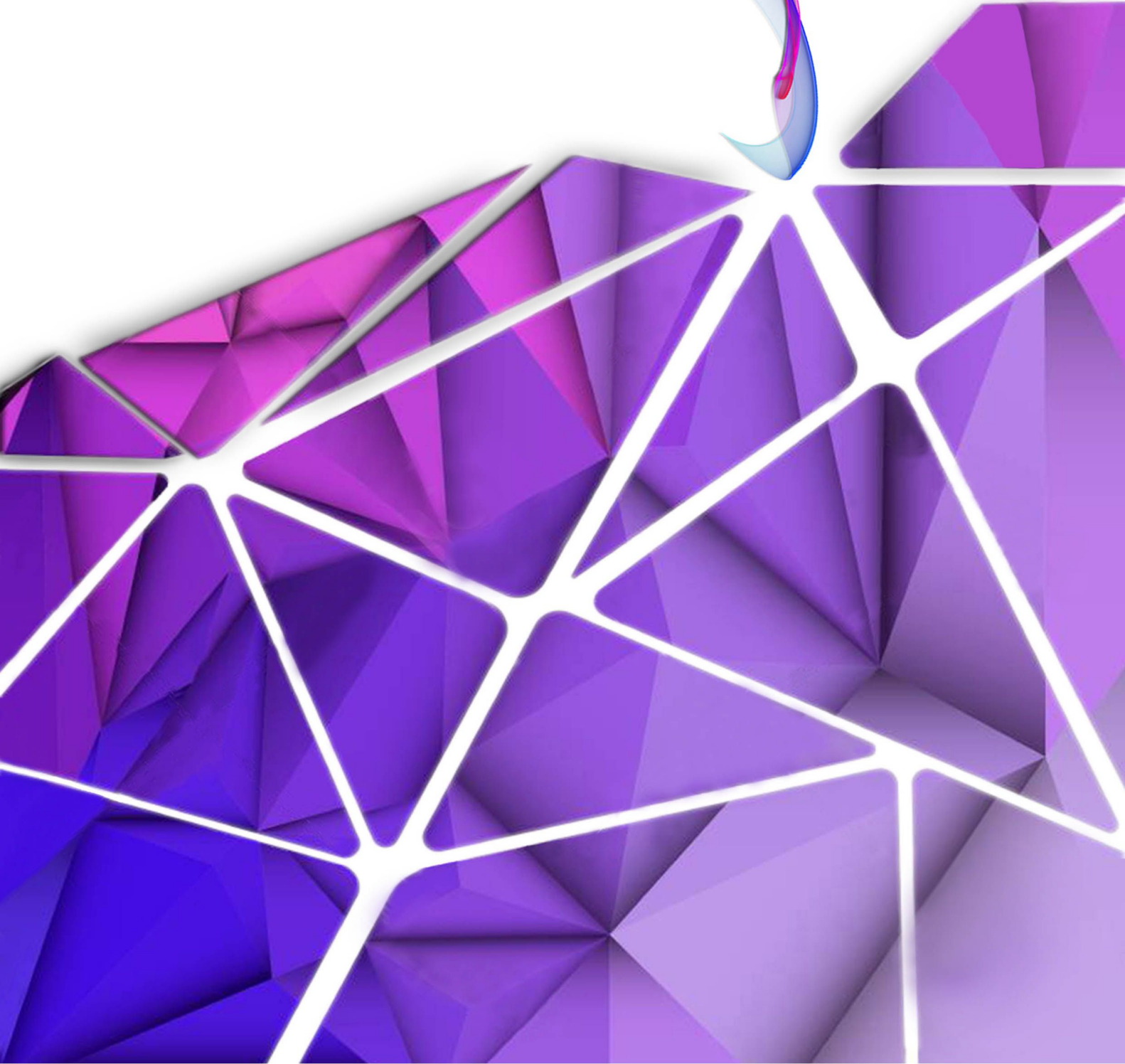


2020, CİLT 1 SAYI 1

e-ISSN 2717-9397

İLERİ MÜHENDİSLİK ÇALIŞMALARI VE TEKNOLOJİLERİ DERGİSİ



İÇİNDEKİLER

Zeynep ORMAN Mert KAVİ	Akan Veri İşleyen Dağıtık Sistemlerde Gecikme Duyarlı Dinamik Ölçekleme İçin Bir Sistem Tasarımı A System Design for Latency Aware Dynamic Scaling at Distributed Data Stream Processing System	1-12
Alper ÇİÇEK Yelda ARSLAN	Müşteri Kayıp Analizi İçin Sınıflandırma Algoritmalarının Karşılaştırılması Comparing Classification Algorithms of Customer Loss Analysis	13-19
Saadet Aytaç ARPACI Banu DİRİ	Birliktelik Kuralı Temelinde Kısaltma Genişletme Abbreviation Expansion on the Basis of the Association Rule	20-27
Berfinsu BARUT Muharrem ÜNVER Celal Türkcan KAYIM Ecem TOPRAK Ezgi UYSAL	Otomotiv Endüstrisinde Akıllı Fabrika Uygulamaları ve Türkiye’de Adaptasyon Süreci Smart Factory Applications in the Automotive Industry and Adaptation Process in Turkey	28-38
Sevda ÖZEL	Afetlerden Sonra Kirlilik ve İkincil Kirliliği Afet Olarak Değerlendirmek İçin Bir Tartışma A Discussion on the Assessment of Pollution and Secondary Pollution as Disasters After Disasters	39-48
Elif ÖZTAD	Meme Kanseri Tespitinde Sınıflandırma ve Sinir Ağları Yöntemlerinin Karşılaştırılması Comparison Between Classification and Neural Network Methods In Breast Cancer Detection	49-54
Bahdişen GEZER Can GEZER	Ag/Aniline blue/n-Si Type Schottky Diode with Organic Dye Interlayers Investigation of Electrical and Photovoltaic Properties Organik Boya Ara Katmanlı Ag/Anilin mavisi/n-Si Tipi Schottky Diyot Elektrik ve Fotovoltaik Özelliklerin Araştırılması	55-62

Akan Veri İşleyen Dağıtık Sistemlerde Gecikme Duyarlı Dinamik Ölçekleme İçin Bir Sistem Tasarımı

Zeynep ORMAN*¹ , Mert KAVI*² 

*İstanbul Üniversitesi-Cerrahpaşa, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 34320, İstanbul

Araştırma Makalesi, Geliş Tarihi: 13.07.2020, Kabul Tarihi: 11.08.2020

Özet

Büyük ölçekli akan veri işleyen dağıtık sistemleri inşa etmek ve operasyonunu sağlamak oldukça karmaşık ve maliyetli bir süreçtir. Sistemlerin veri akışının değişen hızlarına adapte olabilmesi ve gerektiğinde ölçeklenebilmesi gerekmektedir. Bu nedenle, akan veriyi işleyen dağıtık sistemlere entegre edilecek etkin bir otomatik ölçekleme sistemi kullanılması çoğu zaman kaçınılmazdır. Son yıllarda, hızla artan akan veri kaynaklarını işleyebilen sistemlere olan ilgi oldukça artmıştır ve literatürde bu alanda yapılan çok sayıda çalışma bulunmaktadır. Ancak bu çalışmaların çoğu sistemin değişen iş yüklerine adapte olabilmesi ve ölçeklenebilirlik konusu yerine sistemin olağan şartlarda nasıl çalışacağı üzerine yoğunlaşmıştır. Literatürde az sayıda olan ölçeklenebilirlik ile ilgili çalışmalarda ise genellikle ölçeklenebilirlik bir kaynak kümesi ile gerçekleştirilmektedir. Bu çalışmada, literatürdeki bu eksikliklerden yola çıkılarak, Apache Flink üzerinde çalışan, değişen çalışma yüklerine adapte olabilen bir sistem tasarımı önerilmiştir. Apache Flink, hem sistem geliştirme hem de ölçekleme metriklerini hesaplama amacıyla kullanılmıştır. Ölçekleme, Kuyruk Teorisi kullanılarak hesaplanan, sistemde meydana gelmesi beklenen gecikme ve kritik sistem metrikleri değerlendirilerek gerçekleştirilmiştir. Büyük veri işleyen sistemlere entegre çalışabilecek bu model ile sistem performanslarının geliştirilmesi ve kalite kayıplarının azaltılması hedeflenmiştir. Son olarak, sistemin hangi durumlarda ölçeklendiği ve ölçeklemeden sonraki durumu benzetim çalışmaları ile gerçekleştirilerek önerilen sistemin etkinliği gösterilmiştir.

Anahtar Kelimeler: Dağıtık sistemler, Büyük veri, Akan veri işleme, Ölçeklenebilirlik, Kuyruk teorisi.

A System Design for Latency Aware Dynamic Scaling at Distributed Data Stream Processing System

Abstract

Establishing large-scale distributed stream processing systems and ensuring their operations is a very complex and costly process. These systems should be capable of adapting the varying rates of data stream and they must be scaled, if required. It is usually inevitable to use an effective automatic scaling system which can be integrated into such systems. In recent literature, there are numerous studies on this issue. Many of these studies have focused on how these systems will operate under normal conditions. There are limited studies on scalability where scaling is usually implemented with a set of resources. In this study, based on these shortcomings, a system design which can adapt to changing working loads and work on Apache Flink, is proposed. Apache Flink is used for both system development and calculating the scaling metrics. Scaling is performed by evaluating the expected latency calculated with Queuing Theory and some critical metrics. It is aimed to improve system performances and reduce quality losses with this model, which can be integrated into big data processing systems. Pre-scaling and post-scaling cases are also demonstrated by simulations to show the effectiveness of the proposed system.

Keywords: Distributed systems, Big data, Stream processing, Scalability, Queuing theory.

¹Sorumlu yazar ormanz@istanbul.edu.tr, ²mertkavi@gmail.com

1. GİRİŞ

Gelişen teknoloji ile birlikte kuruluşlar büyük hacimde ve hızda veri üretmektedir. Sayısı sürekli artan veri kaynakları birçok veri akışı oluşturmaktadır. Uygulama sunucularından gelen log verileri, web sitelerinden, mobil uygulamalardan gelen klik akışları, Nesnelerin İnterneti'ni oluşturan aygıtlardan gelen ölçüm verilerinin tümü, müşterileri, uygulamaları ve ürünleri daha iyi anlamak, tanımak ve analiz etmek için yardımcı olmaktadır. Oluşan bu veri akışlarını gerçek zamanlı olarak işleyebilmek ve analiz edebilmek, uygulamaları devamlı olarak gözlemleyebilmek ve bu sayede müşterilerine kişiselleştirilmiş teklifler vererek ürün önerileri yapabilmek kurumlara rekabet ortamında büyük avantajlar sağlamaktadır. Web site analitiği ve makine öğrenmesi yöntemleri, daha doğru ve aksiyon alınabilir sonuçlar elde etmek için veriyi saatler ve günler yerine dakikalar ya da saniyeler içinde kullanılabilir hale gelmesini sağlayabilmektedir.

Akan veriyi işleyen uygulamalarda yığın analitikten akan analitiğe geçiş ve gerçek zamanlı uygulamalar inşa edilmesi şeklinde iki tür kullanım senaryosu vardır. Yığın veriyi işleme ve akan veriyi işleme arasında bazı temel farklılıklar vardır. Akan veriyi toplamak, hazırlamak ve işlemek için, yığın veriyi işlemeye göre farklı araçlar kullanılır. Yığın veri işleyen geleneksel sistemlerde veri alınır ve periyodik bir şekilde veri tabanına yüklenir. Veriyi analiz edebilmek saatler, günler hatta bazen haftalar sürebilir. Akan veriyi işleyen sistemlerde ise veri gerçek zamanlı olarak işlenir, bu işlem verinin saklanmasıyla önce gerçekleştirilir. Gerçek zamanlı olarak veriyi işlemek, geleneksel veri işleyen teknolojilere göre çok daha hızlı karar alınmasını sağlar. Fakat bu sistemleri inşa etmek ve operasyonunu sağlamak oldukça karmaşık bir iştir ve sistem kaynaklarının kullanımına sıkı sıkıya bağlıdır. Bu yüzden bu sistemlerin inşası mali açıdan verimli olmak zorundadır. Hesaplama kaynaklarının optimize edilmesi, maksimum verim ve minimum gecikmeyle sistemin çalışması gerekmektedir. Sunucuların yönetilmesi, verinin değişen hızlarına adapte olabilmesi ve gerektiğinde ölçeklenmesi kaçınılmazdır. Tüm bu işlemlerin verimsiz çalışması zaman ve para kaybını da beraberinde getirmektedir. Dağıtık veri işleyen sistemlerde karşılaşılan en büyük zorluklardan bir tanesi de sistemin optimize çalışması için gereken kaynakların hesap edilmesidir. Öngörülemeyen veri hacmi ve hız değişimleri, ihtiyaç duyulan kaynak sayısının hesaplanmasını zorlaştırır. Gerekinden fazla kaynak tahsis etmek kaynak israfına yol açarken, düşük sayıda kaynak tahsisi ise sistem performansının düşmesine

sebeptir. Bu nedenle, dağıtık akan veriyi işleyen sistemlere dahil edilecek etkin bir otomatik ölçekleme sistemi kullanılması gerekmektedir. Bu sistem elastisite prensiplerine uygun olarak servisin çalışma kalitesini ve maliyetleri göz önünde bulundurarak, çalışan sisteme entegre olmalıdır.

Dağıtık akan veri işleme sistemleri son yıllarda oldukça popülerite kazanmış ve literatürde üzerine çok sayıda çalışma yapılmış bir konudur. Bu çalışmaların bir kısmında dağıtık sistemleri kullanarak gerçek zamanlı veriyi analiz edebilmek ve sistem ölçekleyebilmek gibi problemler de ele alınmıştır. (Kombi, Lumineau ve Lamarre, 2017; Farahabady, Samani, Wang, Zomaya ve Tari, 2016; Renart, Diaz-Montes ve Parashar, 2017; Papageorgiou, Poormohammady ve Cheng, 2016; De Matteis ve Mencagli, 2017; Basanta-Val, Garcia, Fernandez ve Fiesteus, 2017; Xu, Peng ve Gupta, 2016; Zhang, Li, Zhu ve Liu, 2016) çalışmalarında Apache Storm üzerinde operatör tıkanıklıkları ve gecikmeler göz önüne alınarak ele alınan sistemlerin ölçeklenmesi amaçlanmıştır. Bu çalışmalarda gerçekleştirilen işlemlerin ayrıntısı ise şu şekildedir: (Kombi, Lumineau ve Lamarre, 2017) çalışması sürekli sorguları göz önüne alarak ve aktif operatörlerin yakın gelecekteki metriklerini tahmin ederek, yerel ve global kapsamda paralelizm derecelerini değerlendirir ve değiştirir. Bu çalışmada, klasik Storm ile topoloji gecikmeleri karşılaştırılarak verimlilik analizi gerçekleştirilmiştir. (Farahabady, Samani, Wang, Zomaya ve Tari, 2016) çalışmasında “çokuzlu ulaşma oranı” ya da “her işçi düğümü için kaynak kullanımı” gibi metriklere bakılarak Storm için varsayılandan farklı bir zamansal programlayıcı geliştirilmiştir. Sistemin verimliliği klasik Storm ile arasındaki servis kalitesini düşürmeden, kaynak kullanımı oranlarına bakılarak gösterilmiştir. (Renart, Diaz-Montes ve Parashar, 2017)'de kullanıcılara programatik olarak iş akışı tasarlanmasına olanak sağlanmıştır ve sistemin akan verinin içeriğine göre reaktif bir davranış sergilemesi amaçlanmıştır. Çalışmada, belirlenen konumlardaki kaynakların artırılarak bir kaldıraç etkisi yaratılması yaklaşımı ele alınmıştır. Bu şekilde gecikmelerin azaltılması sağlanmıştır. Bu yaklaşımın, tek bir konum ile birden fazla konumun karşılaştırılması yapılarak verimlilik sağladığı gösterilmiştir. (Papageorgiou, Poormohammady ve Cheng, 2016) çalışmasında etkileşim, veritabanı, en son gerçekleştirilen görevleri toplayan bir modül, karar verici algoritmayı sağlayan bir modül ve topolojiyi güncelleyen bir modülden oluşan entegre bir çatı tasarlanmıştır. Tasarlanan algoritma kural tabanlı bir yapıda çalışmaktadır. Görev örneklemeleri sayıları karşılaştırılarak verimlilik

gösterilmiştir. (De Matteis ve Mencagli, 2017) çalışmasında servis kalitesini bozmadan, daha düşük maliyetli çözümler bulunması amaçlanmıştır. Tahmini model temelli bir yaklaşım kullanılarak, sentetik ve gerçek veri kümeleri ile gecikmeler gösterilmiştir. (Basanta-Val, Garcia, Fernandez ve Fiesteus, 2017) çalışmasında, operatör sıralama, yük dengeleme, algoritma seçilimi gibi birçok optimizasyon tekniği ve akan veri işleme örgüsü kullanılarak sentetik yüklerle uygulamalar test edilmiştir. Bu makalede sonuç olarak çalışma örüntüleri üzerinden kaynak kullanım oranları ve maksimum giriş frekansına ve cevap zamanına bakılarak verimlilikler karşılaştırılmıştır. (Xu, Peng ve Gupta, 2016) çalışmasında Storm topolojisindeki operatörlerin giriş oranlarına bakılarak bir tıkanıklık metriği hesaplanmıştır. Tıkanıklığın belli bir eşik değerinin üstüne çıkması halinde ölçekleme gerçekleştirilmiştir. Önerilen sistem, farklı topoloji tipleri üzerinde deneyerek, birim zamanda ortalama çıktı sayısına göre verimlilik gösterilmiştir. (Zhang, Li, Zhu ve Liu, 2016) çalışmasında, düğümler arasındaki iç trafik ve işçi düğümler arasındaki etkili olmayan yük dengeleri gözönüne alınarak Storm için varsayılandır farklı bir zamanlama algoritması çıkarılması amaçlanmıştır. Çalışma deneysel yüklerle gerçekleştirilmiş ve varsayılan zamanlama algoritması ile karşılaştırılmıştır. Ortalama gecikme ve ortalama düğüm içi trafik karşılaştırılarak verimlilik durumları gösterilmiştir.

Başka bir çalışmada, Apache Storm üzerinde statik bir kaynak konfigürasyonu ile topolojilerin önceliklendirilmesi ile sistem çıktısının nasıl arttırılabileceği üzerine çalışılmıştır (Chakraborty ve Majumdar, 2016). (Wang, Meng, Guo, Weng ve Yang, 2017; Khoshkbarforousha, Ranjan, Gaire, Abbasnejad ve Wang, 2016; Li, Tang ve Xu, 2016) çalışmalarında Apache Storm üzerinde makine öğrenmesi yaklaşımları kullanılarak kaynakların verimliliğinin arttırılması hedeflenmiştir. Bu çalışmalarda gerçekleştirilen işlemlerin ayrıntısı ise şu şekildedir: (Wang, Meng, Guo, Weng ve Yang, 2017) çalışmasında dağıtımlar için; veri seviyesinde, sorgu planlama seviyesinde, operatör seviyesinde ve küme seviyesinde dört seviye özellik çıkartımı ile birlikte artımlı makine öğrenmesi teknikleri kullanılmıştır. Farklı sorgular, farklı iş yükleri için farklı veri setleri ile eğitilen bir model kullanılarak yeni bir iş yükü geldiğinde ne kadar kaynak tüketileceği makine öğrenmesi yöntemleri ile tahmin edilmiştir. Bu yaklaşım SPS-Storm ile test edilmiştir. (Khoshkbarforousha, Ranjan, Gaire, Abbasnejad ve Wang, 2016)'da olasılık yoğunluğu fonksiyonları kullanılarak, kaynak kullanımı için bir gösterge

oluşturması amaçlanmıştır. (Wang, Meng, Guo, Weng ve Yang, 2017) çalışmasında olduğu gibi burada da geçmişe dayalı sistem kayıtları göz önüne alınarak makine öğrenmesi teknikleri ile model eğitilmiştir. Bu amaçla Yapay Sinir Ağları'nın özel bir türü olan Karışık Yoğunluk Ağları kullanılmıştır. Tahmin edilen ve gerçekte olan işlemci yükü ve hafıza kullanımı oranlarına bakılarak verimlilik gösterilmiştir. (Li, Tang ve Xu, 2016) çalışmasında ortalama çokuzlu işleme zamanı tahmin edilmek istenmiştir. İş parçacıkları arasındaki gecikme temel ağırlıklardan biri olarak alınmıştır. Eğitilmiş öğrenme algoritması kullanılarak, işlemci yükü, hafıza, iş parçacığı yükü gibi özellikler ile sistemin bir öğrenme sürecinden geçmesi sağlanmıştır. Kelime sayma iş yükü kullanılarak topoloji oluşturulmuştur ve ortalama çokuzlu işleme zamanı ile sistemin verimliliği gösterilmiştir.

(Runsewe ve Samaan, 2017) çalışmasında Hidden-Markov modeli kullanılarak değişen yüklerle adapte olabilen bir sistem tasarımı önerilmiştir. (Buddhika, Stern, Lindburg, Ericson ve Pallickara, 2017; HoseinyFarahabady, Zomaya ve Tari, 2017; Qian, Shen, Qin, Yang D., Yang Y. ve Wu, 2016) çalışmalarında ise kaynakların kullanımına göre işlerin önceliklendirilmesine dayalı bir sistem tasarımı amaçlanmıştır. Bu çalışmalarda gerçekleştirilen işlemlerin ayrıntısı ise şu şekildedir: (Buddhika, Stern, Lindburg, Ericson ve Pallickara, 2017) çalışmasında geliştirilen algoritma gözlemeleme, analiz, planlama ve yürütme olarak dört fazdan oluşmaktadır. Duruma müdahale etmek için çeşitli hesaplamalar yapılmıştır ve bu hesaplamalara "tahmin yüzükleri" adı verilmiştir. Bu yüzükler, analiz ve planlama aşamalarında kullanılmaktadır. Çalışmada, mesaj başına işleme zamanı, mesaj boyutu, bant genişliği kullanımı gibi veriler kullanılmış ve işlemci ve hafıza kullanımları karşılaştırılarak diğer sistemlere göre önerilen sistemin avantajları gösterilmiştir. (HoseinyFarahabady, Zomaya ve Tari, 2017) çalışması konuya bir optimizasyon problemi olarak ele almıştır ve maliyet fonksiyonları ile oluşan bir hesaplama kümesi ile tüm makinelerde servisin kalitesini düşüren olayları azaltmak, işlemci kullanımını kabul edilebilecek bir aralıkta tutmak, aynı kümeyi paylaşan uygulamaların birbirini etkilemesini azaltmak gibi hedefleri mevcuttur. Çalışma sonucunda önerilen sistem ile birlikte kümenin kaynak kullanımı artmış, ortalama çokuzlu gecikmeler ve servisin kalitesini bozacak olaylar azalmıştır. (Chen, Zhang ve Jin, 2017) çalışmasında Apache Storm kaynaklarıyla daha verimli çalışmak için Apache Zookeeper metrikleri göz önüne alınarak alternatif bir önceliklendirme sistem tasarımı geliştirilmiştir. (Liu ve Buyya, 2017; Wang,

Tari, HoseinyFarahabady ve Zomaya, 2017) çalışmalarında işlenen veri seti üzerinde, verinin anahtarlandığı durumlarda, anahtarlamayı göz önüne alarak popüler anahtarları önceliklendirerek sistemin değişken durumlara adapte olabilmesi amaçlanmıştır. Bu çalışmalarda gerçekleştirilen işlemlerin ayrıntısı ise şu şekildedir: (Liu ve Buyya, 2017) çalışmasında süreç dağıtım aşamasından bakılmıştır. P-Deployer adında bir dağıtım planlayıcı modül oluşturulmuştur. Akan veri işleyen dağıtık sisteme gelen veri hızını dinleyerek, görevler için profiller oluşturulmaya çalışılmıştır. Bu profiller ile birlikte kaynak öngörüsü yapılarak operatörler ölçeklenmeye çalışılmıştır. Kelime sayma iş yükü kullanılarak sentetik bir yük oluşturulmuştur ve sistemin birim zamanda verdiği çıktıya göre verimlilik avantajları incelenmiştir. Sonuç olarak, literatürdeki bu çalışmalar ve benzer çalışmaların birçoğunun akan veri sistemlerini Apache Storm platformu üzerinde gerçekleştirildiği ve ölçeklendirme probleminin farklı yöntemler ve metrikler kullanılarak ele alındığı gözlemlenmiştir.

Özetle, her geçen gün hızla artan akan veri kaynaklarını işleyebilmek büyük önem kazanmıştır. Akan veriyi işleyen sistemler veriyi gerçek zamanlı olarak işlediği için sistem çok çeşitli senaryolara adapte olabilmelidir. Kaynakların optimize edilip, çıktının artırılması temel amaçtır ve akan veriyi işleyen dağıtık sistemlerde dinamik bir ölçeklemeye gereksinim vardır. Bu çalışmada bu gereksinimler göz önüne alınarak Apache Flink üzerinde dinamik ölçeklemeye olanak sağlayan, entegre olabilir bir harici sistem oluşturulması hedeflenmiştir. Bu hedef doğrultusunda akan veri işleyen dağıtık bir sistemin metrikleri Java Management Extensions aracılığıyla, Akka kullanılarak Kafka üzerinde toplanmış ve Apache Flink kullanılarak, çeşitli hesaplamalarla sistemde meydana gelmesi beklenen gecikmede göz önüne alınarak sistemin ölçeklenmesi gerektiği durumlar belirlenmiştir. Aynı zamanda, Kuyruk Teorisi kullanılarak belirlenen veri trafik yoğunluğuna bağlı olarak veri hacminin artırılıp azaltılabildiği gecikmeye duyarlı bir model önerilmiştir. Literatürde yapılan daha önceki çalışmalardan farklı olarak önerilen sistem, akan veri işleyen dağıtık sistemlerden bağımsız, soyut bir şekilde birçok sisteme entegre olabilecek ve ölçekleme işlemini Apache Flink üzerinde gerçekleştirebilecek şekilde tasarlanmıştır. Aynı zamanda bağımsız metrik toplama modülüne de sahiptir. Bu sistemin en önemli özelliklerinden bir tanesi de tüketen tarafın kritik metriklerini kullanarak gecikmeler, işlemci ve hafıza kullanımları, giren ve çıkan veri hacmi gibi birçok veri ölçeklemenin gerçekleştirilmesi için kullanılabilmesidir. Aynı

zamanda soyut bir yapı olduğu için birçok dağıtık akan veri işleme sistemi ile kolayca entegre olabilir. Bu özellikler ile önerilen ve geliştirilen model hem performans hem de maliyet açısından dağıtık akan veri işleyen bir sistemi olduğundan daha verimli çalışan bir sistem haline getirmektedir.

