of unsupervised feature learning and deep learning for time-series modeling", *Pattern Recognition Letters*, 42: 11-24, (2014).
- [34] Gamboa, J. C. B., "Deep learning for time-series analysis", *arXiv*, (2017).
- [35] Graupe, D., "Principles of artificial neural networks", *World Scientific*, (2013).
- [36] Karayiannis, N. and Venetsanopoulos, A. N., "Artificial neural networks: learning algorithms, performance evaluation, and applications", *Springer Science & Business Media*, Almanya, (2013).
- [37] Malhotra, P., Vig, L., Shroff, G. and Agarwal, P., "Long short term memory networks for anomaly detection in time series", *Proceedings Presses universitaires de Louvain*, (2015).
- [38] Fan, J., Ma, C. and Zhong, Y., "A Selective Overview of Deep Learning", *arXiv:1904.05526*, (2019).
- [39] Grandell, J., "Time Series Analysis Lecture notes", *KTH Stockholm University*, (2015).
- [40] Baltimore Crime Dataset, (2016). <https://data.world/data-society/city-of-baltimore-crime-data>
- [41] Montgomery country emergency call dataset, (2018). <https://www.kaggle.com/mchirico/montcoalert>