Çalışmanın ilerleyen bölümleri şu şekilde organize edilmiştir: İkinci bölümde, dağıtık akan veriyi işlemek için kullanılan platformlar ve önerilen sistem mimarisi, bu mimariyi oluşturan bileşenler ve nasıl bir yöntemle ölçeklemenin gerçekleştirileceği ele alınmıştır. Üçüncü bölümde, akan veriyi işleyen dağıtık bir sistemin benzetim çalışması yapılarak sistemin hangi durumlarda ölçeklendiği ve ölçeklemeden sonraki durumları analiz edilmiştir. Dördüncü bölümde ise sonuç ve gelecekte yapılması planlanan çalışmalardan bahsedilmiştir.

2. ANALİZ METODU

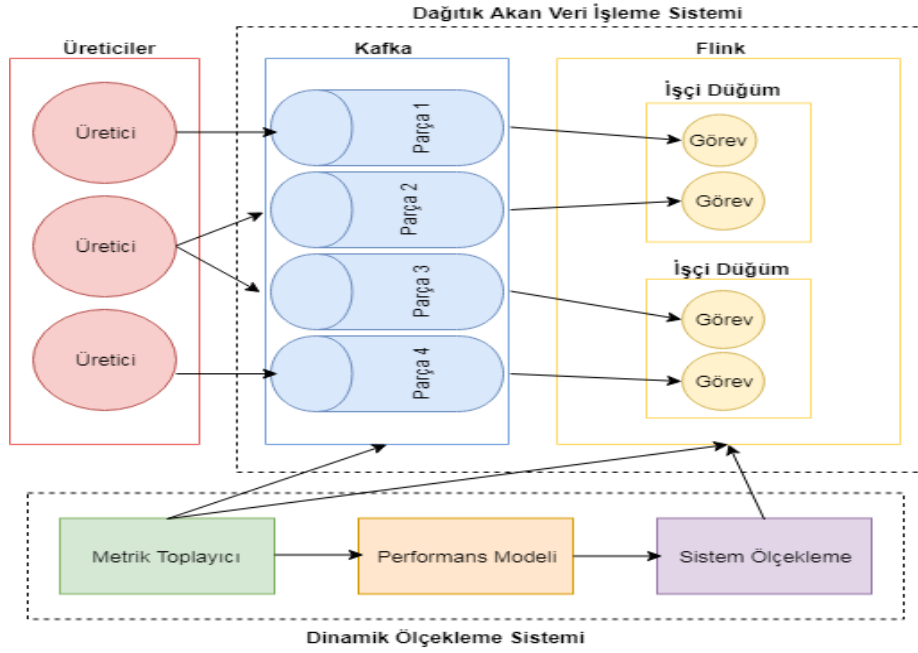
Bu çalışmada, Apache Flink hem geliştirme yapılacak sistem hem de ölçekleme metriklerini alıp hesaplama yapan bir bileşen olarak kullanılmıştır. Akka, Apache Flink metriklerini Apache Kafka'ya gönderen olarak konumlanmıştır. Apache Kafka ise sistem metriklerini toplayan ve dağıtan bir mesaj kuyruğu olarak kullanılmıştır.

Apache Flink açık kaynaklı bir dağıtık akan veri işleme çatısıdır. Akan veriyi işleyen çekirdek dağıtık sistem Java ve Scala ile geliştirilmiştir. Apache Flink düşük gecikmelerle yüksek verimlilik sağlar ve hataları tolere edebilir. Apache Kafka ise dağıtık bir akış platformudur. Akışların kayıtlarını yayınlayabilir ve bu özelliği ile bir mesaj kuyruğuna veya kurumsal mesajlaşma sistemine benzemektedir. Hataları tolere edebilir ve kalıcı bir şekilde akışların kayıtlarını saklayabilir.

2.1. Önerilen Sistem Modeli

Bu bölümde, çalışma kapsamında önerilen ve Şekil 1'de gösterilen sistem mimarisi ve ölçekleme modeli anlatılmaktadır. Oluşturulan bu yapı, literatürdeki mevcut modellere dayanmasının yanı sıra kullanılan yeni formülasyonlar ve yöntemlerle bu modelleri geliştirmektedir.

Şekil 1'de dağıtık Akan Veri İşleme Sistemi ile temsil edilen kısım ölçeklenecek sistemi göstermektedir. Burada işlenecek mesajların bulunduğu bir Kafka kümesi ve bu mesajları işleyen bir Apache Flink kümesi bulunmaktadır.



Şekil 1. Çalışma kapsamında önerilen sistem modeli

Tablo 1. Çalışmada kullanılan sistem metrikleri ve sabitler

Sembol	İsim	Formül	Açıklama
C	İşlemci yükü ortalama değeri	$\frac{1}{n} \left(\sum_{i=1}^n c_i \right)$	c , Java Sanal Makinesi'nde en son elde edilen işlemci yüküdür.
M	Hafıza kullanım oranının ortalama değeri	$\frac{1}{n} \left(\sum_{i=1}^n H_i \right)$	H hafıza kullanım oranıdır.
X	Saniyede gelen kayıt sayısının ortalama değeri	$\frac{1}{n} \left(\sum_{i=1}^n x_i \right)$	x saniyelik birim zamanda, seçilen operatör veya göreve ulaşan kayıt sayısıdır.
Y	Saniyede gelen kayıt sayısının standart sapma değeri	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$	-
R	Saniyede gelen kayıt sayısı eşik oranı	$\frac{Y}{X}$	-
A	Gecikme süresinin ortalama değeri	$\frac{1}{n} \left(\sum_{i=1}^n \tau_i \right)$	-
Q	Gecikme süresi eşik sabiti	-	Dışarıdan parametre olarak verilir, sabit bir değerdir.
B	Gecikme süresi eşik oranı	-	Eğer A değeri Q değerinden büyükse 0.5, küçükse 0 kabul edilir.
J	Lambda değerleri	-	C, M, R, B değerlerini içeren bir dizi
W	Lambda eşik sabiti	-	Dışarıdan parametre olarak verilir, sistemin ölçeklenmeye geçeceği limiti gösterir. Sabit bir değerdir.
λ	Lambda	$\frac{1}{n} \left(\sum_{i=1}^n J_i \right)$	Lambda değerlerinin ortalama değeri

Dinamik Ölçekleme Sistemi kısmında ise *metrik toplayıcı*, *performans ölçüm modeli* ve *sistem ölçekleme* olmak üzere üç temel modül yer almaktadır. *Metrik toplayıcı modülü*, Apache Flink kümesinin metriklerini belirli aralıklarla alır ve Kafka kümesine atar. Ölçekleme hesabının yapılacağı verilerin tamamı burada oluşur. Belirlenen bu kritik metrikler Kafka kümesine gönderildikten sonra bir Apache Flink kümesi bu metrikleri gerçek zamanlı olarak dinler ve veri parçaları üzerinde çeşitli hesaplamalar yapar. Bu hesaplamalar *performans ölçüm modelinde* gerçekleşir ve hesaplamaların sonucuna göre sistemin ölçeklenmesi gerekip gerekmediğine karar verilir. Sistemin ölçeklenmesinin gerektiği durumda *sistem ölçekleme modülü*, veriyi işleyen Apache Flink kümesine REST API ve HTTP protokolü aracılığıyla ölçekleme isteği gönderir. Sistemin ölçeklenmesini tetikleyecek temel kaynak Apache Flink metrikleri ve kritik hesaplamalardır. Apache Flink metrikleri toplama ve harici sistemlere almak için bir metrik sistemi sunar.

Önerilen modelin özelliklerinden bir tanesi de gecikmeye duyarlı bir ölçekleme sistemi sunmasıdır. Akan veri işleyen sistemlerde veri örneklerinin kuyrukta bekleme süreleri ile servis/işleme süreleri gecikmelere yol açmaktadır. Sistem, bu gecikmelerinde göz önüne alınarak veri akış trafiğinin değerlendirilmesi ile işlem hacmi azaltılacak ya da arttırılacak şekilde ölçeklendirilebilir. Kuyruk Teorisi (Botran, Alonso ve Lozano, 2014), sisteme gelen işler ile sistemden ayrılan işler arasındaki ilişkiyi modellemek için Internet uygulamaları ve sunucu modellemede yaygın olarak kullanılmaktadır. Bu teori, kuyruk uzunluğu veya isteklerin ortalama bekleme süresi gibi performans ölçümlerini tahmin etmek gibi görevler için elastik uygulamalarda dinamik ölçeklendirme için de kullanılabilir. Kuyruk sisteminin performansını değerlendirmek için kullanılan en temel ölçümlerden bir tanesi trafik yoğunluğudur. Bu çalışmada da önerilen modelde trafik yoğunluğunu hesaplamak için Kuyruk Teorisi'nden yararlanılmıştır. Model içerisinde her bileşen bir kuyruk sistemi olarak düşünülebilir. t değişkeni, ardışık iki gelen veri örneği arasındaki zaman aralığını ve u ise gelen bir veri örneğinin servis/işleme süresini göstermek üzere trafik yoğunluğu (α) aşağıdaki şekilde hesaplanabilir :

$$\alpha = \frac{E(u)}{E(t)} \quad (1)$$

t değişkeni rastgele olarak seçilirken, u değişkeninin tüm varış zamanları için bağımsız ve aynı şekilde

dağılmış olduğu kabul edilir. Denklem 1'de $E(u)$ ortalama servis süresini, $E(t)$ ise ortalama varışlar arası geçen süreyi göstermektedir. Buna göre, trafiğin sunuculara eşit olarak dağıtıldığı varsayıldığında z adet sunucu için, sunucu kullanımını ve her sunucunun meşgul olduğu sürenin ortalama oranı denklem 2 ile gösterilebilir:

$$\delta = \frac{\alpha}{z} = \frac{E(u)}{zE(t)} \quad (2)$$

Dolayısıyla, α trafik yoğunluğu, gerekli sunucu sayısı (z) ve tıkanıklığın (δ) bir ölçüsüdür. Modelde $\delta \rightarrow 1$ ve $\alpha \rightarrow z$ alınarak sunucu sayısı tıkanıklığı ortadan kaldırmak için gereken örnek sayısına çevrilir. Her örnek tek sunuculu G/G/1 kuyruk sistemi olarak ele alınmıştır. Bu kuyruk sisteminde G, hem varış süreleri hem de servis/işleme sürelerindeki genel dağılımı temsil etmektedir. 1 ise modelin tek sunuculu bir sistem olduğunu gösterir. Farklı varış ve işleme süreleri birbirinden bağımsız olarak düşünülür. Buna göre, $\gamma = \frac{1}{E(t)}$ ortalama varış oranını, $\mu = \frac{1}{E(u)}$ servis oranını ve $x \in \{t, u\}$ olmak üzere $z_x = \frac{\sqrt{\text{Var}(x)}}{E(x)}$ varyasyon katsayısını göstermektedir. $\delta = \alpha = \gamma/\mu$ ve $\delta \rightarrow 1$ olması durumunda yüksek veri akışının olduğu yoğun trafik altındaki bir kuyruk sistemi için, Kingman Formülü (Kingman, 1962) kullanılarak, n veri örneği sayısını göstermek üzere, bir örneğin kuyrukta harcaması beklenen ortalama süre denklem 3 ile hesaplanabilir:

$$T_n = \left(\frac{\delta}{1-\delta} \right) \left(\frac{z_t^2 + z_u^2}{2} \right) E(u) \quad (3)$$

Sonuç olarak, ρ mevcut veri örneği çıktı süresini göstermek üzere sistemde bir örnek için meydana gelmesi beklenen gecikme denklem 4 ile hesaplanır:

$$\tau_n = T_n + (u - p) \quad (4)$$

Sistemin toplam çalışma süresi sistem yükünü göstermektedir ve kuyruk süresi de zamanla çalışma süresine yakınsama eğilimindedir. Kuyruk süresinin çalışma süresinden daha fazla olduğu durumlar veri trafiğinin düşük olduğu şekilde yorumlanabilir. Bu tip senaryolarda, kuyruk süresi çalışma süresine yaklaşıp dek işlem hacmi arttırılacak şekilde sistem ölçeklenebilir. Aksi durumda işlem hacmi azaltılarak sistemde meydana gelebilecek tıkanıklıklar önenebilir. Geliştirdiğimiz sistemin ölçeklenmesine karar vermek için Kuyruk Teorisi ile birlikte çeşitli matematiksel ve

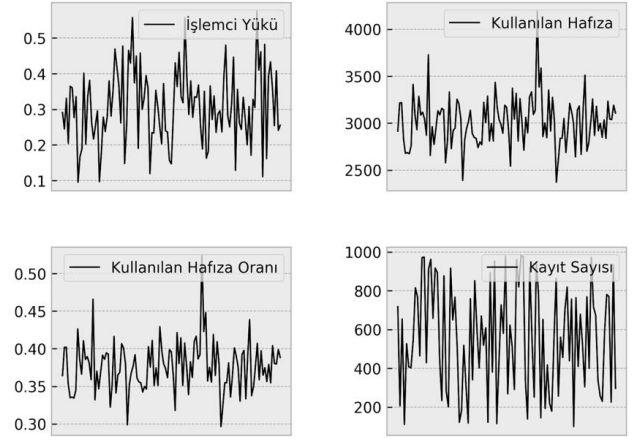
istatistiksel hesaplamalar da yapılmıştır. Bu çalışmada kullanılan sistem metrikleri ve sabitler ile bunlara ait hesaplamalar Tablo 1’de açıklanmıştır.

Tüm bu hesaplamaların sonucunda gecikmeye duyarlı λ değeri performans ölçüm modeli içerisinde değerlendirilir. λ değeri, W sabitinden büyükse ölçekleme işlemi tetiklenir. Önerilen bu dinamik model, halihazırda çalışan uygulamalarda sistem iş yükünün ölçeklendirilmesi için kullanılmasının yanı sıra iş yükünü tahmin ederek ölçeklendirme işlemi tetikleyen modellere de entegre edilebilir. Modelin ön plana çıkan diğer önemli bir özelliği de sistemlerin servis/işleme oranlarını kontrollü bir performans ve kalite kaybıyla önemli ölçüde artırabilmesidir.

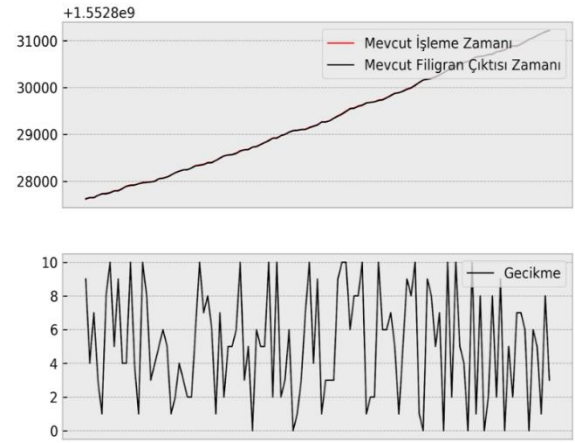
3. ANALİZ

Bu bölümde, önerilen gecikmeye duyarlı ölçekleme modelinin etkinliği, Apache Flink platformu üzerinde Scala programlama dili kullanılarak gerçekleştirilen çeşitli benzetim çalışmaları ile gösterilmiştir. Yapılan benzetim çalışmalarında, Q sabiti 30, W sabiti 0,6 olarak alınmıştır. Sabit değerler keyfi belirlenmiştir. Q sabiti gecikmelere karşı sistemin duyarlılığını saniye cinsinden, W sabiti sistemin tüm eşik dengesinin ne kadar olacağına göre belirlenebilir. W sabiti, 0 ile 1 arasında bir değer kabul eder. Benzetim çalışmasında kullanılan kritik metrikler işlemci yükü, hafıza kullanımı, gelen kayıt sayısı, mevcut işleme zamanı, mevcut çıktı filigranı ve gecikme değerleri olarak belirlenmiştir. Örnek veri setleri olasılık dağılımları kullanılarak sentetik bir yük üretilerek oluşturulmuştur.

Şekil 2a’da verinin geliş sırasıyla birinci pencerede gelen işlemci yükü, kullanılan hafıza, kullanılan hafızanın toplam hafızaya oranı ve saniyede gelen kayıt sayısı metrikleri görülmektedir. Şekil 2b sistemin gelen veriyi işleme zamanı ile filigran oluşturduğu zamanı göstermektedir. Daha önce de bahsedildiği gibi ikisi arasındaki fark ve kuyrukta bekleme süresi gecikmeyi vermektedir ve bu veri de yine Şekilde 2b’de gösterilmektedir. Şekil 2a ve Şekil 2b çizelgelerini oluşturan ve sisteme giriş olarak verilen girdilerin örnek bir kesiti ise Tablo 2’de verilmiştir. Çizelgelerde ve tablolarda bulunan veriler giriş verilerini temsil etmektedir. Tablolarda verinin kesiti gösterilmektedir.



Şekil 2a. Birinci iterasyon için işlemci yükü, kullanılan hafıza, kullanılan hafıza oranı ve kayıt sayısı



Şekil 2b. Birinci iterasyon için mevcut işleme zamanı, mevcut filigran çıktısı zamanı ve gecikme değerleri

Tablo 2’de gösterilen veri örnekleri ve metrikler için λ değeri W sabitinden küçük olduğu için ölçeklemeye ihtiyaç yoktur.

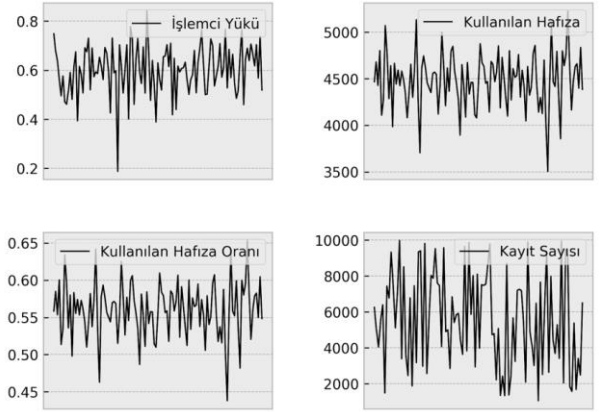
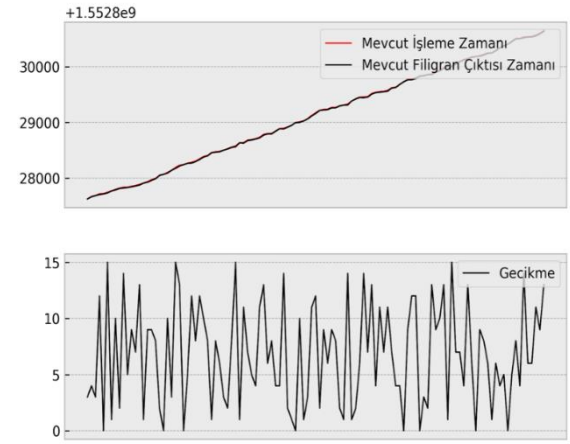
Bir önceki iterasyonda olduğu gibi Şekil 3a’da verinin geliş sırasıyla ikinci pencerede gelen işlemci yükü, kullanılan hafıza, kullanılan hafızanın toplam hafızaya oranı ve saniyede gelen kayıt sayısı metrikleri görülmektedir. Şekil 3b sistemin gelen veriyi işleme zamanı ile filigran oluşturduğu zamanı ve gecikmeyi göstermektedir. Şekil 3a ve Şekil 3b çizelgelerini oluşturan ve sisteme giriş olarak verilen girdilerin örnek bir kesiti ise Tablo 3’te verilmiştir. Buradan birim zamanda gelen kayıt sayısının artışına bağlı olarak, sistemin işlemci yükü ve hafıza kullanımının gözle görülür bir şekilde arttığı görülmektedir. Fakat Tablo 3’de görüleceği üzere λ değeri halen W sabitinden küçük olduğu için ölçeklemeye ihtiyaç yoktur; dolayısıyla ölçekleme tetiklenmez.

Tablo 2. Birinci hesaplama iterasyonu için veri örnekleri

Metrik İsmi	Değer
İşlemci yükü	[0,29081359 0,24484373 0,33099842 0,20422209 0,36440966 0,36044093]
Kullanılan hafıza	[2915,05236578 3215,18824778 3217,2999415 2826,96742778]
Kalan hafıza oranı	[0,36438155 0,40189853 0,40216249 0,35337093 0,33482182 0,33580286]
Saniyede kalan gelen kayıt sayısı	[718, 207, 654, 111, 527, 408, 402, 548, 815, 767, 464, 971, 974, 429, 912, 962, 658, 917]
Gecikme (τ)	[9, 4, 7, 3, 1, 8, 10, 5, 9, 4, 4, 10, 4, 1, 10, 8, 3, 4, 5, 6, 5, 1, 2, 4, 3, 2, 2, 6, 10]
Lambda değerleri (J)	[0,3075763493045144 0,37675745680472644]
λ	0,2926817375716746

Tablo 2’de gösterilen veri örnekleri ve metrikler için λ değeri W sabitinden küçük olduğu için ölçeklemeye ihtiyaç yoktur.

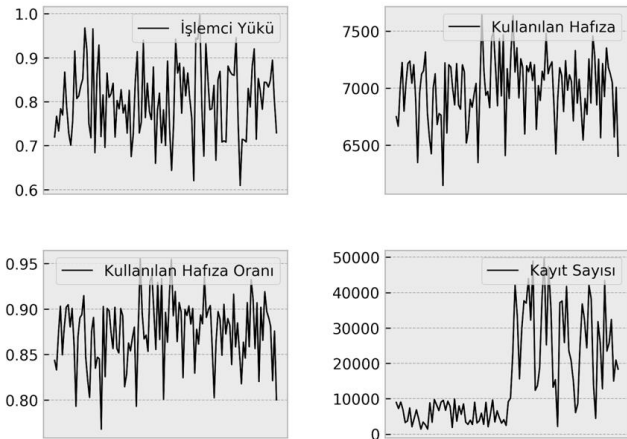
Bir önceki iterasyonda olduğu gibi Şekil 3a’da verinin geliş sırasıyla ikinci pencerede gelen işlemci yükü, kullanılan hafıza, kullanılan hafızanın toplam hafızaya oranı ve saniyede gelen kayıt sayısı metrikleri görülmektedir. Şekil 3b sistemin gelen veriyi işleme zamanı ile filigran oluşturduğu zamanı ve gecikmeyi göstermektedir. Şekil 3a ve Şekil 3b çizelgelerini oluşturan ve sisteme giriş olarak verilen girdilerin örnek bir kesiti ise Tablo 3’te verilmiştir. Buradan birim zamanda gelen kayıt sayısının artışına bağlı olarak, sistemin işlemci yükü ve hafıza kullanımının gözle görülür bir şekilde arttığı görülmektedir. Fakat Tablo 3’de görüleceği üzere λ değeri halen W sabitinden küçük olduğu için ölçeklemeye ihtiyaç yoktur; dolayısıyla ölçekleme tetiklenmez.

**Şekil 3a.** İkinci iterasyon için işlemci yükü, kullanılan hafıza, kullanılan hafıza oranı ve kayıt sayısı**Şekil 3b.** İkinci iterasyon için mevcut işleme zamanı, mevcut filigran çıktısı zamanı ve gecikme değerleri**Tablo 3.** İkinci hesaplama iterasyonu için veri örnekleri

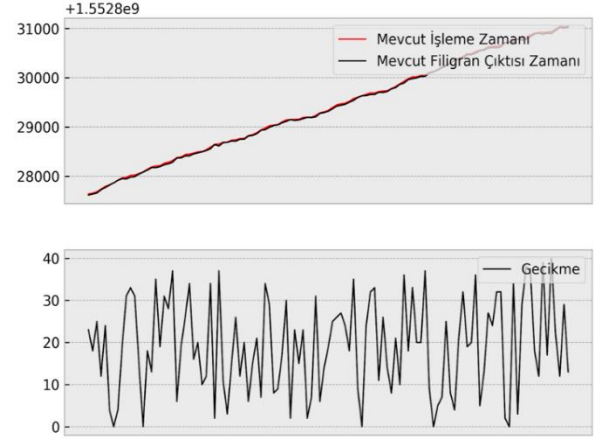
Metrik İsmi	Değer
İşlemci yükü	[0,74792182 0,67820917 0,63737983 0,56141087 0,49553547 0,57619555]
Kullanılan hafıza	[4467,48624105 4679,77961699 4427,77417013 4802,37689798]
Kalan hafıza oranı	[0,55843578 0,58497245 0,55347177 0,60029711 0,51352895 0,53401568]
Saniyede kalan gelen kayıt sayısı	[6248, 4977, 4042, 5477, 6393, 1481, 7435, 6772, 9323, 7411, 5108, 7106, 9981, 3385]
Gecikme (τ)	[3, 4, 3, 12, 0, 15, 1, 10, 2, 14, 5, 9, 7, 13, 1, 9, 9, 8, 2, 0, 10, 3, 15, 13, 0, 5, 12, 8]
Lambda değerleri (J)	[0,6022923124276031 0,560108109236605 0,4695572727535386 0]
λ	0,4079894236044367

Bir önceki iterasyona benzer şekilde Şekil 4a'da verinin geliş sırasıyla üçüncü pencerede gelen işlemci yükü, kullanılan hafıza, kullanılan hafızanın toplam hafızaya oranı ve saniyede gelen kayıt sayısı metrikleri görülmektedir. Şekil 4b'de sistemin gelen veriyi işleme zamanı ile filigran oluşturduğu zamanı ve gecikmeyi göstermektedir. Şekil 4a ve Şekilde 4b çizelgelerini oluşturan ve sisteme giriş olarak verilen girdilerin örnek bir kesiti ise Tablo 4'te verilmiştir. Birim zamanda gelen kayıt sayısının anormal bir şekilde arttığı ve tepe yaptığı görülmektedir. Bu artışa bağlı olarak, sistemin işlemci yükü ve hafıza kullanımı sınırları zorlayacak şekilde artmıştır. Bunun yanında gecikmeler de birim zamanda gelen kayıt sayısını doğru zamanda işlemek için çaba sarf etmektedir; fakat yüksek gecikmeler de oluşmaya başlamıştır. Tablo 4'te görüleceği üzere λ değeri bu iterasyonda W sabitinden büyük olduğu için sisteme ölçkleme isteği gönderilir.

Sistem ölçeklendikten sonra sistemin durumu, işlemci yükü, kullanılan hafıza, kullanılan hafıza oranı ve birim zamanda gelen kayıt sayısı Şekil 5a'daki gibidir. Bu şekilden görüldüğü gibi birim zamanda gelen kayıt sayısı aynı davranışı göstermektedir. Birim zamanda gelen kayıt sayısına rağmen belirli bir süre sonra işlemci yükü, kullanılan hafıza ve kullanılan hafıza oranı düşmektedir. Bu duruma bağlı olarak da Şekil 5b'de görüldüğü gibi gecikmeler de kabul edilebilir bir noktaya gelmiştir. Şekil 5a ve Şekil 5b çizelgelerini oluşturan ve sistemin mevcut durumdaki metriklerinin örnek bir kesiti ise Tablo 5'de verilmiştir.



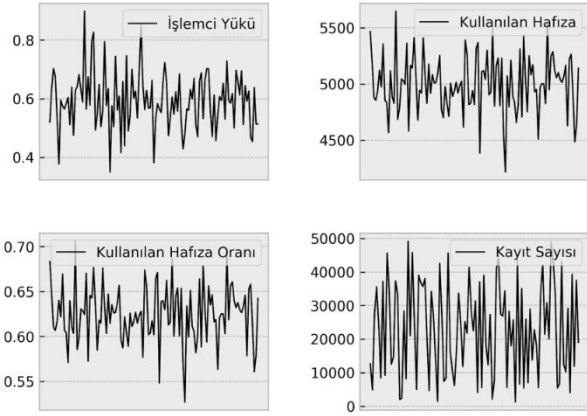
Şekil 4a. Üçüncü iterasyon için işlemci yükü, kullanılan hafıza, kullanılan hafıza oranı ve kayıt sayısı



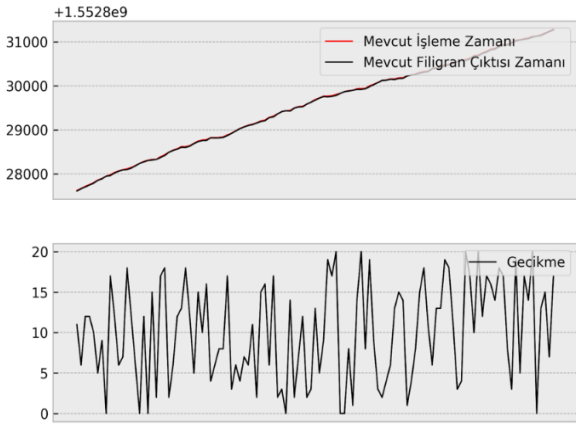
Şekil 4b. Üçüncü iterasyon için mevcut işleme zamanı, mevcut filigran çıktısı zamanı ve gecikme değerleri

Tablo 4. Üçüncü hesaplama iterasyon için veri örnekleri

Metrik İsmi	Değer
İşlemci yükü	[0,71975049 0,76689485 0,7333771 0,78396634 0,76981594 0,86728982]
Kullanılan hafıza	[6748,09165228 6664,69131265 6994,82196045 7223,84370897]
Kalan hafıza oranı	[0,84351146 0,83308641 0,87435275 0,90298046 0,84965956 0,88083314]
Saniyede kalan gelen kayıt sayısı	[8953, 7201, 9061, 6826, 3234, 3706, 7406, 2080, 4420, 6853, 4224, 1388, 3376, 2597]
Gecikme (τ)	[23, 18, 25, 12, 24, 4, 0, 4, 19, 31, 33, 31, 15, 0, 18, 13, 35, 19, 31, 28, 37, 6, 19, 26]
Lambda değerleri (J)	[0,8027545722795377 0,8746252806617266 0,8664009685964649 0]
λ	0,6359452053844323



Şekil 5a. Ölçeklemeden sonra işlemci yükü, kullanılan hafıza, kullanılan hafıza oranı ve kayıt sayısı



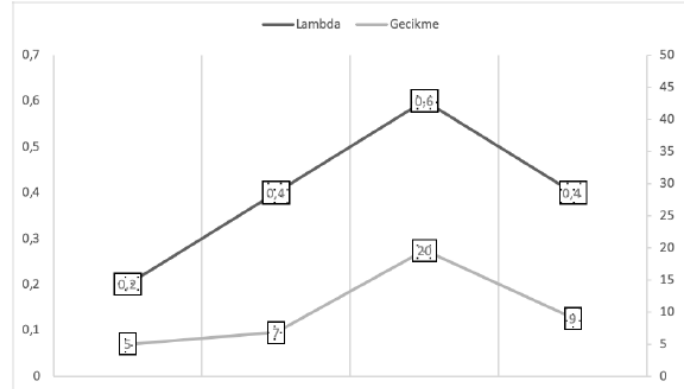
Şekil 5b. Ölçeklemeden sonra mevcut işleme zamanı, mevcut filigran çıktısı zamanı ve gecikme değerleri

Tablo 5. Sistem ölçeklendikten sonra veri örnekleri

Metrik İsmi	Değer
İşlemci yükü	[0,52061993 0,62960358 0,70270973 0,67735602 0,53360672 0,37741859]
Kullanılan hafıza	[5466,4133859 5189,29183283 4881,01291935 4854,00134743]
Kalan hafıza oranı	[0,68330167 0,64866148 0,61012661 0,60675017 0,61677137 0,6403014]
Saniyede kalan gelen kayıt sayısı	[12623, 4870, 27608, 35565, 24276, 8399, 37170, 9201, 45611, 34264, 12502, 14661]
Gecikme (τ)	[11, 6, 12, 12, 10, 5, 9, 0, 17, 12, 6, 7, 18, 12, 6, 0, 12, 0, 15, 2, 17, 18, 2, 6, 12, 13]
Lambda değerleri (J)	[0,5935922476790889 0,6245944830883645 0,5973252562509304 0]
λ	0,45387799675459595

Böylece, sistem için ölçekleme isteği gönderildikten sonra Apache Flink kümesi gecikmeye duyarlı λ değerinin küçülmesi ile birlikte daha etkili bir şekilde çalışır hale gelmiş ve sistemde meydana gelebilmesi muhtemel tıkanıklığın önüne geçilmiştir.

Şekil 6'da λ değeri ile gecikme arasındaki ilişki gösterilmiştir. Buradan görüldüğü gibi 0,6 değerine ulaşan λ değeri, ölçekleme işlemi gerçekleştiği için 0,4 değerine düşmüştür. λ 'nın bu değişimine korele bir şekilde gecikmeler artmakta ve azalmaktadır. Ölçekleme isteği gönderildikten sonra Apache Flink kümesi λ değerinin küçülmesi ile birlikte işlemci kullanımı normal seviyelere gelmiş, hafızayı daha etkin bir şekilde kullanmaya başlamış ve birim zamanda gelen kayıt sayısı artmasına rağmen sistem işleme zamanı, olayın gerçekleşme zamanına yaklaşmıştır. Bunun sonucunda hem kaynak kullanımı optimize edilmiş hem gecikmelerin önüne geçilmiştir.



Şekil 6. Lambda ve gecikme arasındaki ilişki grafiği

4. SONUÇLAR VE TARTIŞMALAR

Kurumlar müşterileri, uygulamaları ve ürünleri daha iyi tanımak ve anlamak için akan veriyi işleme ve analiz etme ihtiyacı duymaktadır. Yığın halinde verinin analiz edilmesine göre, gerçek zamanlı veriyi işlemek çok daha hızlı karar alınmasını ve günün şartlarına hızlı adapte olabilmeyi sağlamaktadır. Fakat gerçek zamanlı akan veriyi işleyen sistemleri inşa etmek ve operasyonunu sağlamak oldukça zordur. Maliyeti de göz önünde bulundurarak sistemin gerçek zamanlı bir şekilde çalışmaya devam edebilmesi için dağıtık sunucuların yönetilebilmesi ve gerektiğinde kolayca ölçeklenebilmesi gerekmektedir.

Bu çalışmada akan veriyi işleyen dağıtık sistemlerin, değişen iş yüklerine adapte olması adına ölçeklenebilirlik problemlerine odaklanarak, Apache Flink üzerinde gecikmeye duyarlı ölçeklendirme

yapabilen bir model önerilmektedir. Önerilen model ile tüketen tarafın kaynakları ve çalışan iş özetindeki metrikleri kullanılarak sistemin o an ölçeklenmesi gerekip gerekmediğine karar verilmesine olanak sağlanmıştır. Bunun yanında, ölçekleme sürecinin daha soyut bir şekilde yapılması ve farklı akan veriyi işleyen dağıtık sistemlere de kolay entegre olabilmesi amaçlanmıştır. Ayrıca, model içerisinde trafik yoğunluğu ve kuyrukta bekleme süreleri, Kuyruk Teorisi ve Kingman Formülü ile hesaplanarak sistemde meydana gelmesi beklenen gecikmelere göre tıkanıklıkların önlenmesi de sağlanmıştır. Sonuç olarak, akan veriyi işleyen dağıtık sistemlerde kullanılabilecek bir sistem önerisinde bulunularak, bu sistemin etkin bir şekilde çalıştığı çeşitli benzetim çalışmaları ile desteklenerek gösterilmiştir.

Gelecek çalışmalarda hesap yapılan metrik verilerini özelleştirip türeterek metrik sayısını arttırmak, tüketen tarafın metriklerini de hesaba katmak ve metrik verilerinin belirli bir matematiksel ve istatistiksel hesap yerine, makine öğrenmesi algoritmaları kullanılarak daha etkin bir ölçekleme stratejisi oluşturmak gibi planlamalar mevcuttur. Bundan sonraki aşamada, mevcut çalışmaya planlar dahilinde belirlenen yeni özellikler sırasıyla eklenerek sistemin geliştirilmesi üzerine çalışılacaktır.

KAYNAKLAR

Basanta-Val. P., Garcia N., Fernandez L., Fiesteus J. 2017. Patterns for Distributed Real-Time Stream Processing, IEEE Transactions on Parallel and Distributed Systems, 28(11), 3243-3257.

Botran T.L., Alonso J.M., Lozano J.A. 2014. A Review of Auto-scaling Techniques for Elastic Applications in Cloud Environments, Journal of Grid Computing, 12(4), 559-592.

Buddhika T., Stern R., Lindburg K., Ericson K., Pallickara S. 2017. Online Scheduling and Interference Alleviation for Low-Latency, High-Throughput Processing of Data Streams, IEEE Transactions on Parallel and Distributed Systems, 28(12), 3553-3569.

Chakraborty R., Majumdar S. 2016. A Priority Based Resource Scheduling Technique for Multitenant Storm Clusters, International Symposium on Performance Evaluation of Computer and Telecommunication Systems, pp1-6, 24-27 Temmuz, Kanada.

Chen H., Zhang F., Jin H. 2017. Popularity-aware Differentiated Distributed Stream Processing on

Skewed Streams, IEEE 25th International Conference on Network Protocols, pp1-10, 10-13 Ekim, Kanada.

De Matteis T., Mencagli G. 2017. Elastic Scaling for Distributed Latency-sensitive Data Stream Operators, 25th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, pp61-68, 6-8 Mart, Rusya.

Farahabady M.R.H., Samani H.R.D., Wang Y., Zomaya A.Y., Tari Z. 2016. A QoS-Aware Controller for Apache Storm, IEEE 15th International Symposium on Network Computing and Applications, pp334-342, 31 Ekim-2 Kasım, ABD.

HoseinyFarahabady M., Zomaya A.Y., Tari Z. 2017. QoS- and Contention- Aware Resource Provisioning in a Stream Processing Engine, IEEE International Conference on Cluster Computing, pp137-146, 5-8 Eylül, ABD.

Khoshkbarforousha A., Ranjan R., Gaire R., Abbasnejad E., Wang L. 2016. Zomaya A.Y., Distribution Based Workload Modelling of Continuous Queries in Clouds, IEEE Transactions on Emerging Topics in Computing, 5(1), 120-133.

Kingman J.F.C. 1962. On queues in heavy traffic, Journal of the Royal Statistical Society. Series B (Methodological), 24(2), 383-392.

Kombi R.K., Lumineau N., Lamarre P. 2017. A preventive auto-parallelization approach for elastic stream processing, IEEE 37th International Conference on Distributed Computing Systems, pp1532-1542, 5-8 Haziran, ABD.

Li T., Tang J., Xu J. 2016. Performance Modeling and Predictive Scheduling for Distributed Stream Data Processing, IEEE Transactions on Big Data, 2(4), 353-364.

Liu X., Buyya R. 2017. Performance-Oriented Deployment of Streaming Applications on Cloud, IEEE Transactions on Big Data, 5(1), 46-59.

Papageorgiou A., Poormohammady E., Cheng B. 2016. Edge-Computing-aware Deployment of Stream Processing Tasks based on Topology-external Information: Model, Algorithms, and a Storm-based Prototype, IEEE International Congress on Big Data, pp259-266, 27 Haziran-2 Temmuz, ABD.

Qian W., Shen Q., Qin J., Yang D., Yang Y., Wu Z. 2016. S-Storm: A Slot-aware Scheduling Strategy for

Even Scheduler in Storm, IEEE 18th International Conference on High Performance Computing and Communications, pp623-630, 12-14 Aralık, Avustralya.

Renart E.G., Diaz-Montes J., Parashar M. 2017. Data-driven Stream Processing at the Edge, IEEE 1st International Conference on Fog and Edge Computing, pp31-40, 14-15 Mayıs, İspanya.

Runsewe O., Samaan N. 2017. Cloud Resource Scaling for Big Data Streaming Applications Using A Layered Multi-dimensional Hidden Markov Model, 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp848-857, 14-17 Mayıs, İspanya.

Wang C., Meng X., Guo Q., Weng Z., Yang C. 2017. Automating Characterization Deployment in Distributed Data Stream Management Systems, IEEE Transactions on Knowledge and Data Engineering, 29(12), 2669 – 2681.

Wang Y., Tari Z., HoseinyFarahabady M., Zomaya A.Y. 2017. QoS-aware resource allocation for stream processing engines using priority channels, IEEE 16th International Symposium on Network Computing and Applications, pp1-9, 30 Ekim-1 Kasım, ABD.

Xu L., Peng B., Gupta I. 2016. Stela: Enabling Stream Processing Systems to Scale-in and Scale-out On-demand, IEEE International Conference on Cloud Engineering, pp22-31, 4-8 Nisan, Almanya.

Zhang J., Li C., Zhu L., Liu Y. 2016. The Real-time Scheduling Strategy Based on Traffic and Load Balancing in Storm, IEEE 18th International Conference on High Performance Computing and Communications, pp372-279, 12-14 Aralık, Avustralya.

İLERİ MÜHENDİSLİK ÇALIŞMALARI VE TEKNOLOJİLERİ DERGİSİ

Müşteri Kayıp Analizi İçin Sınıflandırma Algoritmalarının Karşılaştırılması

Alper ÇİÇEK*¹ , Yelda ARSLAN*² 

*İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, 34295, Türkiye

Araştırma Makalesi, Geliş Tarihi: 06.07.2020, Kabul Tarihi: 22.07.2020

Özet

Bu çalışmada kaggle.com üzerinden yayımlanan toplam iki veri seti ile çalışılmıştır. İlk veri seti bir banka müşterilerine, ikinci veri seti ise telefon operatörü müşterilerine aittir. Kullanılan iki veri seti için de çeşitli nitelikler incelenerek müşteri kayıp analizi yapılmıştır. Çalışmada Logistic Regression, Naive Bayes, Desicion Tree, K-NN, SVM ve LDA sınıflandırma modelleri kullanılmıştır. İncelemede toplam 13.000 müşteri bilgisi üzerinden belirlenen yöntemler ile çalışma yapılmıştır. Algoritmanın uygulanması kolay ve literatür çalışması fazla olduğu Python Programlama Dili seçilmiştir. Kullanılan sınıflandırma yöntemleri her iki veri seti için de uygulanmış ve doğruluk oranları karşılaştırılmıştır. En kararlı sonucu karar ağaçları algoritması olduğu gözlemlenmiştir. Tüm sonuçların %70'ten daha fazla bir doğruluk oranı vermesi başarılı bir çalışma yapıldığını ortaya çıkarmaktadır.

Anahtar Kelimeler: Müşteri Kayıp Analizi, Logistic Regression, Naive Bayes, Desicion Tree, K-NN.

Comparing Classification Algorithms of Customer Loss Analysis

Abstract

Total of two data sets which is published from kaggle.com has been used in this study. The first data set belongs bank customers, and the second data set contains telephone operator customers. Customer loss analysis was conducted by examining various attributes with using the two data sets. In this study, Logistic Regression, Naive Bayes, Desicion Tree, K-NN, SVM and LDA classification modeling are used. The methods which are determined from 13.000 customer information are used in this research. Python Programming Language, which is easy to implement and has many literature studies, has been chosen. The classification methods used were applied for both data sets and their accuracy rates were compared. It has been observed that the most decisive result is the decision trees algorithm. More than 70% of all results give an accuracy rate, a successful study is revealed.

Keywords: Customer Loss Analysis, Logistic Regression, Naive Bayes, Desicion Tree, K-NN.

¹Sorumlu yazar alpercicek@msn.com, ²arslanyelda@hotmail.com

1. GİRİŞ

Günümüzde firmalar için önemli bir konu haline gelen müşteriye elinde tutma, geri kazanma gibi konular popüler haldedir (Accardi ve Petersen, 2011). Artan rekabet ve pazarın odak noktasının müşteri olması, ilgiyi bu alana çekmiştir. Böylelikle her geçen gün müşteriye firmaya bağlı tutma veya geri kazanma gibi yeni geliştirilen ve üzerinde durulan noktalardır. Söz konusu firmanın devamlılığı olunca, tüketicinin de satın alma süreçleri ve karar verme süreçleri de dikkatle incelenmektedir (Kotler, 2003).

Müşteri kaybı kavramını basit bir şekilde “Rekabet nedeni ile müşterinin firmayı tercih etmeyi bırakması” olarak tanımlayabiliriz (Nettleton, 2014). Bu durumdan da anlaşılacağı üzere eldeki müşteriye tutmak zorlaşmaktadır. Müşteri kaybının doğru belirlenmesi firmadan vazgeçip, vazgeçmemesi önem arz etmektedir. Bu yüzden müşteri bırakabilir veya bırakmaz şeklinde sınıflandırılmalıdır (Tsiptsis ve Chorionopoulos, 2009). Bu çalışmanın amacı müşteri profillerine bakarak firmayı hangi tür müşterilerin bıraktıklarını tahmin etmek ve buna karşı alınabilecek önlemleri ortaya çıkarmaktır.

Müşteri kaybı, genellikle hizmet işletmelerini ilgilendirmekte olup, başta telekomünikasyon ve bankacılık olmak üzere birçok müşteriye hizmet eden firmalar için müşterilerin tercihlerinin değişmesi veya firmayı tercihten vazgeçişinin artmasıyla gündem haline gelmiştir (Kim, Jun ve Lee, 2014). Firmalar yeni müşteri kazanabilmek için çeşitli yollara başvurmaktadır. Dolayısıyla maliyetleri de artırmaktadır. Bu yüzden yeni müşteri kazanmak için kampanyalar düzenleyip yeni müşteriler kazanmak yerine, elindeki müşteriye cezbedici ve bırakmaması için yeni yollara başvurmaktadır. En büyük avantaj ise müşteriye kaybetmeden önce belirlemek ve onun üzerine önlemler almaktır. İyi belirlenmiş bir modelleme ile hangi müşterinin bırakabileceği hangi müşterinin sadık olduğunu ortaya çıkarılabilir.

Gelişen teknoloji ile birlikte işletmelerin mevcut müşterilerinin durumunu analiz edebilmek amacıyla birçok çalışma yapılmaktadır. Örneğin; Tosun (2006), bir bankanın müşteri verileri üzerinde çalışmasını yaparak kredi kartı kullanmayı bırak müşterileri tespit etmeye çalışmaktadır. Çalışmasında karar ağacı yöntemlerini kullanmaktadır.

Akbulut (2006), bir kozmetik firmasının müşteri verilerini analiz etmektedir. Akbulut çalışmasında, WEKA kullanarak kümeleme ve sınıflandırma

yapmaktadır. Kullandığı veriler üzerinde en etkili sonucu J.48 karar ağacı ile elde ettiğini anlatmaktadır. Yaptığı analiz sonucunda müşterileri aldıkları ürünlere göre belirli kategorilere ayırmaktadır ve işletme için önemli olabilecek önerilerde bulunmaktadır.

Asilkan (2008), yapmış olduğu çalışmasında yapay sinir ağları kullanarak ikinci el otomobil fiyatlarının gelecek dönemlerdeki durumunu tahmin etmeye çalışmaktadır. Sonuçların karşılaştırılmasında zaman serisi analiz yöntemleri kullanılmıştır. Yapılan çalışmada, yapay sinir ağlarının kullanılmasının nedeni, eldeki az miktardaki veriden anlamlı sonuçların çıkarılması olarak tanımlanmaktadır.

Koçtürk (2010), bireysel emeklilik müşterilerinin kayıp analizini yapmaktadır. Çalışmada SAS Enterprise Miner kullanarak müşteri bağlılık davranışı modellenmiştir. Entropy tabanlı 21 yapraklı karar ağacı ile analiz yapılmıştır. En kaliteli değer sağlandığı 4 dallı ve 46 yapraklı hiyerarşinin sağladığı görülmüştür.

Arifoğlu (2011), bir GSM firmasının müşterilerine ait verileri analiz etmektedir. Çalışmada aynı veri seti ile Navie Bayes, Support Vector Machine, Probabilistic Neural Network ve C-means Algoritmaları karşılaştırılmıştır. Kullanılan veri seti için Adaptif Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) ile birlikte kullanılan kurallarda C-means Algoritması'nın en iyi sonucu verdiği görülmüştür.

Ercan (2015), çalışmasında tek bir veri seti kullanarak Bayesian Network, Logistic Regression, SMO ve Simple CART Algoritmaları'ndan elde ettiği sonuçları karşılaştırmakta ve yeni bir tahmin modeli oluşturmaktadır. Oluşturduğu yeni modelin doğruluğunu arttırmak için Ensemble Yöntemleri uygulanmaktadır. Modelden elde edilen doğruluk oranının %68,2 olduğu açıklanmaktadır.

Çimenli (2015), lojistik sektöründe faaliyet gösteren bir işletmeye ait müşteri verileri ile kayıp analizi yapmaktadır. Çalışmada yapay sinir ağları ve karar ağaçları karşılaştırılmıştır. Çalışma sonucunda karar ağaçlarından %81 doğru sonuç alınıyorken, yapay sinir ağlarından %97 doğru sonuç alındığı vurgulanmaktadır.

2. SINIFLANDIRMA YÖNTEMLERİNE GENEL BAKIŞ

Bu bölümde yukarıda bahsi geçen ve uygulamada kullanılan sınıflandırma yöntemleri hakkında bilgi verilmektedir.

Logistic Regression

Lojistik Regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkinin şeklini belirleyerek tahmin etmek ve sınıflandırmak amacıyla kullanılan bir algoritmadır (Teng, Lin ve Ho, 2007). Doğrusal Regresyon Algoritması'nın uygulanması için bazı koşulların sağlanması gerekmektedir. Bu koşullardan bir tanesi bağımlı değişkenin sürekli olmasıdır. Ancak, uygulamalarda bağımlı değişken genellikle kategorik ve iki düzeylidir. Bağımlı değişkenin iki düzeyli olduğu durumlarda Lojistik Regresyon Analizi kullanılmaktadır (Gujarati, 1999).

Naive Bayes

Naive Bayes (Denklem 1), Bayes Teoremi baz alınarak oluşturulan kolay uygulanabilirlik ve anlaşılabilirlik yönünden avantajlı olan basit makine öğrenme algoritmalarındandır. Bu algoritma niteliğin hangi sınıflandırma bölümüne ait olduğunu tahmin etmektedir.

Sınıflandırma

$$(s_1, s_2, \dots, s_n) = \text{argmax}_k p(K=k) \prod_{i=1}^n p(S_i = s_i / K=k) \quad (1)$$

1 numaralı denklem Naive Bayes Formülü'nden anlaşılacağı üzere S_1 ile S_n arasındaki sınıflardan seçim yaparken bu koşulları yerine getiren k çarpımından farkı yoktur. Her koşulun bir sınıf değeri vardır ve hangi koşulun hangi sınıftan olduğunu bulunması için olasılık değeri çarpılır.

Decision Tree (CART)

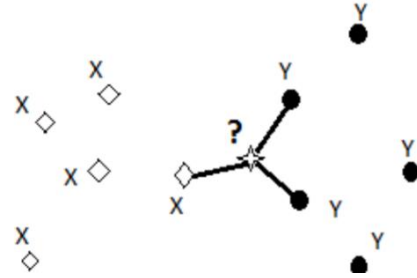
Sınıflandırma modellerinden olan karar ağaçları kolay kurulumu, etkili sonuçları ve oldukça fazla olan literatür çalışmasından dolayı en çok kullanılan yöntemdir. Karar ağaçları basit kararlar verebilme ve çok büyük verileri çok küçük sınıflara bölmek için kullanılır. Sıklıkla kullanılan karar ağacı modellerinden biri de CART Algoritması'dır. CART, kategorik sürekli değişkenli kullanarak sınıflama problemlerinin çözümünde karar ağaçlarını kullanan istatistiksel bir metottur. Bağımlı değişkenler kategorikse sınıflandırma ağaçları, eğer sürekliyse lojistik regresyon ağaçları olarak tanımlanır (Deconinck ve Hancock, 2005). CART, kümeyi ikiye ayırarak daha homojen ikili karar ağacı oluşturmak isteyen yapıya sahiptir. Bu yüzden CART veriyi iki düğüme ayırıp bağımsız değişkenleri kullanarak alt gruplara ayırmayı hedefler.

K-Nearest Neighborhood (K-NN)

Sınıflandırma işleminde temel amaçlardan biri en basit haliyle nesnelerin özelliklerine bakılarak nesnelerin hangi sınıfa ait olduğu belirlenmesidir. Çeşitli sınıflandırma algoritmaları vardır. Bunların içinde en çok ve yaygın kullanılan makine öğrenim algoritması (Denklem 2 ve Şekil 1) KNN Algoritmasıdır. Bir özellik seçilerek ona en yakın özellikle arasındaki yakınlığa göre sınıflandırma yapar ve nesnelere arasındaki mesafelerin hesaplanmasında kullanılır.

$$d_{(i,j)} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (2)$$

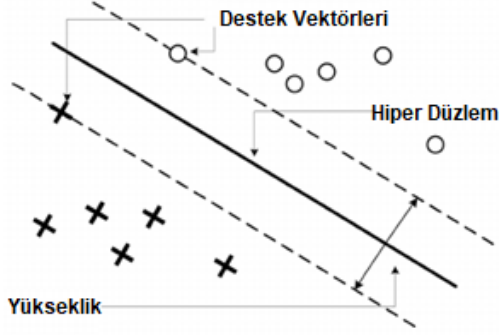
Çalışma prensibine bakıldığı zaman yeni bir nesne sınıflandırılmak için k değerine bakılır. Burada herhangi bir eşitlik olmaması için k değeri tek sayı olarak belirlenir. Yeni veri ile var olan veri hesaplanırken Manhattan uzaklığı veya Öklid gibi yöntemler kullanılır (Kılınç, Borandağ, Yücelar, Tunali, Şimşek ve Özçift, 2016).



Şekil 1. K-Nearest Neighborhood (K-En yakın komşu Algoritması)

Support Vector Machine (SVM)

Sınıflandırma sorunlarını çözmek için son yıllarda geliştirilen makine öğrenimi algoritmalarından biridir. Aynı zamanda örüntü tanımadaki popüler bir genelleme yöntemidir (Ayhan ve Erdoğan, 2014). Bu problemlerin çözümünde başarıyla uygulanmış ve yüksek başarımla elde edilmiş makine öğrenimi algoritmalarından biridir (Sengur, 2019). Veri madenciliği uygulamalarında sınıflandırma, doğrusal veya doğrusal olmayan bir işlev ile yerine getirilmektedir. Kernel İşlevi ile doğrusal olmayan dönüşümlerde, verileri doğrusal olarak ayırmaya olanak tanınmış olunur. Şekil 2'de Destek Vektör Makine Algoritması'nın genel yapısı verilmiştir (Karagülle, 2008).



Şekil 2. Destek Vektör Makine Algoritması'nın genel yapısı

1960'lı yıllara dayanan istatistiksel öğrenme algoritmalarından biri olan Destek Vektör Makinesi, Vapnik ve Chervoneskis tarafından tanımlanmış ve çok sayıda problemi başarıyla çözmüştür. Orijinal ifadesi Support Vektör Machine (SVM) olan Destek Vektör Makine Algoritması'nın temel mantığı, verileri doğrusal olarak ayırabilen en iyi çözümün belirlenmesidir (Sezer, Erçil ve Keskinöz, 2005). Doğrusal olarak ayırılmayan veriler dönüşüm tekniği ile farklı bir boyuta taşınarak çözülür. SVM'lerde kullanılacak örnek sayısı mühim değildir. SVM öğrenim sırasında görülmemiş verileri de sınıflayabilmektedir. Bunu genelleme özelliği sayesinde yapabilmektedir. Bu özelliği diğer tekniklerden ayıran en önemlisi olarak görülmektedir. Son yıllarda yüz tanıma, veri madenciliği, biyoloji, gen analizi ve örüntü tanıma sınıflandırılmalarında SVM kullanılmaya başlanmıştır. Bir olayı tanımlamak için özellikler vektör ile çağrılarak doğru seçilmelidir. Bu sayede doğru sonuçlar elde edilebilmektedir (Kecman).

Linear Discriminant Analysis (LDA)

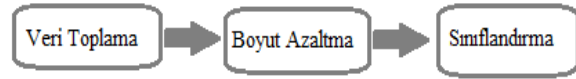
Linear Discriminant Analysis (LDA), verilerin sayısallaştırılmadığı modellerde kaplanan boyutu küçültmek, hesaplama verimliliğini artırarak algoritma başarı oranını yükseltmek için kullanılır (Martinez ve Kak, 2001). LDA, sınıflandırmayı en iyi duruma getiren özellik alt uzayını bulmaktır. LDA, bir veri kümesindeki boyut sayısını azaltmak ve sınıflandırmak için kullanılan doğrusal dönüşüm teknikleridir. LDA sonucu iki sınıf oluşur (Hua ve Yang, 2001).

LDA, bağımsız değişkenlerin aralarında nasıl bir etkileşim olduğunu ve bu etkileşimin bağımlı değişkeni nasıl etkilediğini ortaya çıkarmaya çalışan bir Varyans Analizi (ANOVA) ile yakından ilgilidir (Fisher, 1936). Kısaca LDA verilerin sınıflandırılmasından sonra bu

verilerin arasındaki farkı tespit etmeye çalışır (Perriere ve Thioulouse, 2003).

3. ANALİZ METODU

Kullanılan sınıflandırma algoritmaları: Logistic Regression, Naive Bayes, Decision Tree, K-NN, SVM ve LDA'dır. Verilerin analizinde Python ortamının seçilmesinin sebebi algoritma üzerinde değişiklik yapılabilmesi ve yapılan literatür çalışmaları etkili olmuştur. Yapılan çalışmada sınıflandırma modelleri ile müşteri kayıp tahmini yapılmak istenmiştir. Çalışma sonunda elde edilen değerler müşterinin kayıp edilip edilmediğini göstermektedir. İzlenen yöntemler Şekil 3'teki gibi özetlenebilir.



Şekil 3. Yöntem

3.1. Veri Toplama

Analizlerde kullanılacak veriler, şirketlerin belirli bilgilerini ya da çalışma verileri yayımlayan kaggle.com'dan temin edilmiştir. Veri setinin %80'ni eğitim, %20'si test için kullanılmıştır. Tablo 1'de özellikleri verilen veri seti bir bankaya aittir ve müşteri bilgilerinin içermektedir. Tablo 2'de bulunan veri seti ise bir operatör firmasına ait verilerden oluşmaktadır. Tabloda veri setinde bulunan bütün özellikler yer almaktadır. Tablo 1'de görünen ilk seti için boyut azaltma işlemi yapılarak önemli olduğu bilinen özelliklerden faydalanılmıştır.

Tablo 1. Banka veri setindeki özellikler

Özellik	En Küçük Değer	En Büyük Değer
Satır Sayısı	1	10000
Müşteri Numarası	15528523	15745083
Soyadı	-	-
Kredi Skoru	490	800
Yaşadığı Ülke	-	-
Cinsiyet	-	-
Yaş	24	58
Firma ile Çalıştığı Süre (Yıl)	0	13
Hesap Tutarı	10134,88	99183,9
Kullandığı Ürün Sayısı (Kart Sayısı vb.)	1	9

Kredi Kartı Olması Durumu	0	1
Firma ile Aktif Çalışma Durumu	0	1
Tahmini Maaş	2924,56	92865,41

Tablo 2. Telefon operatörü veri setindeki özellikler

Özellik
Durum
Firma ile Çalışılan Süre
Alan Kodu
Telefon Numarası
Uluslararası Plan
Sesli Posta Planı
Vmail Mesajları
Toplam Gün
Toplam Yapılan Günlük Arama
Toplam Yapılan Günlük Ücret
Toplam Yapılan Ev Telefonu Dakikası
Toplam Yapılan Ev Telefonu Araması Sayısı
Toplam Yapılan Ev Telefonu Araması Ücreti
Toplam Yapılan Gece Görüşme Dakikası
Toplam Yapılan Gece Görüşme Ücreti
Toplam Yapılan İnternet Dakikası
Toplam Yapılan Arama
Toplam Yapılan Ücret
Toplam Müşteri Hizmetleri Çağrı Sayısı

3.2. Boyut Azaltma

Tablo 1’de yer alan veri seti için boyut azaltma çalışmasından sonra kullanılan özelliklerin değişimi Tablo 3’te bulunmaktadır.

Tablo 3. Banka veri setinde kullanılan özellikler

Özellik	En Küçük Değer	En Büyük Değer
Kredi Skoru	490	800
Yaşadığı Ülke	-	-
Cinsiyet	-	-
Yaş	24	58
Firma ile Çalıştığı Süre (Yıl)	0	13
Hesap Tutarı	10134,88	99183,9
Kullandığı Ürün Sayısı (Kart Sayısı vb.)	1	9
Kredi Kartı Olması Durumu	0	1
Firma ile Çalıştığı Süre (Yıl)	0	13
Tahmini Maaş	2924,56	92865,41

3.3. Sınıflandırma

Sınıflandırma işlemlerinden sonra her müşteri özelinde bir sonuç beklenmektedir. Kullanılan sınıflandırma Tablo 4’te yer almaktadır.

Tablo 4. Sınıflar

Sonuç	Sınıf (Class)
0	Kaybedilen Müşteri
1	Kazanılan Müşteri

4. SONUÇ

Belirlenen tüm sınıflandırma algoritmaları ile iki veri seti üzerinde de çalıştırılmıştır. Bankaya ait olan ve üzerinde boyut azaltma işlemi yapılan Tablo 3’teki veri setinden, Tablo 5’teki sonuçlar elde edilmiştir. Elde edilen sonuçlara göre ilk veri seti için en iyi sonucu %81,00’lik doğruluk oranı ile LDA ve %76,40’lık bir değer ile en az doğruluk oranı olan algoritma K-NN’dir.

Tablo 5. Banka veri setinden elde edilen sonuçlar

Algoritma	Accuracy Score
LDA	%81,00
Decision Tree	%80,30
Logistic Regression	%79,75
SVM	%79,75
Naive Bayes	%78,45
K-NN	%76,40

Tablo 2’de yer alan bir operatör firmasına ait veri seti ile elde edilen sonuçlar Tablo 6’da gösterilmektedir. Tablo 6’da görüldüğü gibi bu veri seti için en iyi sonucu %80,4’lik oran ile Decision Tree algoritması ile elde edebilmekteyiz. En düşük doğruluk oranı ise %72,16 ile LDA ve Logistic Regression algoritmalarından elde edilmiştir.

Tablo 6. Operatör firmanın veri setinden elde edilen sonuçlar

Algoritma	Accuracy Score
Decision Tree	%80,41
Naive Bayes	%76,80
K-NN	%74,74
SVM	%74,23
LDA	%72,16
Logistic Regression	%72,16

Tablo 5'te ve Tablo 6'da görünen performans sonuçlarına göre uygulamalarda kullanılacak yöntemler ve veri setleri arasında doğru orantı olduğunu söylemek mümkündür.

5. TARTIŞMA

Sınıflandırma yaparken kullanılan algoritmaların seçimi, kullanılacak veri tipiyle uyumlu olmalıdır. Böylece daha doğru sonuçlar elde edebilmemiz mümkündür. İki veri seti de sınıflandırma ile uyumlu olmasına rağmen bazı algoritmalar iki veri seti için de beklenenden daha farklı sonuçlar vermiştir. Bunun sebebi iki veri setindeki özelliklerin sayıca fazlalığından olabilir. Çalışmada, 1. veri setinde (Tablo 1 Banka veri seti) 10 özellik var iken 2. veri setinde (Tablo 2 Telefon operatörü veri seti) 19 özellik vardır. İki veri seti için de özellik sayılarının birbirinden farklı olması, sınıflandırma algoritmalarında yaklaşık olarak %10'luk farklı bir oran vermiştir. Algoritmalarından elde edilen sonuçlar aşağıdaki gibi değerlendirilmiştir.

- LDA Algoritması'nda daha fazla özelliğe sahip olan 2. veri setinin başarı oranı daha yüksek iken daha az özelliğe sahip olan 1. veri setinde en az doğruluk oranı sağlayan 2. algoritma olmuştur.
- K-NN Algoritması ise daha az özelliğe sahip olan 1. veri seti için en az doğruluk oranına sahip algoritma olmuştur. 2. veri seti için ise, Tablo 6'da görüleceği üzere ortalama bir doğruluk oranı elde etmiştir.
- Decision Tree Algoritması, iki veri seti için de yaklaşık olarak %80 doğruluk oranı vermiştir.
- Naive Bayes Algoritması, 1. veri setinin en az doğruluk oranına sahip algoritmalarından biri olurken, 2. veri seti için diğer algoritmalara göre daha yüksek oranda doğruluk elde edilmiştir.
- SVM, iki veri setinin de karşılaştırma tablolarında ortalama sıralamada yer almaktadır. Bunun sonucu olarak; seçilen veri setlerinin SVM için daha uyumlu olabileceği düşünülmektedir.
- Logistic Regresyon, 1. veri seti için daha iyi sonuç verirken, 2. veri seti için en az doğruluk oranı veren algoritma olmuştur.

Verinin analizinin güvenilirliği, özelliklerinin iyi belirlenmesi ve sınıflandırmanın tam yapılması algoritmanın başarısını olumlu yönde etkileyeceği düşünülmektedir. Her iki veri seti için de elde edilen sonuçların %70'ten fazla olması kullanılan algoritmaların veri setleri için uygun olduğunu göstermektedir. Bu nedenle Tablo 5'te ve Tablo 6'da gösterilen sonuçlarda algoritmaların doğruluk oranları farklı olsa da, yapılan iki çalışma da genel olarak başarılı kabul edilebilir.

KAYNAKLAR

Akbulut S. (2006). *Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri segmentasyonu*, (Yüksek lisans tezi). Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Arifoğlu E. (2011). *Churn management by using Fuzzy C-Means*, (Yüksek lisans tezi). Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Asilkan Ö. (2008). *Veri madenciliği kullanılarak ikinci el otomobil pazarında fiyat tahmini*, (Doktora tezi). Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya.

Ayhan, S. ve Erdoğan, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *İibf Dergisi*, 9(1), 175-198.

Çimenli S. (2015). *Churn analysis and Prediction with decision tree and artificial neural network*, (Yüksek lisans tezi). Kadir Has Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Deconinck, E., Hancock, T., Coomans, D., Massart, D.L. and Heyden, Y.V. (2005). Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. *Journal of Pharmaceutical and Biomedical Analysis*, 39, 91-103.

Ercan P. (2015). *Detection of churners in internet games using crm approach: a case study on pishti plus*, (Yüksek lisans tezi). Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179-188.

Gujarati, D.N. (1999). *Temel ekonometri*, (Şenesen, Ü. ve Şenesen, G.G., Çev.), İstanbul: Literatür Yayıncılık (Orijinal çalışma basım tarihi 1978).

Hua, Y. and Yang, J. (2001). A direct LDA algorithm for high-dimensional data-with application to face recognition, *Pattern Recognition*.

Karagülle, F. (2008). *Destek vektör makinelerin kullanarak yüz bulma*. (Yüksek lisans tezi).

Kecman, V. (2001). *Learning and Support Vector Machines Neural Networks And Fuzzy Logic Models*.

Kılınç, D., Borandağ, E., Yücelar, E., Tunalı, V., Şimşek, M. ve Özçiftçi, A. (2016). KNN algoritması ve R dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, (3), 89-94.

Kim, N., Jung, K. H., Kim, Y. S., and Lee, J. (2012). Uniformly subsampled ensemble (USE) for churn management: *Theory and implementation*. *Expert Systems with Applications*, 39(15), 11839–11845.

Koçtürk Y. (2010). *Veri madenciliğinde bağlılık*, (Yüksek lisans tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Kotler, P. (2003). *Marketing Insights from A to Z*. New Jersey.

M. Accardi-Petersen. (2011). Agile marketing. New York: Apress. Commercial data mining: processing, *Analysis and Modeling for Predictive Analytics Projects*.

Martinez, A. M. and Kak, A. C. (2001). PCA versus LDA(PDF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228-233.

Perriere, G. and Thioulouse, J. (2003). Use of correspondence discriminant analysis to predict the subcellular location of bacterial proteins. *Computer Methods and Programs in Biomedicine*, 70, 99–105.

Sengur, A. (2019). Ulticlass least-squares support vector machines for analog modulation classification, *expert systems with applications*, 36(3), 6681-6685.

Sezer, O., Erçil, A. ve Keskinöz, M. (2005). Destek vektör makinesi kullanarak bağımsız bileşen tabanlı 3B nesne tanıma, *Sabancı University Computer Vision and Pattern Analysis Laboratory*.



Teng, J., Lin, K. and Ho, B. (2007). Application of classical tree and logistic regression for the management and health intervention plans in a community-based study, *Journal of Evaluation in Clinical Practice*, 13, 741-748.

Tosun T. (2006). *Veri madenciliği teknikleriyle kredi kartlarında müşteri kaybetme analizi* (Yüksek lisans tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Tsipsis, K., and Chorianopoulos, A. (2009). Data mining techniques in CRM: *Inside Customer Segmentation*. John Wiley & Sons, Ltd.

İLERİ MÜHENDİSLİK ÇALIŞMALARINI VE TEKNOLOJİLERİ DERGİSİ

Birliktelik Kuralı Temelinde Kısaltma Genişletme

Saadet Aytaç ARPACI*¹ , Banu DİRİ*² 

*Yıldız Teknik Üniversitesi, Elektrik Elektronik Fakültesi, Bilgisayar Mühendisliği Bölümü, 34220, İstanbul, Türkiye

Araştırma Makalesi, Geliş Tarihi: 31.05.2020, Kabul Tarihi: 09.08.2020

Özet

Metinlerde yaygın olarak kullanılan kısaltmaların açık karşılıklarının bulunması, bilginin elde edilmesi ve anlaşılması açısından önemli bir gerekliliktir. Metinde kullanılan kısaltmaları, herkesin bildiği düşünülüyorsa metinde bu kısaltmaların açık karşılıklarına yer verilmeyebilir. Bununla birlikte bazen kullanılan kısaltma birden fazla açık karşılığa sahip olabilir ve bu durum anlaşılabilirliği zorlaştırır. Kısaltmalardan doğru açılımın oluşturulabilmesi halen üzerinde çalışılan bir konu olarak farklı yöntemlerle incelenmektedir. İncelenen literatürde Apriori Algoritması'nın kısaltma açılımının bulunmasına yönelik kullanımına rastlanmaması nedeniyle, bu çalışmada *PubMed* özetlerinde bulunan kısaltmaların açık karşılıklarının elde edilmesi için Birliktelik Kuralı temelinde bir yöntem önerilmiştir. İncelenen veri kümesi ve kısaltmalar dahilinde kısaltmanın birden fazla açık karşılığı olsa dahi uygulanan yöntem, ortak bir minimum destek değeri ile %87,5 farklı minimum destek değerleri ile %87,5'ten daha yüksek doğrulukla kısaltmanın açılımını bulabilmektedir.

Anahtar Kelimeler: Birliktelik kuralı, Apriori, Kısaltma genişletme.

Abbreviation Expansion on the Basis of the Association Rule

Abstract

Finding clear expansion of abbreviations commonly used in the texts is an important requirement for obtaining and understanding the information. If the abbreviations used in the text are thought to be known to everyone, these abbreviations might not be used with clear expansion in the text. However, sometimes used the abbreviation can have more than one clear equivalents and this makes understanding difficult. To be able to create correct expansion from abbreviations is still examined by different methods on as a subject being studied. In the literature reviewed, since the use of the *Apriori* algorithm for abbreviation expansion is not encountered, in this study, a method is proposed based on the association rule to obtain the clear expansion of the abbreviations in *PubMed* abstracts. Within the examined data set and abbreviations, even if the abbreviation has more than one clear expansion, the applied method can find the clear expansion of the abbreviation with 87.5% accuracy by a common minimum support value or more than 87.5% accuracy by different minimum support values.

Keywords: Association rule, Apriori, Abbreviation expansion.

¹Sorumlu yazar saadeta99@gmail.com, ²diri@yildiz.edu.tr

1. GİRİŞ

İngilizcede “*acronym*” kelimesinin anlamına karşılık gelen “Birkaç kelimenin baş harflerinin veya ilk hecelerinin bir araya gelmesi ile oluşan kelime” bu çalışmada kısaltma olarak anılacaktır. Kısaltmalar, bilimsel makaleler, klinik notlar ve kullanıcı sorguları gibi özellikle biyomedikal literatürde yaygın olarak kullanılmaktadır (Jin, Liu, ve Lu, 2019). Kısaltmaların açık karşılıklarıyla birlikte metin içindeki yerleşimleri Liu, Lussier ve Friedman (2001)’e göre farklılık göstermekle birlikte, çalışmamızda kısaltmanın açılımı devamında parantez içerisinde kısaltmanın olduğu özetler kullanılmıştır. Örneğin, *Prosthetic Joint Infection (PJI)*. Metin içinde bir kez açık karşılığıyla birlikte verilen kısaltmalar daha sonra metin içerisinde sadece kısaltma hali ile kullanılmaktadır. *DNA* vb. gibi çoğunlukla açık karşılıklarının bilindiği düşünülen kısaltmaların bazen metin içerisinde açılımları verilmez. Bu durum yazının anlaşılabilirliğini azaltabilir. Ayrıca, kullanılan kısaltmanın *PDA (Patent Ductus Arteriosus, Posterior Descending Artery)* (Moon, Pakhomov, ve Melton, 2012) gibi birden fazla açık karşılığının olması durumunda ise yazının anlaşılabilirliği zorlaşabilir.

Kısaltmaların belirlenmesine yönelik yapılan çalışmalardan; (Yu, Hripcsak ve Friedman, 2002) çalışmalarında bir kısaltmanın açık biçim karşılığının makalede bulunması için bir dizi kural ile çalışan bir uygulama geliştirilmiş ve kısaltmaların dört genel kısaltma veri tabanından herhangi birinde karşılığının olup olmadığı araştırılmıştır (*GenBank LocusLink, SWISSPROT, LRABR* ve *BioABACUS*). Çalışmada, tanımlanmış kısaltmalar %95 kesinlik ve %70 hassaslık ile tespit edilmiştir. Bir kelimenin anlamının, bağlamıyla uyumluluğunu belirlemek için yapılan çalışmalardan (Moon, Pakhomov, ve Melton, 2012), belirli bir pencere boyutu ve yönelimi dikkate alarak elde ettiği çevre kelimelerin anlamsal ilgisini bulmak için kullandığı denetimli öğrenme yöntemleriyle yaklaşık %90 doğruluk elde etmiştir. (Stevenson, Guo, Amri ve Gaizauskas, 2009) çalışması, kısaltmaların doğru olan genişletilmiş halini belirlemek için problemi kelime anlam ayrımı (*Word Sense Disambiguation*) sorunu olarak görmüşler ve önerdiği sistemde belirsiz kelimenin çevresindeki kelimelerden çıkardığı özellikler ile *Vektör Uzay Modeli (VUM)*, *Naive Bayes (NB)* ve *Destek Vektör Makinesi (DVM)* gibi denetimli bir öğrenme yaklaşımıyla problemi incelemiştir. Çalışma, %99’a varan bir doğrulukla kısaltma açılımını belirlemiştir. (Wu, Xu, Zhang, ve Xu, 2015) çalışması ise, tıbbi metinlerde yer alan kısaltmaların açılımlarında

özelliği olarak sinir kelime gömmelerini (*neural word embedding*) kullanarak iki yeni kelime yerleştirme özelliği önermiş ve %95,79’luk başarı elde etmiştir.

(Li, Ji, ve Yan, 2015) çalışması, kısaltmaları belirlemek için iki kelimedenden oluşan gömme modeli önermiştir. Sözcük gömme işlemi, sözcükleri sürekli ve çok boyutlu bir vektör uzayında temsil etmek olup, çalışma sözcükler arasındaki vektörel uzaklığı hesaplayarak kelime benzerliklerini farklı iki veri seti üzerinde %93 ile %95’lik bir tutarlılıkla elde etmiştir.

(Zheng, Xiao, Wang, Zhu, ve Yang, 2019) çalışması, Çince kısaltmaları tanımak için *Evrimsel Sinir Ağı-Çift Yönlü Uzun Kısa Süreli Bellek-Şartlı Rasgele Alanlar’a (CNN-BLSTM-CRF)* dayanan bir sinir ağı modeli önermiştir. Alınan sonuçlar, makine öğrenmesi yöntemlerine göre daha iyi sonuç vermiştir. Derin öğrenme modellerinin farklı kombinasyonları ile yaklaşık %75 ile %79 arasında başarı elde edilmiştir. (Jin, Liu, ve Lu, 2019) çalışması, belirsiz kısaltmaların anlamlarının kesin tanımlarını bulabilmek için sadece *Çift Yönlü Uzun Kısa Süreli Bellek* modelini kullanmış ve beş farklı veri seti üzerinde yapılan testlerde en yüksek başarı *macro-f1* ile $98,3 \pm 3,5$ olarak alınmıştır.

Birliktelik Kuralı algoritmaları (Mahgoub, Rösner, Ismail, ve Torkey, 2008; Reátegui, ve Ratté, 2019) gibi çalışmalarda, metinden çıkarılan kelimelere kelimeler arası ilişkileri bulmak için uygulanmıştır. Bu çalışmalar gibi farklı konularda da Birliktelik Kuralının geniş bir uygulama alanı bulunmaktadır. İncelenen literatürde *Apriori* algoritmasının kısaltmaların açık karşılıklarının elde edilmesine yönelik kullanımına rastlanılmadığı için çalışmamızın katkısı, Birliktelik Kuralının bu yönüyle de incelenmiş olması ve kısaltmaların açık karşılıklarının bulunması işleminde öğrenme algoritmaları haricinde de yeni bir yaklaşımın kullanılmasındadır.

2. VERİ KÜMESİ

Çalışmada kullanılmak üzere yedi farklı kısaltma ve sekiz farklı açılım (Tablo 1) belirlenmiş ve her kısaltma için içerisinde geçtiği en az altı adet doküman toplanarak bir veri seti hazırlanmıştır. Dokümanların hazırlanmasında Tablo 2’deki web adreslerinden (web adreslerinin her biri <https://www.ncbi.nlm.nih.gov/pubmed> ile başlamaktadır) 46 adet özet tıp metinleri alınmıştır. Metinlerin içerisinde en az bir adet veya daha fazla kısaltma, açılımlarıyla birlikte yer almaktadır.

Tablo 1. Çalışmada incelenen kısaltmalar

Kısaltma	Açık Karşılığı	Kısaltma	Açık Karşılığı
DASH	Disabilities of the Arm, Shoulder and Hand	CA	Carbohydrate Antigen
TEA	Total Elbow Arthroplasty	PDA	Posterior Descending Artery
PJI	Prosthetic Joint Infection	POLST	Physician Orders for Life Sustaining Treatment
ORIF	Open Reduction Internal Fixation	PDA	Patent Ductus Arteriosus

Tablo 2. Veri kümesinin elde edildiği web adresleri

#	Web Adresi	Kısaltma	#	Web Adresi	Kısaltma	#	Web Adresi	Kısaltma
1	/31942319	DASH	16	/32191561	PJI	31	/21853470	PDA
2	/31686586	DASH, ORIF	17	/32110908	ORIF	32	/30725892	PDA
3	/31272266	DASH	18	/32158578	ORIF	33	/10642773	PDA
4	/31382027	DASH	19	/32282417	ORIF	34	/12910941	PDA
5	/30257773	DASH, TEA	20	/31949231	ORIF	35	/29298109	POLST
6	/29813168	DASH	21	/32106732	ORIF	36	/27696173	POLST
7	/31565456	TEA, PJI	22	/32057623	ORIF	37	/25441841	POLST
8	/31149971	TEA	23	/31472701	CA	38	/23865958	POLST
9	/30784119	TEA	24	/31593132	CA	39	/16948957	POLST
10	/24145266	TEA	25	/17933139	CA	40	/24743101	POLST
11	/23818030	TEA	26	/31826182	CA	41	/32211435	PDA
12	/32049563	PJI	27	/32090248	CA	42	/32175342	PDA
13	/31965585	PJI	28	/32226491	CA	43	/32143511	PDA
14	/31877146	PJI	29	/26197905	PDA	44	/32011298	PDA
15	/31857290	PJI	30	/23365974	PDA	45	/31983349	PDA
						46	/31724543	PDA

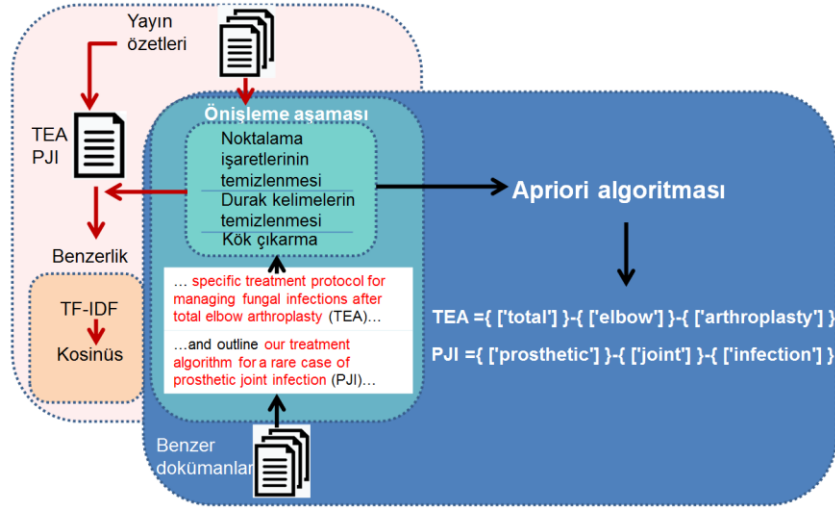
Çalışmada kullanılan kısaltmalar 2 ile 6 arasında karakterden ve büyük harflerden oluşmaktadır. Ayrıca, veri seti içerisinde *PDA* kısaltmasının iki farklı açılımına yer verilmiştir. *PDA*'nın ilk açılımı olan *Posterior Descending Artery* [29-34] dokümanlarında, ikinci açılımı olan *Patent Ductus Arteriosus* [41-46] dokümanlarında bulunmaktadır. İleriki çalışmalarda veri kümesi aynı kurallar dikkate alınarak genişletilebilir.

3. ANALİZ METODU

Önerilen yöntemin ilk aşamasında, veri kümesinin içerisindeki herhangi bir dokümanda yer alan kısaltmalar belirlenerek hangi kısaltmaların açık karşılıklarının bulunacağı bilgisi alınmıştır. (Medical Abbreviations, 2020)'deki İngilizce tıbbi kısaltmalar sözlüğünde, küçük harf içeren *kg=kilogram*, *syr.=syrup*,

Se=selenium gibi kısaltmalara göre, *MRI=Magnetic Resonance Imaging*, *DI=Diabetes Insipidus* gibi çoğunlukla hastalık veya teşhis cihazlarını tanımlayan ifadeler sözlüğün yaklaşık %70'ini oluşturduğundan, bu tür yaygın ifadelerle odaklanmak için kısaltmaların seçiminde 2 ile 6 karakter uzunluğunda ve büyük harflerden oluşmuş olmaları kriter olarak belirlenmiştir.

İkinci aşamada, kısaltmaları elde edilen kaynak dokümanın veri kümesindeki diğer dokümanlarla benzerlikleri ölçülmüştür. Bu işlem öncesi dokümanlar ön işlem adımı ile noktalama işaretlerinden, metin içinde sıkça tekrarlanan ve tek başına kullanıldıklarında anlam taşımayan durak kelimelerden (a, an, the, vb.) temizlenir, kök çıkarma (*lemmatizasyon*) işlemi ile kelimelerin köklerinin elde edilmesi sağlanır.

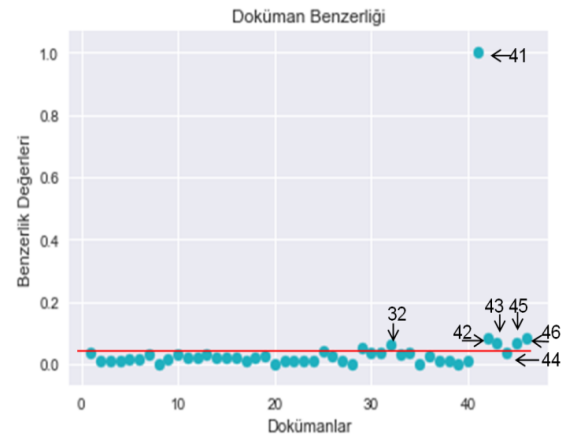


Şekil 1. Sistemin akış şeması

Çalışmamızda kelimenin ön veya son ekini atarak kök haline çeviren gövdeleme (*stemming*) işlemine (örneğin, studies, was, studying→studi, wa, studi) kıyasla, kelimeleri sözlükteki kökenine göre elde eden kök çıkarma işlemi (örneğin, studies, was, studying→study, be, study) tercih edilmiştir. Ön işlem aşamasında *NLTK* kütüphanesinden (Bird, Tan, Garrette, Ljunglöf, Nothman, Korobov, ve Dimitriadis, 2020) yararlanılmıştır. Benzerliği ölçülecek tüm dokümanlar için ön işlemin yapılmasının amacı, dokümanları kendilerine özgü kısaltma ve açık karşılıkları açısından ayırabilmektir. Önerilen sistemin genel akışı Şekil 1’de verilmiştir.

Çalışılan veri kümesi içinde “*DASH, TEA, PJI, ORIF, CA, PDA, POLST*” olmak üzere 7 kısaltma yer almaktadır. Her kısaltmanın belirli bir konuya özgü olması, aynı kısaltmaların bulunduğu dokümanları benzerliklerine göre ayırmakta tamamen olmasa da yardımcı olmaktadır. Özellikle *PDA* gibi birden fazla açık karşılığı olan kısaltmalarda doküman benzerliklerinin bulunması, işlemi kolaylaştırmaktadır. Dokümanların benzerliği, *tf-idf* değeri ve *kosinüs* açısının hesaplanmasıyla elde edilmiştir. Terim frekansı (*tf*), bir terimin bir dokümandaki geçiş sıklığı olup, bir dokümanda çok sayıda olan bir kelimenin o doküman için belirgin bir değere sahip olduğu söylenebilir. Ancak, diğer dokümanlarda da bu kelimenin geçmesi kelimenin değerini ölçmek açısından daha önemlidir. Ters doküman frekansı (*idf*) hesaplanarak kelimenin tüm dokümanlar bazında değeri elde edilir. Bu iki ölçütün birleşimiyle elde edilen *tf-idf* değeri, kelimenin dokümanların tümünde geçmesi durumunda en düşük değerini alır. Dokümanların *tf-idf* vektörel değerleri

kullanılarak dokümanlar arasındaki benzerlik, *kosinüs* benzerliği ile hesaplanır. Dokümanlar aynı ise, aralarındaki açının sıfır olması nedeniyle benzerlik değeri 1, benzer ise bire yakın olur. Dokümanların benzerliğinin bulunması için gerekli *tf*, *idf*, *kosinüs* değerlerinin hesaplanmasında *scikit-learn* kütüphanesinden (Cournapeau vd., 2020) faydalanılmıştır. *PDA* kısaltmasının bulunduğu 41. dokümanın diğer dokümanlarla olan benzerlik sonucu Şekil 2’de sunulmaktadır. Şekil 2’de grafiğin altında 41. dokümana en benzer olan dokümanlardan en az benzer olanlara doğru dokümanların sıralaması da sunulmuştur.



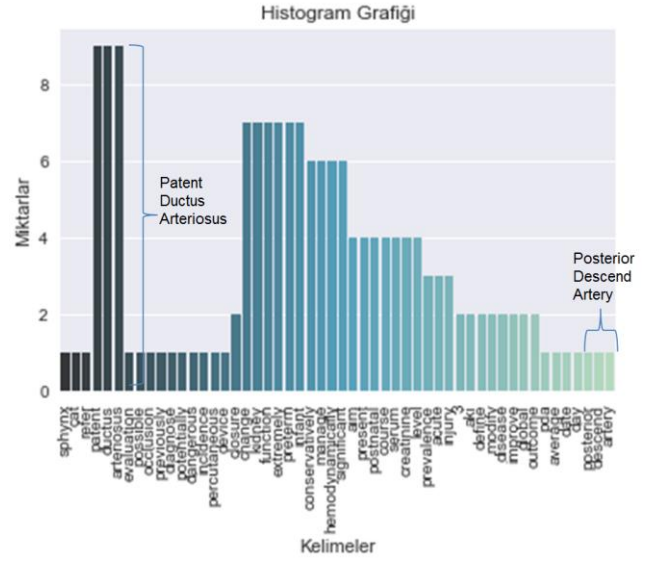
41 nolu dokümana en benzer olanlardan az benzer olanlara doğru sıralama: 41, 42, 46, 45, 43, 32|29, 25, 31, 34, 44, 30, 1, 13, 33, 10, 7, 19, 26, 36, 16, 14, 11, 15, 12, 18, 6, 5, 9|37, 40, 2, 3, 21, 27, 4, 17, 24, 22, 38, 23, 8, 20, 28, 35, 39 'PDA' kısaltması için alınan dokümanlar: 41, 42, 46, 45, 43, 32

Şekil 2. 41. Dokümanın diğer dokümanlarla olan benzerliğinin grafiksel gösterimi

Şekil 2’de de görüldüğü gibi *PDA=Patent Ductus Arteriosus*’ un bulunduğu 41. doküman 44. dokümanı daha benzer olarak bulması gerekirken 32. dokümanı benzer bulmuştur. Bu durum gibi diğer kısaltmalar için de yapılan benzerlik işlemi, farklı kısaltmaya sahip olan dokümanların benzer olarak algılandığı bir sonuç oluşturabiliyordu. Aranılan kısaltmaya sahip dokümanlar sıralandığında çoğunlukla ilk 25 içinde yer almıştır. Bu nedenle aranılan kısaltmanın bulunduğu dokümanların seçimi genel olarak bu sınır içerisinde yapılmıştır. Bu sınır değerinin verilmesi veri kümesinin daha büyük olması durumunda bir avantaj sağlamaktadır. *PDA* kısaltmasının birden fazla açık karşılığı olduğu için ilk 25 dokümandan seçim yapmak, diğer açık karşılığa sahip dokümanların da alınmasına neden olacağından bu kısaltmaya sahip dokümanların ilk 6’sı için inceleme yapılmıştır. *Apriori* algoritmasının temel mantığına göre kelime sıklığının yüksek olması, bu kelimelerin oluşturacağı alt kelime kümelerinin de görünme ihtimalini yükseltir, şeklindedir. Bu anlamda algoritmaya verilecek verilerin elde edilmesi için aranılan kısaltmaya özgü benzer dokümanlar ile devam eden işlemler yapılmıştır. Veri kümesindeki dokümanlarda kısaltma ve açık karşılıkları Şekil 1’de görüldüğü gibi parantez içinde bir kısaltma ifadesi ve parantezin sol tarafında da kısaltmaya ait açık karşılıkları verilmiş şekildedir. Doküman içindeki kısaltmaların bu açık karşılıklarına ulaşabilmek adına “(” işareti öncesi 11 eleman alınıp bu kelime grupları ön işlem basamağı ile noktalama işaretlerinden, durak kelimelerden temizlenmiş, kök çıkarma işlemi ile kelimelerin sözlükteki köklerinin elde edilmesi sağlanmıştır. Kelime grupları ön işlem basamağı ile daha belirgin kelime grupları haline getirilmiştir. *PDA* kısaltmasının bulunduğu 41, 42, 46, 45, 43 ve 32. dokümanlardan elde edilmiş kelimelere ait histogram grafiği Şekil 3’te sunulmaktadır.

Her benzer doküman için elde edilen bu kelime grupları temelinde Birliktelik Kuralı uygulanmıştır. Şekil 4’te genel olarak ifade edilmiş sözde koda dayanan *Apriori* algoritması çalıştırılmıştır. İncelenen kısaltmalar için algoritmanın üç adım ilerlemesi yeterli olmuştur.

Birliktelik Kuralına göre kelime-1 varsa kelime-2 ile birlikte ortaya çıkacağı, *kelime-1 → kelime-2* ifadesi ile belirtilir. Eşik destek değerinin üstündeki öge alt kümelerinin elde edildiği algortmada güçlü birliktelik kuralları oluşturmak için güven değerinin de yüksek olması beklenir. Destek ve güven değerleri için denklem 1 ve denklem 2 kullanılmıştır.



Şekil 3. 41, 42, 46, 45, 43 ve 32. dokümanlardan elde edilmiş kelimelere ait histogram grafiği

k=adım

$L_k = k.$ adındaki sık geçenler kümesi

Basla

$L_1 = \{\text{sık geçenler-1 kümesi}\}$

$k \leftarrow 2$

while $L_{k-1} \neq \emptyset$ *do*

$C_k \leftarrow$ aday kelime kümesi (L_{k-1})

$C_k.\text{frek} \leftarrow$ aday kelime kümesi frekansı (C_k)

$L_k \leftarrow C_k$ ’yi eşik desteğe göre ayırma ($C_k.\text{frek}$, *mindestek*)

$k \leftarrow k+1$

return L

Son

Şekil 4. Apriori algoritması sözde kodu

$$\text{Destek} (\text{kelime} - 1 \rightarrow \text{kelime} - 2) = \frac{\alpha}{\epsilon} \quad (1)$$

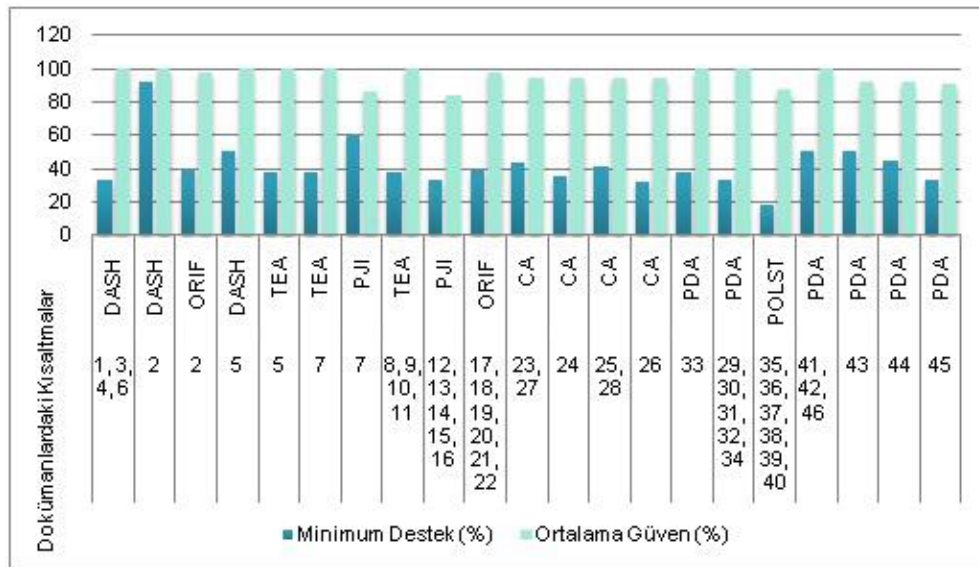
$$\text{Güven} (\text{kelime} - 1 \rightarrow \text{kelime} - 2) = \frac{\alpha}{\phi} \quad (2)$$

α , kelime-1 ve kelime-2’nin bir arada bulunduğu küme sayısı; ϵ , kaynak kısaltmayı içeren benzer dokümanlardan elde edilen toplam küme sayısı; ϕ , kelime-1’in bulunduğu küme sayısını ifade etmektedir.

4. DENEYSEL SONUÇLAR

Tablo 3. Kısaltmaların çalışmadan elde edilen açık karşılıkları

Doküman No	Kısaltma ve Açık Karşılığı
1, 2, 3, 4, 5, 6	DASH={['disability']}-{'arm'}-{'shoulder'}-{'hand'}}
5, 7, 8, 9, 10, 11	TEA={['total']}-{'elbow'}-{'arthroplasty'}}
7, 12, 13, 14, 15, 16	PJI={['prosthetic']}-{'joint'}-{'infection'}}
2, 17, 18, 19, 20, 21, 22	ORIF={['open']}-{'reduction'}-{'internal'}-{'fixation'}}
23, 24, 25, 26, 27, 28	CA={['carbohydrate']}-{'antigen'}}
29, 30, 31, 32, 33, 34	PDA={ ['posterior']}-{'descend'}-{'artery'}}
35, 36, 37, 38, 39, 40	POLST={['physician']}-{'order'}-{'life'}-{'sustain'}-{'treatment'}}
41, 42, 43, 44, 45, 46	PDA={['patent']}-{'ductus'}-{'arteriosus'}}



Şekil 5. Her kısaltma için gereken minimum destek ve ortalama güven değerleri

Metin içindeki kısaltmanın açık karşılığının kural tabanlı bulunmasına yönelik yapılan çalışmamızda elde edilen kısaltma açılım sonuçları Tablo 3'te sunulmaktadır. Sistem tüm kısaltmaların açık karşılığını farklı minimum destek değerleriyle elde edebilmiştir. Tüm kısaltmalar için ortak kullanılan %50 minimum destek değeri ile açık karşılıkları aranan 8 kısaltmanın (*POLST* kısaltması hariç) 7'si için sonuca ulaşılmıştır. Bu anlamda sistemin genel doğruluğu %87,5'tir. *POLST* kısaltması için %18'lik bir minimum destek değeri kullanılmıştır.

Apriori algoritması ile kısaltmaların elde edilmesi için gereken minimum destek değerleri ve oluşan kurallar dahilinde elde edilen güven değerlerinin ortalaması da Şekil 5'teki grafikte sunulmaktadır.

Çalışmamızda 2 harfli (*CA*), 3 harfli (*TEA*, *PJI*, *PDA*), 4 harfli (*DASH*, *ORIF*), 5 harfli (*POLST*) kısaltmaların,

aynı dokümandaki birden fazla kısaltmanın (örneğin, *DASH-ORIF*, *DASH-TEA*, *TEA-PJI*), aynı harf dizilimine fakat farklı açılımlara sahip (örneğin, *PDA*) olan kısaltmanın açık karşılıkları bulunmuştur. Dokümanlardan elde edilen kelime gruplarının *Apriori* algoritmasına verilmeden önce ön işlem aşamasında çıkarılan noktalama ve durak kelimeleri *DASH* kısaltmasının açık karşılığında (*Disabilities of the Arm, Shoulder and Hand*) yer alan kelimelerin işleme alınmamasına neden olmuştur.

Aynı doküman (örneğin, 2, 5, 7) içerisindeki iki kısaltmanın açılımlarının elde edilmesinde minimum destek α değeri her iki kısaltmanın açık karşılıklarının bulunabilmesine göre ortak verildiğinden grafikte 2 ve 5 nolu dokümanlardaki *DASH* kısaltması ile 7 nolu dokümandaki *PJI* kısaltması daha yüksek minimum destek değeri ile işleme alınmıştır.

Aynı harf diziliminde farklı açılımlara sahip kısaltmaların (örneğin, *PDA*) bulunduğu dokümanlar benzerlik işlemiyle ayrıldığında kısaltmanın aynı açılımına sahip diğer dokümanların yanında farklı açılımına sahip dokümanlarda işleme eklenebilmiştir bu durum minimum destek değerini artırma gerekliliğine neden olmuştur. Aynı kısaltmayı içeren benzer dokümanlar üzerinden işlem yapılması nedeniyle Şekil 5'te görüldüğü gibi *ORIF* kısaltmasında aynı sabit minimum destek değerine göre ortak sonuç alınmıştır, farklı dokümanların işleme eklenmesi ile *CA* kısaltmasında farklı minimum destek değerinin verilmesi gerekmiştir.

Kısaltmaların açık karşılıklarının bulunması işleminde, elde edilen ortalama güven değerleri *PJI* ve *POLST* için %80 üzerinde, *CA* ve *ORIF* için %94 üzerinde, *PDA*, *DASH* ve *TEA* için %100 olarak bulunmuştur.

Çalışmada tüm veri seti içinden dokümana benzer ilk 25 dokümana yönelik bir sınır değer üzerinden işlemlerin devam etmesi, veri seti içine eklenecek farklı kısaltmaların (özellikle birden fazla açık karşılığa sahip) bulunmasında bir kısıt olarak görülebilir, bu durumda tüm veri seti için en uygun sınır değeri/değerleri belirlenerek işlemler yapılabilir.

5. SONUÇ

Bu çalışmada *PubMed* özetlerinde bulunan kısaltmaların açık karşılıklarının elde edilmesine yönelik Birliktelik Kuralı temelinde bir yaklaşım sunulmuştur. İki ile altı karakter arasında ve büyük harften oluşan kısaltmaların açık karşılıkları, bir dokümanda bir veya birden fazla kısaltma olması durumunda bulunabilmiştir. Birden fazla açık karşılığa sahip olması nedeniyle açılımının bulunması sorun olan kısaltmaların doğru açılımları *Apriori* algoritması ile yapılabilmektedir. Önerilen yöntem tüm kısaltmaların açık karşılıklarını farklı minimum destek değerleriyle %100'lük bir başarı ile elde edebilmiştir. Ortak belirlenen bir minimum destek değerine göre işlem yapıldığında elde edilen başarı %87,5 olmuştur. Kısaltma açılımlarının bulunması için kullanılan öğrenme algoritmalarının eğitim için uzun zaman gerektirmesi nedeniyle önerilen bu yöntem ile daha kısa sürede sonuca ulaşıldığından alternatif olarak önerilebilir.

KAYNAKLAR

Bird, S., Tan, L., Garrette, D., Ljunglöf, P., Nothman, J., Korobov, M., ve Dimitriadis, A. (2020). Natural

Language Toolkit. 31 Mayıs 2020 tarihinde <https://www.nltk.org/> adresinden erişildi.

Courneau, D., Brucher, M., Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V.,... Boisberranger, J. (2020). scikit-learn. 31 Mayıs 2020 tarihinde <https://scikit-learn.org/stable/> adresinden erişildi.

Jin, Q., Liu, J., ve Lu, X. (2019). Deep Contextualized Biomedical Abbreviation Expansion. 31 Mayıs 2020 tarihinde <https://arxiv.org/pdf/1906.03360.pdf> adresinden erişildi.

Li, C., Ji, L., ve Yan, J., (2015). *Acronym Disambiguation Using Word Embedding*. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (ss. 4178–4179).

Liu, H., Lussier, Y. A., ve Friedman, C., (2001). *A study of abbreviations in the UMLS*. Proc AMIA Symp (ss. 393-397).

Mahgoub, H., Rösner, D., Ismail, N., ve Torkey, F. (2008). A Text Mining Technique Using Association Rules Extraction. 31 Mayıs 2020 tarihinde <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.212.8624&rep=rep1&type=pdf> adresinden erişildi.

Medical Abbreviations, (2020). 31 Mayıs 2020 tarihinde https://www.tabers.com/tabersonline/view/Tabers-Dictionary/767492/all/Medical_Abbreviations adresinden erişildi.

Moon, S., Pakhomov, S., ve Melton, G.B., (2012). *Automated Disambiguation of Acronyms and Abbreviations in Clinical Texts: Window and Training Size Considerations*. AMIA Annu Symp Proc. (ss. 1310–1319).

Reátegui, R., ve Ratté, S., (2019). *Analysis of Medical Documents with Text Mining and Association Rule Mining*. International Conference on Information Technology & Systems (ss.744–753).

Stevenson, M., Guo, Y., Amri, A., ve Gaizauskas, R. (2009). Disambiguation of Biomedical Abbreviations. 31 Mayıs 2020 tarihinde <https://dl.acm.org/doi/10.5555/1572364.1572374> adresinden erişildi.

Wu, Y., Xu, J., Zhang, Y., ve Xu, H., (2015). *Clinical Abbreviation Disambiguation Using Neural Word Embeddings*. Proceedings of the 2015 Workshop on

Biomedical Natural Language Processing (BioNLP 2015) (ss.171–176).

Yu, H., Hripcsak, G., ve Friedman, C. (2002). Mapping Abbreviations to Full Forms in Biomedical Articles. *Journal of the American Medical Informatics Association*, 9(3), 262–272.

Zheng, J., Xiao, X., Wang, B., Zhu, Y., ve Yang, L. (2019). A New Method for Abbreviation Prediction viaCNN-BLSTM-CRF. 31 Mayıs 2020 tarihinde <https://iopscience.iop.org/article/10.1088/1742-6596/1267/1/012001> adresinden erişildi.

İLERİ MÜHENDİSLİK ÇALIŞMALARI VE TEKNOLOJİLERİ DERGİSİ

Otomotiv Endüstrisinde Akıllı Fabrika Uygulamaları ve Türkiye’de Adaptasyon Süreci

Berfinsu BARUT*¹ , Muharrem ÜNVER*² , Celal Türkcan KAYIM*³ , Ecem TOPRAK*⁴ 
Ezgi UYSAL*⁵ 

*Karabük Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Karabük, 78050, Türkiye

Araştırma Makalesi, Geliş Tarihi: 24.07.2020, Kabul Tarihi: 06.08.2020

Özet

Rekabetin arttığı globalleşen dünyada birçok gelişme yaşanmaktadır. Bu gelişmeler içinde Sanayi Devrimleri de yer almaktadır. Zamanımızda Endüstri 4.0 hakimiyetini sürdürmektedir. Endüstri 4.0 ile insan gücünün etkinliği devam edip, geleneksel üretim sistemleri modernleşmiştir. Bu modernleşme ile otomasyon sistemleri geliştirilmiştir ve böylece birbiriyle iletişim kurabilen, öğrenebilen, karar verebilen makineleri de beraberinde getirmiştir, akıllı fabrikanın temellerinin de atılmasını sağlamıştır. Akıllı fabrikalarda çalışan sayısı en az sayıda ve üretimden daha çok üretim yapan makinelerle ilgilenilmektedir. Buradaki amaç ise, insan kaynaklı üretim hatalarını azaltmak, yavaşlamaları engellemek ve uzun süren karar verme sürecini ortadan kaldırmaktır. Yapılan bu çalışma, dinamik ve hızlı gelişen otomotiv sektöründe amortisör üretimi gerçekleştiren işletmenin mevcut durumu incelenmiştir. Yapılan incelemeler neticesinde kullanılabilir teknolojiler hakkında tespitler yapılmış olup, bu teknolojilerin adaptasyon sürecinde yaşanabilecek avantajlar ve dezavantajlar açıklanmıştır.

Anahtar Kelimeler: Endüstri 4.0, Akıllı fabrika, Otomotiv sektörü.

Smart Factory Applications in the Automotive Industry and Adaptation Process in Turkey

Abstract

There are many developments in a globalizing world where competition is increasing. Industrial Revolutions are among these developments. Industry 4.0 continues to dominate in our time. With Industry 4.0, the efficiency of manpower continues and traditional production systems have been modernized. With this modernization, automation systems have been developed, thus bringing machines that can communicate, learn, and make decisions together, and have laid the foundations of the smart factory. Machines dealing with the minimum number of employees in smart factories and producing more than production are dealt with. The aim is to reduce man-made production errors, prevent slowdowns and eliminate long-term decision making. In this study, the current situation of the enterprise that produces shock absorbers in the dynamic and rapidly developing automotive sector is examined. As a result of the examinations made, determinations have been made about the technologies that can be used, and the advantages and disadvantages that may occur during the adaptation process of these technologies have been explained.

Keywords: Industry 4.0, Smart factory, Automotive industry.

¹Sorumlu yazar berfinsubarut@gmail.com, ²muharremunver@karabuk.edu.tr, ³turkcankayim@gmail.com,
⁴toprakecemm@gmail.com, ⁵ezgiuysal48@gmail.com

1. GİRİŞ

Endüstri 4.0 ile ağırlıklı olarak insan gücünün etkin olduğu merkezi sistemli geleneksel üretim sistemleri beyin gücüyle çalışan daha esnek sistemlere dönüşmeye başlamıştır. Birbirleriyle haberleşebilen makinalardan oluşan üretim sistemlerinde insana gereksinim olmadan hızlı karar verebilme, esnek üretim, daha kaliteli üretim, olası sorunları ön görme veya daha erken müdahale edebilme gibi avantajlar sağlanmaktadır. Endüstri 4.0 sağladığı faydalarla birlikte hızla dünyada yayılmaya başlamıştır ve üreticileri akıllı fabrikalar kurmaya teşvik etmiştir. İlk olarak Almanya’da ortaya çıkan Endüstri 4.0 kavramının temellerini Siber-Fiziksel Sistemler ve Nesnelerin İnterneti oluşturmaktadır (Koca, 2018).

Yaklaşık 6 yıldır hayatımızda olan Endüstri 4.0’a adaptasyon süreci tüm büyük ülkelerde devam etmektedir. Akıllı fabrikalara geçiş özellikle AR-GE alanında yüksek yatırım gerektirmektedir. Hızlı seri üretim sağlanması, verimlilik ve zaman kazancı gibi getirileri olsa da Endüstri 4.0, teknoloji ve internete yüksek bağımlılıktan dolayı küçük bir dijital açık sebebiyle yüksek maliyetli sorunlar yaratabilecektir (Çırak, 2019) ve bazı ülke ekonomilerinde yüksek maliyetli bu çalışmaları karşılamayabilecektir.

Fabrikalarda Endüstri 4.0 alanında çalışmak üzere bölümler kurulmalıdır. Endüstri 4.0 alanında çalışmalar yapılabilmesi için ülkelerin, eğitim, profesyonel bilginin sürekliliği, yeni yapıya uygun görev tanımları ve örgüt organizasyonu, donanım mimarisinde yapılandırma sağlanması gibi gerekli birçok değişiklik söz konusudur (Kılıç ve Alkan, 2018). Türkiye’de de bu çalışmalar henüz yeterli seviyede olmasa da ülkedeki büyük firmalar tarafından çalışmalar yürütülmektedir. Türkiye’nin piyasa rekabetinde üstünlük sağlama, ülkemizde üretim yapan yabancı firmalarla çalışmalarını sürdürebilmesi, günümüz teknolojilerini yakalayarak gelişmiş ülkeler arasında yer alması için Endüstri 4.0’a adaptasyonu ve akıllı fabrikaların çoğalması oldukça önemlidir.

Koca (2018)’nın Sanayi 4.0’ın Türkiye açısından fırsatlar ve tehditlerini değerlendirdiği çalışmasında da belirttiği gibi, Türkiye Asya ve Avrupa arasında köprü görevi gören bir üretim sahası konumundadır. Türkiye’nin Endüstri 4.0 teknolojisini sağlayamaması, işçilik maliyetlerinin Endüstri 4.0 anlayışındaki üretimlere göre yüksek olmasından dolayı ülkeleri Türkiye’deki üretim sahalarından çekilmeye itecektir. Bu da pazarda Türkiye için büyük bir tehdit oluşturacaktır. Ancak Koca, süreçteki yıllık gelirin

%3-4 oranında AR-GE yatırımlarına harcanmasının Türkiye’nin Endüstri 4.0 adaptasyonunu kolaylaştıracağını savunmuştur.

Kamber ve Bolatan (2019)’ın Endüstri 4.0 farkındalığına yönelik 2019 yılında gerçekleştirdikleri, 199’u sektör belirtmiş olan toplam 202 katılımcılı araştırmada, farkındalığın otomotiv ve elektronik sektörlerinde yüksek olduğu ve Endüstri 4.0 alanında çalışmalarını başlatan firma sayısının fazla olduğu görülmüştür. Kurulan hipotezlere göre, mevcut bilgi sistemleri altyapısının Endüstri 4.0’a uyumlu olduğu, tedarikçiler ve paydaşların teknolojik seviye farklılıklarının söz konusu olduğu, firmaların kaynak ve kabiliyetlerinin Endüstri 4.0 dönüşümünde yeterlilikleriyle ilgili farklar olduğu, Endüstri 4.0 çalışmaları olan ve olmayan firmaların, dönüşüm sürecinde benzer teknolojik yaklaşımların firmalarına katkısı olacağı düşüncesine sahip olmaları gibi sonuçlara ulaşılmıştır.

Eryer ve Eryer (2019) çalışmalarında, Endüstri 4.0 dönüşümünü sağlayabilecek önemli teknolojik gelişmeleri şöyle sıralamışlardır: büyük veri, akıllı robotlar, simülasyon, entegrasyon, nesnelerin interneti, siber güvenlik, bulut teknolojisi ve artırılmış gerçeklik. Bu teknolojilerin, birçok sektörde önemli etkiler sağlayacağını belirttikleri çalışmalarında lojistik alanını ve bu alanda Endüstri 4.0’ın etkilerini incelemişlerdir. Endüstri 4.0’ın lojistik sektöründe gerçek zamanlı iletişim ile iş yapış şeklini etkileyeceğini, özellikle de maliyetleri azaltırken tedarik risklerini ortadan kaldırdığını belirtmişlerdir.

Otomotiv yan sanayi alanında gelişmiş olan Bursa, otomotiv sektöründe Endüstri 4.0 adaptasyonu üzerine yapılan SWOT Analizi’nde bilgi birikim eksikliği, projelere yönelik kaynak sıkıntısı, Endüstri 4.0 süreçlerine ayrılacak vakit kısıtı gibi zayıf yönler ortaya konmuştur. Bunun yanı sıra coğrafi konum, ulusal ve uluslararası projelere dahil olabilme imkanı, üniversite-sanayi işbirlikleri, bakanlık teşvikleri gibi fırsatlar mevcuttur. Ancak ülkede yaşanan ekonomik krizlerin etkisi olarak Bursa otomotiv endüstrisinde faaliyette olan işletmelerin %90’ının iş hacimlerini etkileyecek ekonomik sıkıntılar yaşamış olmalarından, teknolojik yatırımlar için uygun şartlar sağlanmasının güç olduğu belirtilmiştir (Sürmen, 2019).

Dünyada yaşanan teknolojik gelişmelerin hızı her geçen gün artmaktadır. Bu gelişmelere uyum sağlayabilmek ve gelişmiş ülkelerin arasında yer alabilmek için Türkiye’de de hızlı gelişmeler yaşanmalıdır. Ulusal Yenilik Sistemi oluşturmak için ülkede üniversite

sayıları arttırılmış, KOSGEB destekleri genişletilmiş, Teknoloji Geliştirme Bölgeleri kurulmuştur. Fakat tüm bunlar yeterli olmamıştır. Ulusal Yenilik Sisteminin etkin bir şekilde ilerlemesi için yeni teknolojilerin kullanılması yaygınlaştırılmalı, AR-GE yatırımı arttırılmalı ve nitelikli iş gücü geliştirilmesi yolunda çalışmalar yapılmalıdır (Bilgin ve Işık, 2018).

Endüstri 4.0 alanındaki yeniliklere ülkemizin uyum sağlaması için yapması gereken birçok çalışma vardır. Ülkede küçük ve orta ölçekli firmalar fazla olduğu için dijital dönüşümde öncelik bu firmalara verilmelidir. Aynı zamanda bu dönüşüm devlet tarafından politikalarla desteklenmeli ve takip edilmelidir. Bu sayede gelişen fabrikaların uluslararası pazarda da rekabet gücü artacaktır (Akben ve Ös, 2019).

Bahçekapılı, Berigel ve Calp (2018)'in yapmış oldukları "Endüstri 4.0 Kapsamında Akıllı Fabrikaların İncelenmesi" adlı çalışmada, Endüstri 4.0 ve temel bileşenleri açıklanmıştır. Endüstri 4.0 kapsamında "Akıllı Fabrikalar" detaylıca incelenmiş olup, geleneksel fabrikaların akıllı fabrikaya dönüşüm süreci ve gerekli teknolojiler hakkında tartışmalar gerçekleştirilmişlerdir. Tüm bu incelemeler neticesinde, akıllı fabrikaların tüm üretim süreçlerine adapte edilerek, desteklenmesi gerektiği kanısına varmışlardır.

Endüstri 4.0'ın günümüzde üretim alanında etkisinin olduğu ve gün geçtikçe bu etkinin artacağını düşünülmektedir. Yıldız (2018) çalışmasında ülkelere göre Endüstri 4.0'ı değerlendirmiş ve Almanya'nın yapmış olduğu yatırımlar karşısında 4. Sanayi Devrimi için öncü davrandığını, Finlandiya'nın ise adından her ne kadar bahsedilmese de ciddi hamleler yaptığını belirtmektedir. Türkiye'nin Endüstri 4.0 alanında öncü olabilmesi için, tüm üretim tesislerinde dijital teknoloji erişiminin kolaylaştırılıp, dijital sanayi platformlarının oluşturulup akıllı fabrika uygulamalarının yaygınlaşması için gerekli çalışmaların yapılması gerektiğini savunmaktadır.

Türkiye'nin otomotiv endüstrisinde Endüstri 4.0'a adaptasyon sürecinin incelendiği bu çalışmada öncelikle konunun anlaşılabilmesi için sanayi devrimleri kısaca açıklanmış ve akıllı fabrikalar hakkında bilgiler verilmiştir. Sonrasında ise dünyada ve Türkiye'de otomotiv sektörünün gelişiminden bahsedilmiştir. Yüksek otomasyon teknolojileriyle üretim yapmakta olan bir otomotiv işletmesindeki geleneksel yöntemlerden kaynaklanan aksaklıklar incelenmiştir. Tespit edilen bu sorunların akıllı fabrika teknolojileriyle nasıl iyileştirilebileceği hakkında öneriler sunulmuştur. İşletmenin bu adaptasyon sürecinde karşılaşılabileceği

zorluklar ve çalışma sonucu elde edeceği faydalar tartışılmıştır.

1.1. Sanayi Devrimleri ve Tarihsel Gelişim Süreci

Teknoloji, hayatımızdaki hemen hemen her alanda oldukça önemli bir yere sahiptir. Özellikle ürün/hizmet üretimine teknolojinin entegre olması, ekonomik anlamda ülkeler için önemli bir kaynaktır. Günümüzdeki üretim sistemlerinde kullanılan teknolojilerin temelleri 18. ve 19. yüzyıllarda sömürge devleti olan İngiltere'de başlayan, ardından Batı Avrupa ve Amerika'ya yayılan sanayi devrimlerinin gelişmiş bir noktasıdır. Sanayi devrimleri, gelişimlerin yaşandığı çağlar boyunca toplumsal ve ekonomik anlamda büyük değişimlere sebep olmuştur.

Sanayi Devrimi öncesi üretim, basit aletlerle ve aile üyelerinin katılımıyla evlerde ya da atölyelerde yapıyordu. Üretimde kullanılan enerji kaynağı insan ya da hayvan gücü yani kas gücü idi (Günay, 2002). Bu dönemlerde, tarımda yaşanan gelişmeler köyden kente göçü beraberinde getirmiş ve istihdam için büyük bir nüfus oluşmuştur. Fabrikalaşma ise bu dönemlerden sonra ortaya çıkmıştır. Sömürge devletlerinin yağmaları, sanayileşmedeki en büyük ekonomik kaynak olmuştur. Bu kaynaklarla makine üretimi için yatırım yapılmıştır. Makineler eve sığmayacak büyüklükte olduğundan evde çalışmanın mümkün olmadığı anlaşılacak üretim alanları oluşturulmuş ve işçiler burada çalışmaya başlamışlardır. Bu süreçte sömürgelerden elde edilen hem ekonomik kaynaklar hem de ham maddeler işlenerek sömürgelere tekrar satılmış ve kapital sağlanmıştır. Kapital sağlanması, yeni yatırım alanları arayışını da beraberinde getirmiştir. Bu da sanayi devriminin yaşanmasına neden olan faktörlerden biri olmuştur.

Fabrikalaşma gelişiminin olumlu etkilerinin yanı sıra, orta sınıfın zenginleştiği bu dönemde, fabrikalarda çalışan işçilerin haklarının sağlanmamasından dolayı sınıfsal sorunlar yaşanmıştır. Bu sorunlar özellikle, ucuz işçilik sağlamak için çalıştırılan kadın ve çocuk işçiler ile ilgili olmuştur. Şartların ağırlığı ve ücretlerin yetersiz olması ile kadınlar ve çocuklar bu koşullar altında zorluklar yaşamış ve ezilmişlerdir. İşçi sınıfında yaşanan bu sıkıntılı süreç sosyalizmin temellerini oluşturmuştur.

1.1.1. Endüstri 1.0

İlk sanayi devriminin yaşandığı bu dönem İngiltere'de maden işlerinde kullanılmak üzere geliştirilen buhar makinesinin (1712), İskoçyalı Mühendis James Watt tarafından makineyi verimli hale getirecek şekilde

geliştirmesi ile başlamıştır (Sürmen, 2019). Bu dönem makineleşme çağı olarak da anılmaktadır. Bu dönemde küçük aile üretim yapısı kendini fabrika ortamlarına bırakmıştır. Dönemde, buhar ve kömür ile çalışan makinelerin kullanıldığı tekstil ve demir-çelik sektörleri ön plandadır.

1.1.2. Endüstri 2.0

Sanayi gelişimindeki 2. dönem, dünyanın seri üretim kavramıyla tanıştığı bir dönemdir. Akış bantlarını üretimde kullanmaya başlayan Henry Ford, seri üretimin öncüsü olarak kabul edilmektedir. Bu dönemde buhar ve kömür ile çalışan makineler yerini, elektrik, petrol ve doğal gaz kullanımı ile çalışan daha gelişmiş makinelere bırakmaya başlamıştır. Birinci çağda da var olan demir-çelik sanayisi gelişerek ağır sanayinin de gelişmesini sağlamıştır. Endüstri 2.0, telgraf ve telefonların kullanımının başladığı teknoloji devrimi olarak da anılmaktadır. Dönemde gelişen demiryolları ulaşımı, ham madde tedarik edilmesini ve ürünlerin farklı pazarlara açılmasını sağlamıştır. Aynı zamanda ulaşımın ve iletişimin gelişmesi insanları birbirine daha da yakınlaştırmış ve toplumların hayat standartları farklılaşmaya başlamıştır (Pamuk ve Soysal, 2018).

1.1.3. Endüstri 3.0

Endüstri 3.0, 20. yüzyılda otomasyonun üretim sistemlerine entegre olmaya başladığı endüstri dönemidir. Bu dönemde üretim sistemleri matematik tabanlı ve programlamaya yönelik olmaya başlayarak bilgi ve iletişim teknolojileri üretimlerde önem kazanmıştır. Üretimde yaşanan bu gelişim sayesinde tüketicinin isteklerini karşılayan bir üretim sistemi sağlanmıştır. Bu gelişmeler ile üretimlerde insan gücüne duyulan ihtiyaç azalmaya başlamıştır (Gabaçlı ve Uzunöz, 2017). Ancak otomatikleşen sistemlerle daha pratik ve yüksek adetlerde üretim yapılarak üretim maliyetlerinin düşürüldüğü daha verimli ürün/hizmet sağlanmaya başlanmıştır.

1.1.4. Endüstri 4.0

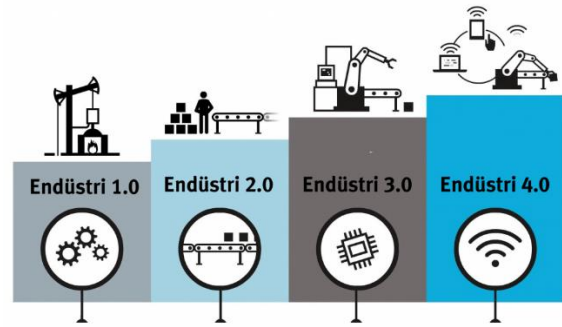
Endüstri 4.0 terimi, ilk olarak 2011 yılında Alman hükümeti tarafından ileri teknoloji stratejisi bağlamında kullanılmıştır (Rodic, 2017). Üretim sistemlerinde optimizasyonun sağlanacağı, dijitalleşmenin üretime entegre edildiği bir dönemdir. Bu dönemde nesnelere interneti (IoT) son derece önem kazanmıştır. Makinelerin iletişimde olacağı, insansız üretimlerin sağlandığı bir üretim sistemi anlayışı ile üretim sürelerinin düşürüleceği, birim başı maliyetlerin azaltılacağı sistemler hedeflenmiştir. Endüstri 4.0, akıllı

fabrikaların oluşturulduğu, yazılım ve donanımın ön planda olduğu bir endüstri dönemidir.

Endüstri 4.0, mühendislik, planlama, lojistik ve operasyonel süreçlerde kaliteyi arttırıp, kalite standartlarının daha iyi olmasının yanında, uzun solukluluk ve esneklik sağlamaktadır. Ayrıca verimliliği arttırarak, gerçek zamanlı ve dinamik olarak optimize edilmiş, kendi kendini organize eden değer zincirinin optimizasyonu olarak da tanımlanabilir (Soylu, 2018).

Endüstri 4.0'ın hedefleri, bilgi teknolojileriyle üretimi gerçekleştirilen ürünlerin özelleştirilmesini, değer zincirinin otomatik ve esnek uyumunu sağlamak, makineler ve ürünler arasındaki iletişimi kolaylaştırıp, insan-makine etkileşimini sağlamak, akıllı fabrikalarda ise nesnelere interneti ile üretim optimizasyonu, yeni iş modelleri ve hizmetler sağlamak olarak sıralanabilmektedir (Yıldız, 2018).

Günümüzde yaşanan bu endüstri devrimi (Şekil 1), üretim sistemlerini dijitalleştirilmesinden ve insansız üretimin hedeflenmesinden dolayı istihdamda düşüş yaşanacağına yönelik soru işaretlerini de barındırmaktadır ancak süreçte yeni iş alanlarının açılacağına yönelik öngörüler de bulunmaktadır.



Şekil 1. Endüstri Devrimler (Festo, 2020).

1.2. Akıllı Fabrikalar

Günümüzde Endüstri 4.0 ile yaşanan gelişmelerle birlikte insan gücüne dayalı üretim azalmaya başlamıştır. Otomasyon sistemleri geliştirilmiş, birbirleriyle iletişim kurabilen, öğrenebilen ve karar verebilen makineler geliştirilmiştir. Bu yeniliklerin uygulanmasıyla elde edilen ve süreçlerin kendi kendine işlenmesini sağlayan sistemlere akıllı fabrika denmektedir. Akıllı fabrikalarda insan gücü kullanılmamaktadır. Çalışan insan sayısı minimumdur ve üretimden çok üretimi gerçekleştiren makinelerle ilgilenen çalışanlar bulunmaktadır. Bu sebepten akıllı fabrikalar aynı zamanda karanlık fabrikalar olarak da

bilinmektedir. Akıllı fabrikaların temel amacı insandan kaynaklanan hataları, yavaşlamaları, uzun süren karar verme süreçlerini ortadan kaldırmak, daha hızlı ve daha az hatalı üretim sağlamaktır. Bunların yanında bir diğer etkisi ise üretim maliyetini düşürmektir.

Marr (2017) “Veri Stratejisi” adlı kitabında bir fabrikanın veya sistemin Endüstri 4.0 olarak düşünülebilmesi için sıraladığı müşterek çalışma, bilgi şeffaflığı, teknik destek, merkezi olmayan karar verme süreci maddelerinde, birbiriyle bağlantılı ve iletişim kurabilen makineler, cihazlar, sensörler ve insanlar; fiziksel dünyanın sanal bir kopyasını yaratacak sistemler; karar veren, sorun çözme süreçlerinde destekleyici, tehlikeli görevlerde insanlara yardımcı olma becerilerine sahip sistemler ve basit kararlarda karar verici otonom sistemler sağlanmasından bahsetmiştir.

Akıllı fabrikayı oluşturmak için gerekli olan bileşenler Şekil 2’deki gibidir ve en temel bileşenlerden biri Nesnelerin İnterneti (IoT) kavramıdır.



Şekil 2. Akıllı fabrika bileşenlerinin şematik gösterimi

Nesnelerin İnterneti endüstriyel otomasyon dışında tedarik zinciri, ev otomasyonu, ulaşım, giyilebilir teknoloji alanlarında da kullanılmaktadır. Otomotiv sektöründe ise akıllı fabrikalar dışında kendi kendini yöneten veya uzaktan yönetilebilen araçlarda kullanılmaktadır. Türkiye’de bu alanda Turkcell, Triomobil, Arvento Mobile System, Mobiliz gibi firmalar Nesnelerin İnterneti Teknolojisi’ni kullanan öncü firmalardandır (Şimşek, 2019). Akıllı fabrikanın uygulanmasında IoT, temeldeki ekipman kaynaklarını entegre etmek için kullanılır. (Calp vd., 2018).

Bu kavram ile makineler internet aracılığıyla birbirleriyle ve insanlarla iletişim kurar. Bu sayede makineler algılayabilir, bağlantı kurabilir ve sürekli verileri güncelleyebilir ve analiz edebilir. Burada önemli olan bir diğer nokta makinelerin sadece insanlara gerekli bilgileri aktarabilmesinin dışında kendi aralarında da iletişim kurabilmesidir. Böylelikle düşünebilen, hissedebilen ve karar verebilen makineler ortaya çıkmaktadır. Nesnelerin İnterneti sayesinde üretimde insan sayısı minimuma düşürülür, üretim kolaylıkla takip edilir, oluşabilecek arızalar önceden belirlenir, boş süreler azaltılır, üretim için gerekli kaynak ihtiyaçları daha hızlı belirlenir ve temin edilir.

Akıllı fabrikaları oluşturan bir diğer temel bileşen ise Siber Fiziksel Sistemler’dir. Siber Fiziksel Sistemler fiziksel dünya ile sanal dünyayı birbirlerine bağlar. Çeşitli teknolojiler kullanarak nesnelere tanımlar ve bu bilgileri bilgisayara aktarır. Böylelikle nesnelerin internetiyle sağlanan makine-makine ve makine-insan iletişimi gerçek zamanlı olarak ağa aktararak kontrol edilebilir, ulaşılabilir hale dönüştürülür. “Siber Fiziksel Sistemler” sayesinde akıllı makineler stok seviyesi, problemler, hatalar, üretimdeki değişiklikler gibi konularda bilgi verebilmektedirler.

Akıllı fabrikalardaki bir diğer temel kavram ise Bulut Bilişim Teknolojisi’dir. Makinelere elde edilen tüm veriler bulut tabanlı sistemlerde depolanmaktadır. Bulut bilişim teknolojisi sayesinde tüm veriler tek bir alanda toplanmakta ve her yerden ulaşılabilir ve paylaşılabilir hale gelmektedir. Bu toplanan verilerin tamamı büyük veriyi oluşturmaktadır. Büyük veriler analiz edilerek yeni kararlar alınabilir veya farklı stratejiler oluşturulabilir.

Bir diğer temel bileşen ise yapay zeka kavramıdır. Kullanılan makineler yapay zeka tabanlı çalıştıklarında büyük veriyi de kullanarak karar ve tavsiye verebilen makineler haline gelmektedir. Bu sayede makineler hem kendi hatalarını hem de iletişim halinde oldukları için diğer makinelerin hatalarını anlık kontrol edebilmekte ve üretimi sürekli iyileştirebilmektedirler. Bu durumda yapay zeka, ileri bilişsel becerilere sahip olma dışında akıl yürütme, problem çözme, çıkarım yapma gibi insan davranışları gösteren bir bilgisayar olarak da tanımlanabilir (Arslan, 2020).

Akıllı fabrikalarda kullanılan bir diğer kavram ise simülasyondur. Simülasyon sayesinde fabrikanın tamamı veya bir kısmı bilgisayar ortamında gerçeğe uygun bir şekilde oluşturulur. Bu sayede sistemdeki olası aksaklıklar önceden simülasyonun test edilmesiyle belirlenebilir. Sistemdeki değişikliklerin, aksaklıkların,

hassas analizlerin denenmesi, gözlenmesinde ve analitik çözümlerin ispatı için kullanılmaktadır (Öztürk, 2004). Simülasyon ile sağlanan bu süreç takibi, zaman, maliyet ve risk yönetimi bakımından avantajlar sağlamaktadır (Çelen, 2017).

Son olarak belirteceğimiz bir diğer temel kavram ise Artırılmış Gerçeklik Teknolojisi'dir. Artırılmış Gerçeklik sayesinde gerçek ortamdaki görüntülerle sanal ortamdaki görüntüler birleştirilebilmektedir. Bu sayede ürünler daha öncesinde sanal ortamda test edilebilmekte, bir makinedeki aksaklık makine sökülmeden bulunabilmekte, olası durumlar sanal ortamda denenip önceden çözülebilmektedir.

Tüm bunların dışında akıllı fabrikalar için siber güvenlik oldukça önemlidir. İşletmenin varlıklarını, bilgilerini, uygulamalarını, personelini korumak için tüm bu bilgiler siber ortamda saklanmalıdır. Siber güvenlik sayesinde işletmelerin tüm bu varlıkları korunmaktadır. Başlangıçta elektronik para birimi Bitcoin'in güvenliği için geliştirilen Block Zinciri Teknolojisi son yıllarda sistemlerin güvenliği içinde kullanılmaktadır. (Şekkeli ve Bakan, 2018).

1.3. Otomotiv Sektörünün Dünya'daki Gelişimi

Otomotiv sektörünün tanımı için kısaca motorlu taşıtları üreten sanayi denilmektedir. Bu sektör içinde demir-çelik, petro-kimya, lastik gibi birçok sektörü barındırmaktadır ve sektörlerin teknolojik gelişimleri için öncülük etmektedir, ekonomiye büyük katkısı bulunmaktadır. Otomotiv sektörünün temeli otomobil üretiminden geçmektedir.

İlk otomotiv sanayi 1900'lü yıllarda kurulmuştur. İlk başlarda ustaların yetenek ve el becerilerinden yararlanılarak, sipariş üzerine yapılan üretim sonrasında Henry Ford'un "Kitlesel Üretim" yaklaşımı ile büyük bir sanayi olmuştur. Yeni üretim sistemine geçiş ile maliyetler azaltılırken kalite ve verimlilik artmaya başlamıştır. ABD bu gelişmeler sonucunda büyük pazar paylarına sahip olmuştur (Polat, 2020). İkinci Dünya Savaşı'nın başlaması ile 1940'lı yıllarda gelişmeler durmuştur.

Savaş sonrası yıllarda ise Almanya'nın öncülüğünde İngiltere, Fransa ve İtalya gibi ülkeler, kitlesel üretimden müşteri odaklı üretime geçiş çalışmalarına başlamıştır. 1960'lı yıllarda ABD'li firmaların Avrupa'da kendi kurdukları ya da iş birliği yaptıkları firmalar ile artık Avrupa halkına hitap edecek üretime başlamışlardır. Yapılan çalışmalar sonucunda Avrupa ekonomisi artmıştır aynı zamanda Batı Avrupa'nın

otomotiv sanayisi büyüme göstermiştir. Bu büyümeyle Batı Avrupa'nın 6,2 milyon adet olan ürün sayısı, ABD'nin 7 milyon adet olan ürün sayısına yaklaşmıştır. Daha rakip olamayan Japonya'nın ise üretim adedi 165 bindir. 1970'li yıllarda Japonya'nın yalın üretime verdiği önem artmıştır ve bu esnada ABD, Batı Avrupa Ülkeleri yatırım alanlarını genişletmekteydi.

1980 yılında Avrupa Ekonomi Topluluğunun gösterdiği hızlı gelişmeler ile diğer üye ülkelerde de otomotiv sektörü yaygınlaşmaya başlamıştı. Japonya'da bu gelişmeler yaşanırken ABD ve Avrupa Birliği ülkelerine tesisler açıp, denizleri aşan bir pazar oluşturmuşlardır. 1980'lerin ikinci yarısında ise, Japonya'ya karşı ABD'nin ve Avrupa Birliğinin üreticileri Pazar kayıplarına uğramaya başlamışlardır. Ama bu kayıp kısa süreli olmuştur, bu sürede diğer ülkelere boş durmayarak yalın üretim tekniklerini inceleyip kendilerine empoze etmişlerdir.

1997'li yıllarda Güney Kore'de otomotiv sektöründe hızlıca büyümüş ve üretim adedi 2,4 milyona ulaşmıştır. Ama 1998'de Güney Doğu Asya'da yaşanan krizle, Güney Kore'nin büyüme hızı yavaşlamıştır.

1970-1999 yılları arasında otomotiv sektöründe yaşanan hızlı gelişmeler sonucunda Japonya'nın dünya üretimindeki payı 15'den 21'e yükselmiş olup, Batı Avrupa'nın dünya üretimindeki payı 48'den 38'e, ABD'nin ise 34'ten 21'e düşmüştür.

1999 yılının verilerine bakılarak Japonya ve Güney Kore için şöyle bir çıkarımda bulunulabilmektedir. Bu ülkelerin sadece sanayi için değil tüm alanlar için ithalata kapalı olduğu anlaşılmıştır. Japonya buna karşılık olarak dünyanın her yerine yayılmış, Güney Kore ise bu konular üstünde gelişmeler sağlamaya çalışmaktadır.

Soydal (2007)'in 2007 yılında yayımlanan çalışmasına bakıldığında, 2005 yılı verilerinin incelenmesi sonucu toplam 64,7 milyon motorlu taşıt aracı üretildiği bilgisine varılmıştır. 2007 yılındaki çalışmada yer alan bilgiye göre son 4 yılda dünya otomobil üretiminin %11 artış göstererek gelişmekte olan bölgelerdeki otomobil üretim artışı %54 olmuştur. Bu durumda dünya üretimindeki pay %15 den %20,6'ya yükselmiştir. Türkiye'nin dünya otomobil üretimi içindeki payı ise %0,4 den %1'e yükselmiştir.

Haziran 2020'de OSD tarafından yayınlanan Otomotiv 2019 yılı Küresel Değerlendirme Raporunda 2019 yılı önceki yıllarla karşılaştırıldığında, dünya genelinde otomotiv sanayi üretimi %5, küresel pazar ise %4'lük

daralma göstermiştir. Türkiye'deki 2019 yılına ait durum incelendiğinde ise toplam üretimin önceki yıllara kıyasla %6 oranında düşüş gösterdiği görülmüştür. Türkiye otomotiv endüstrisi pazarında ise 2018 yılının aynı dönemine göre %23'lük bir gerileme söz konusudur (Otomotiv Sanayi Derneği, 2020).

1.4. Türk Otomotiv Sektörünün Gelişimi

Ülkemizde 1954 yılına kadar otomotiv sektöründen bahsedilememiştir. Cumhuriyetin ilan edildiği zaman diliminde ülkemizde demiryolu tercih edilmekteydi ve motorlu taşıt ihtiyaçlarımızı ithalatla gerçekleştirmekteydik. Demiryolu ulaşımından, karayoluna geçmemizdeki sebep, İkinci Dünya Savaşı'ndan sonrasında Batılı Ülkelerle ilişkilerimizin kuvvetlenmesi ve iletişim halinde olunmasıydı. İhtiyaçların ithalatla karşılandığı ilk yıllarda döviz rezervlerinde erime olmuştur bu sebeple daha az döviz ile parça ithalatı gerçekleştirilerek montaj yapma yoluna gidilmiş ve maliyet düşürülmüştür. 1954 yılında Willys Overland Co.'ya verilen montaj izni, ithal ikamesini seçen Türkiye'de otomotiv sektörünün de başlangıcı olmuştur (Soydal, 2007) ve BMC, Otosan, MAN gibi birçok büyük marka da bu yolu tercih etmiştir. Böylece karayolu ulaşım yolları ve motorlu araç parkı genişlemiştir. 1990 yılına kadar Türkiye'de imal edilen motorlu araçlar otomotiv yan sanayi sektörümüzün gelişiminde önemli rol oynamaktadır. 1990'lı yıllarda, otomotiv sektörünü, döviz kazandıran sektör haline getirecek önlemler alınmış, yeni model araç üretimine dönük yatırımlara teşvik sağlanmıştır. Türkiye'deki maksimum üretim kalitesi ve minimum maliyetler gibi avantajlar dünyada otomotiv sektörü öncülerinin Türkiye'ye yatırım yapmasında büyük rol oynamıştır. Avrupa'ya yakın oluşu ve büyük potansiyele sahip pazarı ile Türkiye 1990'lı yılların sonunda otomotiv üretiminde önemli bir konuma gelmiştir (Soydal, 2007).

2. ANALİZ METODU

Çalışmada amortisör üretimi yapmakta olan bir otomotiv işletmesinin mevcut üretim şekli incelenmiştir. İncelenen bu sistemde üretimde zaman ve kalite kaybı yaratan iyileştirilmesi gereken operasyonlar tespit edilmiştir. Tespit edilen bu operasyonların hangi akıllı teknolojilerle iyileştirilebileceği üzerine araştırmalar yapılmıştır. İşletmenin akıllı fabrika olma yolunda uygulayabileceği teknolojilerle ilgili önerilerde bulunulmuş ve bu adaptasyon sürecindeki avantajlar ve karşılaşılabilecek zorluklar üzerine tartışılmıştır.

Uygulamanın yapıldığı işletme Türkiye'de üretim yapmaktadır. Bu işletme önde gelen yurtdışı ve yerel

çok sayıda OEM üreticilerine yıllık yaklaşık 6 milyon üretim kapasitesiyle amortisör üretmektedir.

Amortisörlerin amacı araçlardaki sarsıntıyı ve titreşim etkisini minimuma indirmek, konfor ve sürüş güvenliği sağlamaktır. İşletmede birçok farklı referanslarda amortisör üretimi yapılmaktadır fakat bunları genel olarak iki farklı tip olarak ayırabiliriz. Bunlar teleskopik tip amortisör ve strut tip amortisörler olmak üzere Şekil 3 ve Şekil 4'te gösterildiği gibidir.



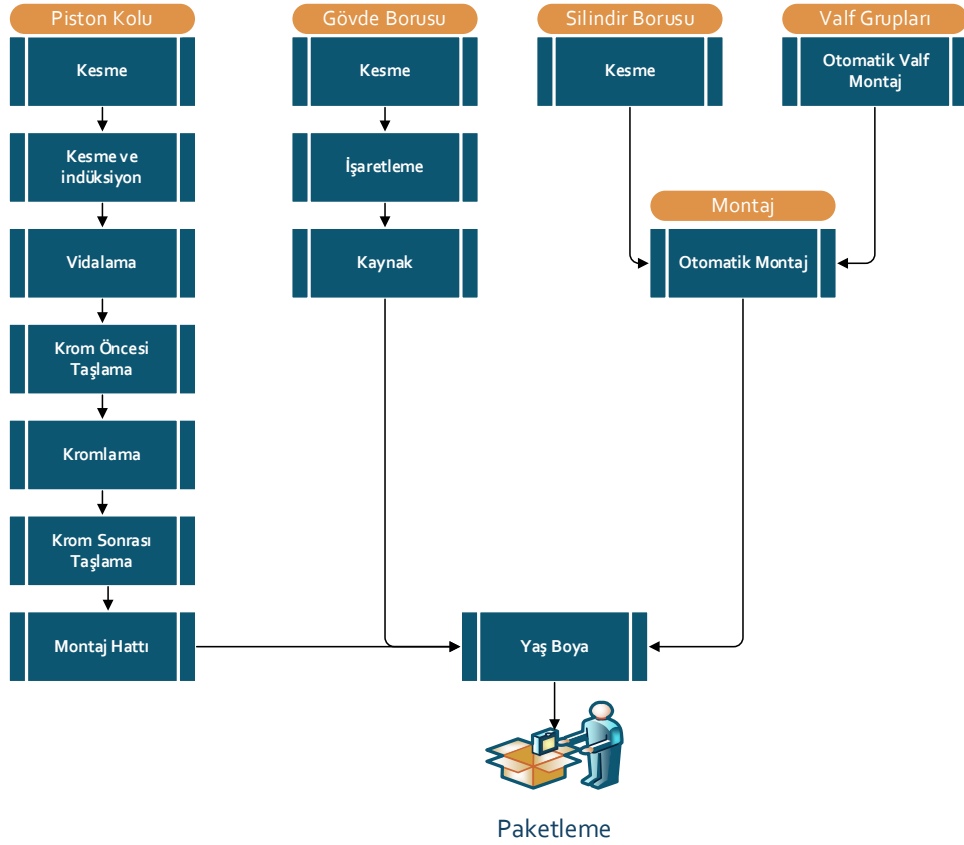
Şekil 3. Teleskopik tip amortisör



Şekil 4. Strut tip amortisör

Amortisörler burçlu lastik, silindir borusu, hidrolik yağ, gövde borusu, alt kapak, yüzük, piston kolu, toz borusu kapağı, keçe, tüm kılavuz, durdurucu pul, segman, tüm piston valfi, tüm taban valfi ve toz borusu gibi parçalardan oluşmaktadır. Üretimde ileri imalat sistemleri metodolojisi kapsamında, sürdürülebilir metodolojilerle inovatif bir imalat süreci adapte edilmeye çalışılmaktadır.

Amortisörün üretim akış şeması Şekil 5'te gösterilmektedir. Üretimin ilk ve en uzun aşaması piston kolu üretimidir. Daha sonra fason olarak fabrikaya gelen gövde boruları ve silindir boruları gerekli işlemlerden geçer. Bir diğer parça olan valf grupları ise temiz oda olarak adlandırılan alanda insan eli değmeden otomatik bir hatta üretilir. Ayrı ayrı işlemlerden geçen bu parçalar ve dışarıdan tedarik edilen parçalar montaja hazır bir şekilde montaj hatlarına getirilir. Parçaların birleştirilmesi için 7 adet montaj hattı bulunmaktadır. Bu hatlarda montaj işlemleri tamamlandığında amortisörler testlerden geçer ve paketleme hattına gönderilir. Paketleme hattından sonra amortisörler sevkiyata hazır hale gelmiş olur.



Şekil 5. Amortisör üretim akış şeması

3. ANALİZ

Amortisör yapısı gereği oldukça hassas bir üründür. İç kısımda oluşacak ufak bir hasar veya gözden kaçan bir hata amortisörün hurda olmasına sebep olmaktadır. İşletme bu yüzden kaliteye çok önem vermekte ve sıfır hata felsefesiyle çalışmaktadır. Üretimin çoğu alanında yüksek otomasyon teknolojileri kullanılmakta ve hemen hemen her istasyonda makineler kalite kontrol yapmaktadırlar. Yine de fabrikada belli noktalarda insan gücü kullanılmakta ve bu da bazı problemlere yol açmaktadır.

Yüksek teknolojilerle üretim yapan işletmenin akıllı fabrikaya dönüşme aşamasında en temel eksiği makineler tarafından tüm verilerin toplanması fakat makineler arası iletişim olmadığı için verilerin makinelerde kalmasıdır. Buna bağlı olarak bakım zamanları çalışanlar tarafından ayarlanmakta, olası hata ve arızalar otomatik tespit edilememektedir. Hatlarda üretim otomatik olarak gerçekleşse bile aralarda insan gücüne ihtiyaç duyulmaktadır.

İşletmede üretim planları üretim planlama mühendisleri tarafından alınan siparişlere ve tahminlere göre hazırlanmaktadır. Hazırlanan bu planlara göre satın

alma departmanı tarafından ham madde siparişleri verilmektedir. Tüm bu işlemler ve verilen siparişlerin takibi, stok durumu, depodan malzemenin üretim hattına çekilmesi gibi işlemler bilgi akışları bilgisayar üzerinden olsa da çalışanlar tarafından karar verilerek yapılmaktadır.

İşletmenin stok sistemi ise şu şekildedir; mamuller için müşteriye göre tanımlanmış alanlara, etiketler yerinden çıkartılmadan barkodların el terminalleri ile okutulması suretiyle stoklanır. Sevkiyatlar tüm müşteriler için IFS Sistemi tarafından öngörülen FIFO bilgisine uygun olarak, en eski lot numarasına öncelik verilecek şekilde sistem tarafından gösterilen adreslerden yapılır. Ham maddeler ise üretim gruplarına göre yerleştirilir. Yerleşimler ve parça aktarımları FIFO yönlerine uygun olarak gerçekleştirilir. Raflarda malzemeler; kasalar ve sevk edildikleri kutular içinde istiflenir. Parçalar üretimden gelen iş emirleri doğrultusunda üretim alanına aktarılır, fiili stok takibi kasalar üzerindeki göz kartları ve sisteme veri girişi yapılarak sağlanır. Raflar üzerindeki adresleme şablonu ile minimum stok, lot büyüklüğü, kasa cinsi bilgilerine ulaşılabilir.

İşletmede malzeme taşıma sistemi ise taşıma arabaları, forkliftler, transpaletler ile gerçekleşmektedir. Fabrikada

makine teçhizat yerleşimi sürece göre yapılmıştır ve hatlar istasyonlara ayrılmıştır. Her istasyonda hangi yarı mamul montajlanıyorsa onunla ilgili teçhizat ve malzemeler bulunmaktadır. İstasyonlar arası yarı mamuller ve malzemeler bu araçlarla taşınmaktadır. Bu da gereksiz insan gücü, vakit kaybı, fazla veya yanlış malzeme taşıma, taşıma sırasında sarsıntı veya düşmelerden kaynaklı hasarlar gibi sorunlar ortaya çıkarmaktadır.

İnsan gücünün etkin olduğu bir diğer hat ise paketleme hattıdır. Hatları tamamlayan ürünler paketleme hattına götürülür. Paketleme için ürünlere özel tasarlanmış kasalar, modüller bulunmaktadır. Katlı gelen modüllerin açılması alt kapama işlemi, amortisörlerin yerleştirilmesi, üst kapama işlemi ve transpalet yerleştirilmesi aşamalarının hepsi manuel olarak yapılmaktadır. Bu durumda paketleme hızının düşük olması, ürünü paketlere yerleştirirken düşürme çarpma gibi oluşabilecek insan kaynaklı etkenler amortisörün hasar görmesi gibi problemler yaratmaktadır.

3.1. Önerilen Akıllı Sistemler

İşletme insan gücünden yaşanan zaman kayıplarını, hataları azaltmak için ve değişen sisteme daha fazla uyum sağlamak için akıllı sistemlerden faydalanmalıdır. Bu alanda atılacak adımlardan biri *Nesnelerin İnterneti Teknolojisi* olabilir. Yüksek otomasyonla çalışan işletmede makineler arası iletişim ve veri akışı sağlanarak makinelerin topladığı veriler ulaşılabilir hale getirilmelidir. Yapılacak bir diğer değişiklik ise her üründe kullanılan barkod sistemi *RFID (Radyo Frekans Tanımlama) Barkod Teknolojisi* ile değiştirilmesidir. Kablosuz ağa bağlanma özelliği bulunan ve nesnelere tanımlayan bu etiketler nesnelerin interneti teknolojisinin temel bileşenlerindedir. Bu sayede nesnelere fabrika içinde gerçek zamanlı izlenebilir ve ürünler hakkında bilgilere ulaşılabilir (Gökalp, Gökalp ve Eren, 2019).

Üretim planlama departmanı: Önerilen akıllı teknolojiler sayesinde işletmede gerçek zamanlı üretim planlama yapılır. Nesnelerin interneti sayesinde otomasyondaki makinelerden, RFID barkodları sayesinde ürünlerden gerçek zamanlı veriler toplanır. Bu veriler *Büyük Veri Yaklaşımları* ile analiz edilir. Elde edilen bilgiler ile hızlı ve *gerçek zamanlı üretim planları* oluşturulabilir, esnek üretim sağlanabilir, üretim hızını etkileyen etkenler gözlemlenebilir, darboğazlar hemen tespit edilerek hızlı çözümler üretilebilir. Ek olarak normalde oldukça zor olan üretim hattı dengeleme etkinliği sistem tarafından iyileştirilir ve verimlilik artar.

Malzeme taşıma sistemleri: İşletmede malzeme taşıma alanında AGV Sistemleri'nin (Otomatik Yönlendirmeli Araçlar) kullanılması önerilmektedir. AGV'ler operatöre ihtiyaç duymadan, lazer veya yeni nesil farklı teknolojilerle taşıma görevini yerine getiren otomatik araçlardır. Önlerine çıkan engelleri anında fark eden bu araçlar sayesinde taşıma yollarını insanlarda rahatlıkla kullanabilir ve yollara herhangi bir ray sistemi veya kablo döşenmesi gerekmez. Bu sayede insan bağımlılığı ortadan kalkar, araçlar ürünleri en kısa yoldan ve optimum hızla taşıyacakları için vakit kayıpları azalır. Aynı zamanda ürünlerde düşme, çarpma, sarsıntı gibi sebeplerden kaynaklanan hasarlarda minimuma indirilmiş olur.

Depolama sistemleri: İşletmenin otomatik depolama sistemlerini kullanması önerilmektedir. Otomatik depolama sistemleri malzemeleri raflara yerleştirmek ve hızlıca erişmek konusunda tasarlanmış sistemlerdir. Bu sayede otomatik makineler aracılığıyla her zaman doğru ürünlere ulaşılır, aramakla vakit kaybedilmez, depolama sırasında ürüne zarar gelmez. Ürün doğrudan istenen noktaya getirilir ve yürüme bekleme gibi zaman kayıpları yaşanmaz. *Gerçek zamanlı stok kontrolü* yapılır.

Paketleme hattı: Otomatik paketleme hattı sayesinde ürün insan eli değmeden paketlenmektedir. Öncelikle insan gücüne gerek kalmadan koli açma makinesi tarafından modüller açılır ve dolmuş için uygun konuma getirilir. Daha sonra robotik dolmuş hattında ürünler modüllere yerleştirilir. Sonrasında ise dolmuş yapılmış modüller koli kapama ve yapılandırma makinesi tarafından kapatılır. Makineden çıkan modül robotik paletleme hattına gelir ve otomatik taşıyıcı araçlar ile sevkiyat alanına götürülür.

Satın alma departmanı: Bu alanda da önceki siparişlerin analizlerini yapmak için *yapay zeka uygulamaları* kullanılır. Bu sayede En uygun tedarikçi belirlenebilmektedir. Ham maddelerin üzerine yerleştirilen RFID barkodları sayesinde ürünün siparişi için gerekli tüm bilgiler saklanır. Depoda kullanılan *kablosuz sensör ağı* sayesinde ham maddenin deponun hangi kısmında ne miktarda olduğu, hangi sipariş için kullanılacağı kayıt altında tutulur ve *gerçek zamanlı depo yönetimi* sağlanmış olur. Üretim için gerekli olduğunda insan gücüne gerek kalmadan ürünler robotlar tarafından depodan getirilir. Bir diğer gelişme ise tedarik alanında yaşanmaktadır. *Nesnelerin İnterneti* ile toplanan veriler *Büyük Veri Yaklaşımları* ile analiz edilir. Depoda bulunan ürün miktarı ile sipariş miktarları eş zamanlı olarak karşılaştırılır ve malzeme

ihtiyaçları karşılanır. Bu sayede stok miktarlarında azalma, üretim sürecinin aksamaması, daha hızlı sipariş kararları ve bunlara bağlı olarak da süreçte hızlanma ve stok maliyetlerinde azalma yaşanır.

4. SONUÇLAR VE TARTIŞMALAR

Bu çalışmada amortisör üretimi yapan bir firmanın üretim sistemleri incelenmiştir. Yapılan incelemeler sonucu işletmenin akıllı fabrikalara adaptasyon sürecinde faydalanması gereken teknolojiler belirlenmiştir. Belirlenen akıllı fabrika teknolojilerinin işletme için yaratacağı avantajlar ve dezavantajlar üzerine tartışılmış ve gerekli önerilerde bulunulmuştur.

İşletme, öncelikle alternatifleri detaylı bir şekilde incelemeli ve önceliklerini belirlemelidir. İşletmenin mühendisleri ve uzman kişiler ile planlı bir şekilde yürütülecek bir çalışma sonucu akıllı teknolojiler işletmede etkin bir şekilde uygulanacaktır. Önerilen akıllı teknolojiler beraberinde birçok fayda ve zorluklarda getirmektedir. Bu zorlukların üstesinden doğru stratejilerle geldiğinde işletme için büyük faydalar sağlanmaktadır. Akıllı teknolojiler sayesinde verimlilikte artışlar, olası sorunlarda hızlı çözüm oluşturabilme yeteneği, operasyon maliyetlerinde azalma, kalitede artış, müşteri memnuniyetinde artış, teslim sürelerinde azalma gibi birçok avantaj elde edilecektir. Aynı zamanda işletme ülke ekonomisine de katkı sağlayacak ve yenilikçi yönüyle rekabet avantajı elde edecektir.

Faydalarının yanında işletmenin karşılaşabileceği ilk yatırım maliyeti, teknik zorluklar, sosyal zorluklar gibi birçok olumsuz etken de vardır. Bu zorlukların arasında işletmeyi en çok zorlayabilecek etken ilk yatırım maliyetleri olacaktır. Robotik, Kablosuz Sensör Ağı, Büyük Veri Teknolojileri, Nesnelerin İnterneti Teknolojisi, Otomatik Taşıma ve Depolama Sistemleri'nin kurulumu ve bunların entegrasyonu oldukça maliyetli uygulamalardır. Tüm sistemlerin aynı anda kurulması veya yanlış noktadan başlanması işletmeyi ileri götürmek yerine ciddi zorluklarla karşı karşıya bırakabilir. Bu yüzden işletme fayda maliyet analizleri yaparak işletme için en çok fayda getirecek olan uygulamadan başlamalıdır ve bu şekilde aşamalı bir geçiş gerçekleştirmelidir. Bir diğer zorluk ise teknik zorluklardır. Çalışanlar mevcut sisteme alışkın oldukları için yeni bir sisteme adaptasyon konusunda zorluk yaşayabilmektedirler veya yetersiz kalabilmektedirler. Bu durum sistemin düzgün bir şekilde işlememesi konusunda oldukça önemlidir. Bunun için çalışanlara eğitim ve gerekli motivasyon sağlanmalı, uygulanan

sistemlere hazırlanmalıdır. Akıllı teknolojilerin henüz yeni olması ve yaygın örneklerinin bulunmaması da teknik zorluklara neden olmaktadır. İşletme uzman kişilerden yardım almalı ve planlı bir şekilde ilerlemelidir. Bu geçiş aşamasında proje yönetimi yaklaşımlarından da faydalanmalıdır.

KAYNAKLAR

Akben, İ. ve Ös, M. (2019). Akıllı ve veriye dayalı tedarik zincirleri. *3.International EMI Entrepreneurship & Social Sciences Congress*, Lefkosa.

Arslan, K. (2020). Artificial intelligence and applications in Education. *Western Anatolia Journal of Educational Sciences*, 11(1), 71-80.

Bilgin, O. ve Işık, H. B. (2018). Dördüncü Sanayi Devrimi ve Türkiye, Ulusal yenilik sistemi çerçevesinde bir inceleme. *Uluslararası Sosyal Araştırmalar Dergisi*, 11(60), 860-867.

Calp, M. H., Bahçekapılı, E. ve Berigel, M. (2018). Endüstri 4.0 kapsamında akıllı fabrikaların incelenmesi. *5th International Management Information Systems Conference*, (116-120. ss.). Ankara.

Çelen, S. (2017). Sanayi 4.0 ve simülasyon. *International Journal of 3d Printing Technologies And Digital Industry*, 1(1), 9-26.

Çırak, A. N. (2019). Yoksullukla mücadelede Endüstri 4.0. *II. Uluslararası Kahramanmaraş Yönetim, Ekonomi Ve Siyaset Kongresi* (106-113. ss.).Kahramanmaraş Sütçü İmam Üniversitesi.

Eryer, A. ve Eryer, S. (2019). Endüstri 4.0 ve Türk lojistik sektörüne etkileri. *II. Uluslararası Kahramanmaraş Yönetim, Ekonomi Ve Siyaset Kongresi*, (84-96. ss.). Kahramanmaraş.

Festo Endüstri 4.0'ın ortaya çıkışı ve temel taşları, (2020, 30 Ocak). Erişim adresi <https://www.blogfesto.com/endustri-4-0-ortaya-cikisi-ve-temel-taslari/>.

Gabaçlı, N. ve Uzunöz, M. (2017). IV.Sanayi Devrimi: Endüstri 4.0 ve otomotiv sektörü. *International Congress on Political, Economic and Social Studies*.

Gökalp, E., Gökalp, M. O. ve Eren, P. E. (2019). Hazır giyim ve konfeksiyon sektöründe endüstri 4.0 devrimi: akıllı konfeksiyon fabrikası. *Online Academic Journal of Informaion Technology*, 10(37), 74-90.

Günay, D. (2002). Sanayi ve sanayileşmenin tarihi. *Mimar ve Mühendis Dergisi*, Sayı 31, 8-14.

Kamber, E. ve Bolatan, G. İ. (2019). Endüstri 4.0 Türkiye farkındalığı. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(30), 836-847.

Kılıç, S. ve Alkan, R. M. (2018). Dördüncü sanayi devrimi endüstri 4.0: Dünya ve Türkiye değerlendirmeleri. *Girişimcilik İnovasyon ve Pazarlama Araştırmaları Dergisi*, 2(3), 29-49.

Koca, K. C. (2018). Sanayi 4.0: Türkiye açısından fırsatlar ve tehditler. *Sosyoekonomi Dergisi*, 26(36), 245-252.

Marr, B. (2017). *Veri stratejisi*. İstanbul: Kapital Medya Hizmetleri A.Ş.

Otomotiv Sanayi Derneği. (2020). *Otomotiv sanayi 2019 yılı küresel değerlendirme raporu*. Otomotiv Sanayi Derneği.

Öztürk, L. (2004). Monte-Carlo simülasyon metodu ve bir işletme uygulaması. *Fırat Üniversitesi Doğu Anadolu Bölgesi Araştırma ve Uygulama Merkezi*, 3(1), 116-122.

Pamuk, N. S. ve Soysal, M. (2018). Yeni sanayi devrimi endüstri 4.0 üzerine bir inceleme. *Verimlilik Dergisi*, Sayı 1, 41-66.

Polat, M. A. (2020). Türk otomotiv sektörünün ekonomik büyümeye etkisi: karşılaştırmalı bir ekonometrik analiz. *Electronic Journal of Social Sciences*, 19(73), 504-521.

Rodic, B. (2017). Industry 4.0 and the new simulation modelling paradigm. *De Gruyter*, 193-207.

Soydal, H. (2006). Türkiye'de doğrudan yabancı sermaye yatırımlarının verimlilik analizi: otomotiv sektörü üzerine bir uygulama. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Sayı 16, 577-604.

Soylu, A. (2018). Endüstri 4.0 ve girişimcilikte yeni yaklaşımlar. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Sayı 32, 44-57.

Sürmen, Y. E. (2019). *Endüstri 4.0 ve otomotiv endüstrisi: Bursa İli Swot Analizi ile Değerlendirilmesi*. Bursa.

Şekkeli, Z. H. ve Bakan, İ. (2018). Akıllı fabrikalar. *Journal of Life Economics*, 5(4), 203-220.

Şimşek, A. (2019). Otomotiv sektöründe nesnelerin interneti uygulamaları üzerine bir derleme. *Black Sea Journal of Engineering and Science*, 2(2), 66-72.

Yıldız, A. (2018). Endüstri 4.0 ve Akıllı Fabrikalar. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22(2), 546-556.

Afetlerden Sonra Kirlilik ve İkincil Kirliliği Afet Olarak Değerlendirmek İçin Bir Tartışma

Sevda ÖZEL*¹ 

*Sivas Cumhuriyet Üniversitesi, Mühendislik Fakültesi, Jeofizik Mühendisliği Bölümü, Sivas, 58140, Türkiye

Araştırma Makalesi, Geliş Tarihi: 23.07.2020, Kabul Tarihi: 19.08.2020

Özet

Afetler, özellikle büyük doğa olaylarından sonra veya yapay ya da antropojenik sebeplerle oluşabilirler ve birden çok çevresel sorunlara sebep olabilirler. Bu olaylar sırasında, kentlerdeki atık uzaklaştırma işlemlerinin ilgili yönetimlerde ve sanayilerde yapılamaması durumunda, kirlilik ve ikincil kirlenme afetleri ortaya çıkabilir. Bu afetler ani olarak gelişebilirler, büyük bir alanı etkileyebilirler, tekrar oluşabilirler, çevresel sorunlara dönüşebilirler ve ikincil kirlenmeye kaynak oluşturabilirler. Ortaya çıkan katı-sıvı-gaz kirleticilerin, sızma-yağış-sel-rüzgar etkisiyle taşınarak, çevre felaketlerine temel sebepler ve kaynaklar oluşturduğu bilinmektedir. Bu nedenle bilinen afetlere ve sorunlara karşı geliştirilen hazırlıklara, kirlilik ve ikincil kirlenme afetleri konusunun da eklenmesi gerektiği gözlenmiştir. Bu afetlerde, katı-sıvı-gaz kirleticilerin daha az bir etkileme alanı oluşturmasına ya da yayılımının durdurulmasına yönelik yeni hazırlıklar yapılması gerektiği düşünülmüştür ve önerilmiştir. Bu konuda, toplum bilinçlenmesine yönelik yeni çalışmalar, planlamalar ve eğitimler tasarlanarak, riskler belirlenmeli ve önlemler alınmalıdır. Çünkü yaşam ortamları ve yaşamsal kaynaklar sürdürülebilir olmalıdır ve sürdürülebilir toplumsal refahın ve çevrenin korunması için bu gereklidir. Bu çalışma ile kirlilik ve ikincil kirlenme bir afet türü olarak kabul görürse, çalışma hedefine ulaşmış olur. Böylece hedeflenen farkındalık oluşabilir, riskleri öngörülebilir ve başka çalışmalar için bir kaynak olabilir.

Anahtar Kelimeler: Afetler, Kirlilik-ikincil kirlenme, Sürdürülebilirlik, Çevre koruma.

A Discussion on the Assessment of Pollution and Secondary Pollution as Disasters After Disasters

Abstract

Disasters can also be occurred especially after major nature events or artificial or anthropogenic event, and cause the multiple environmental problems. During these events, pollution and secondary pollution disasters may occur if the waste disposal processes in cities cannot be carried out by the relevant administrations-industries. These disasters can develop suddenly, affect a large area, re-occur, turn into environmental problems and cause secondary pollution. It has been known that the solid-liquid-gas pollutants transported under the influence of infiltration-precipitation-flood-wind, creating causes and resources that will be the basis for environmental disasters. Therefore, it has been observed that the issue of pollution and secondary pollution disasters should be added to the preparations developed against known disasters and problems. In these disasters, it was thought and suggested that new preparations should be made for the solid-liquid-gas pollutants to create a less impact area or to stop their spread. The new studies, planning and trainings for public awareness should be designed, risks should be determined and measures should be taken. Because living environments and vital resources must be sustainable. This is essential for sustainable social welfare and environmental protection. As a result, if pollution and secondary pollution are accepted as a disaster type with this study, the study will reach its target. Thus, the targeted awareness can be created, its risks are predictable and it can be a source for new studies.

Keywords: Disasters, Pollution-secondary pollution, Sustainability, Environmental protection.

¹Sorumlu yazar sozel@cumhuriyet.edu.tr

1. GİRİŞ

Günümüzde, özellikle gelişmiş ülkelerde, kirlilik afet türleri için bilimsel ve toplumu bilinçlendirme çalışmaları ve hazırlıkları yeterli olsa da bazen tüm bilgi, dikkat ve önlemlere rağmen bu sınırları aşan doğa (deprem, heyelan, tsunami, volkanizma, sel gibi) ve yapay kaynaklı (santral/tehlikeli madde tesisi patlamaları, petrol/radyoaktif madde sızıntıları, kimyasal yangınlar) afetler yaşanmaktadır. Afetlerden etkilenme, bazen de antropojenik (insan kaynaklı) nedenlerle, uygulamada yönetmeliklere uyulmaması veya mevcut yönetmeliklerin değişen koşullara göre zamanında güncellenmemesi nedeniyle de olmaktadır. Bu nedenle bu olaylar sonrasında ortaya çıkan yapı hasarlarıyla ve suyla, çamurla, gaz bulutu veya yağış, sel ve rüzgar etkisiyle gelen yada taşınan kirlleticiler, çevresel etkileri ve riskleri olan büyük-orta-küçük ölçekte alanları etkileyerek, çeşitli türde çevre felaketlerine temel sebepler oluşturabilmektedirler. Çeşitli çevre sorunu da oluşturan bu olaylar, özellikle 1990'lı yıllardan sonra bilimsel araştırmalardaki disiplin çeşidinin de artmasıyla, çok sayıda araştırmaya konu olmayı sürdürmüştür. Özellikle toprak, yer altı-yer üstü suları, hava, hidrokarbon, nükleer-radyoaktif sızıntı, vadoz zon kirlilikleri başta olmak üzere bu konularda yer bilimcilerin çok sayıda çalışma örnekleri Reynolds (1997) ve Rubin ve Hubbard (2005)'in çalışmalarından incelenebilir. Benzer olarak toprak ve su kirliliği, çevresel izleme ve planlama çalışmaları için Karlık ve Kaya (2001), Özürlan ve Çekirge (2007), Özel (2010), Özel vd. (2017) ve Özel (2018) vardır. Diğer disiplin alanlarında ise yine çok sayıda çalışmalar bulunmakta olup, bu makale için afetler kaynaklı çeşitli kirlililerin boyutlarını ve türlerini inceledikleri Silva ve Weerapperuma (2006), Yalçın vd. (2007), VanMeir (2009), Takada (2014) ve Bellonova vd. (2020)'nin çalışmaları örnek verilebilir. Kirlilik ve ikincil kirlenmenin de bir afet türü olarak tanımlanacağı bu makalede, afet ve çevresel sorun olarak yakın zamanda iki büyük olay yaşanmıştır. Birinci olayda, yakın zamanda Türkiye'de bir havai fişek fabrikasında (3 Temmuz 2020, Sakarya/Türkiye) bir patlama oldu ve sonrasında ortaya çıkan yangında insanların yanı sıra önce patlama nedeniyle ve sonrasında yangının da büyümesiyle birlikte büyük bir alanda çevre ve canlı yaşamının olumsuz etkilendiği gözlenen bir afet yaşandı. Gürültü, koku, hava ve görüntü kirliliği sonrasında maddi ve manevi kayıplarla birlikte patlamanın etkisiyle çevreye yayılan gaz sızıntıları ve yangın kaynaklı dumanın insan ve diğer canlıların yaşamını olumsuz etkilediği gözlemlendi. Bu tür bir kirlilik kaynağının etkileri toprak, su ve bitkilerdeki kalıntılarla

uzun bir zaman bu bölgede etkisini gösterebilir. Uzun yıllar sürecek bu etkiler, geçmişte yaşanan büyük Sumatra depremi ve tsunamisi gibi doğa olayları sonrasında veya yapay ve antropojenik (insan) kaynaklı olaylar sonrasında Çernobil, Hiroşima, Fukuşima, gibi büyük felaketlerde de hep gözlenmiştir (Bogdevitch, 2002; Szczeniński vd., 2005; Silva ve Weerapperuma, 2006; Yablokov vd., 2009; Kashparova vd., 2012; Chartin vd., 2013; Takada, 2014; Chandrappa ve Kulshrestha, 2015; Bellonova vd., 2020). Bu araştırmalarda, özellikle Çernobil felaket bölgesine yakın Avrupa ülkelerinde ve Türkiye'nin kuzeyinde yaşayan insanlarda kanser ve genetik bozukluk hastalıklarını arttırdığı, bu ve diğer felaketlerde toprak, su ve bitkiler üzerindeki çevresel etkilerinin (toprak, hava ve sulara gaz ve sıvı kimyasal madde sızıntıları, yine toprak su ve denizlere karışan farklı boyutlarda katı maddeler, radyasyon, vb.) devam ettiği bildirilmiştir. Havai fişek fabrikasındaki patlama olayına benzer yeni bir olay ise yaklaşık bir ay sonra, yine yönetmeliklerle belirlenen uygun depolama ve taşıma koşullarına uyulmayan bir amonyum nitrat (NH_4NO_3) deposunun patlaması (5 Ağustos 2020, Beyrut/Lübnan) ile yaşanmıştır. Benzer afet ve çevresel sorunların yaşandığı ve gözlemlendiği bu patlamanın etkisiyle, canlı yaşamını olumsuz etkileyecek oranlarda zehirli gaz sızıntıları olduğu da gözlenmiştir. Maddi ve manevi kayıpların yanı sıra kaza yaşanan tesisin yakın çevresini ve deniz kenarında olması sebebiyle denizel ortamı da etkileyecek katı-sıvı gaz kirleticilerin yayıldığı gözlenmiştir. Bu nedenle, bu ani gelişen aşırı yada kısmen kirlenme olayları için de yine ülkelerin yeni planlamalar yapılması konusunda, bilimsel ve toplumu bilinçlendirme olarak çalışmaları ve hazırlıkları önemli olmaktadır. Dolayısıyla bilinen afetlere karşı geliştirilen ve öğrenilen hazırlıklara kirlilik afeti konusu da eklenebilir.

Bu makalede ikincil afet türü olarak tanımlanan kirlilik ve ikincil kirlenme afetleri için katı-sıvı-gaz atıkların (kirleticilerin) yerel ve bölgesel ölçekte acil müdahale ile daha az bir etkilenme alanı oluşturması ya da yayılımının hızla durdurulması yönünde hazırlıklar yapılması gerektiği ileri sürülmüştür. Aynı şekilde doğal afetlerde olduğu gibi bu konuda da toplum bilinçlenmesi oluşturulması gerektiği gözlenmiştir. Bu nedenle yapılacak çalışmalar, risklerin belirlenmesi ve gerekli önlemlerin alınması için önemlidir. Çünkü yapılacak çalışmalar, sürdürülebilir yaşam ortamlarının yanı sıra yaşam için gerekli kaynakların sürdürülebilir olması ve sürdürülebilir toplum refahını korumak ve çevre korumada başarılı olmak için de gerekli olacaktır. Bu konuda örneğin depremlerden önce ve sonra kentsel

dönüşüm, kent ve deprem/afetler konulu seminer/sempozyumlar/çalıştaylar ve toplantılar yapılmaktadır. Bunlardan biri de yaklaşık 20 yıldan fazladır her yıl tekrarlanan Türkiye'nin Afet Risk Yönetimi Yuvarlak Masa Toplantısı ile konuya dikkat çeken bu toplantıdır. Bu toplantı gibi diğer seminer, sempozyum ve çalıştaylarda kaynağı doğa olaylarına veya riskli işletmelere bağlı olarak ortaya çıkan çevre sorunları, bu makalede genel olarak yeniden düzenlenen ve özetlenen izlencedeki konu başlıkları temel alınarak kirlilik afetleri tartışılmıştır (www.odtu.edu.tr, 2019).

- *doğal/insan kaynaklı ve teknolojik afetlerin çevresel etkileri,
- *deprem, volkanik patlama, tsunami, fırtına, hortum, kuraklık, küresel ısınma ve küresel iklim değişikliği,
- *afet risk yönetimi, zarar azaltma/sakınım planları,
- *kentsel planlama ve kentsel dönüşüm projeleri, çevre kirliliği, tehlikeli atıklar,
- *tehlike ve hasar görebilirlik modelleri, afet risk analizi, erken hasar tahmini,
- *temel afet eğitimleri, afetlerle başa çıkabilme bilincini artırma,
- *kapsayıcı afet risk azaltma ve özel gereksinimli bireyler.

Bu başlıklar bu makalede, ikincil afet türü olarak adlandırılan kirlilik afetinin ve algısının oluşmasına-artmasına yönelik konulara temel oluşturacak şekilde kirlilik afetini açıklamak için incelenirken, aynı zamanda yeni ikincil kirlenme afetinin kaynakları olarak da tartışılmıştır.

2. AFET VE KİRLİLİK

2.1. Afetlerden Sonra Kirlilik ve Çevresel Sorunlar

Günümüz bilgi ve önlemlerine rağmen hala doğal ve yapay kaynaklı afetlerden sonra ya da kentsel yaşam gereği olan endüstri ve sanayi kaynaklı insan faaliyetleri sonucunda, maddi-manevi kayıplar ve ciddi çevre sorunları yaşanmaktadır. Bilindiği gibi kaynağı ve türü ne olursa olsun kirlilik konusu önemli bir çevre sorunudur ve günümüzde her türlü yasal önlemlere, uyarılara, kirlilik uzaklaştırıcı bilimsel ve teknolojik gelişmelere rağmen hala gittikçe artan bir sorundur. Dolayısıyla ister küresel ölçekte ister bölgesel ve yerel

ölçekte olsun, deprem, volkanizma, tsunami, sel gibi doğa olaylarından sonra veya endüstri ve sanayi kaynaklı olsun, kentsel yaşam ve yapılaşma alanlarından kaynaklanan ikincil bir afet yaşanmaktadır. Bu afet kirliliktir ve yeni bir ikincil kirlenme için de kaynak oluşturabilir. Bu ikincil afet durumu, ani olarak ve büyük ölçekte bir alanı da etkileyebilir. Diğer yandan etkilediği alan başlangıçta küçük olsa bile çok kısa bir zaman içinde bulunduğu girişim alanları yerlerinden hızlı bir yayılımla daha geniş bir etkileme alanı da oluşturabilir. Bu ikincil afet türü, ilk aşamada bir anda ortaya çıkan hava, su, toprak, koku ve görüntü kirliliği olayı olabilir. Ancak doğa olayları sonrası ortaya çıkan kirlilik afeti, yapay yollarla da oluşabilir. Bunlar santral patlamaları, petrol/radyoaktif madde sızıntıları, kimyasal yangınlar (petrol-kimya üretim tesisi yangını, maden kazaları vb.), toplumsal çatışma ve kargaşa dönemindeki sorunlar kaynaklı olabilir. Dolayısıyla sebebi ne olursa olsun doğal ya da yapay kaynaklı kirlilik ve ikincil kirlenme afetleri ortaya çıkabilmektedir. Her ne kadar tüm bu bilinen sorunlar için başlangıçta her türlü önlem alınmış olsa da, aktif deprem kuşakları üzerindeki ülkelerde büyük depremler, okyanus veya denizlere kıyası olan ülkelerde depremlerden sonra tsunami felaketleri, aktif volkanizma yaşanan ülkelerde (örneğin İzlanda ve Pasifik bölgesi ada ülkelerinde) zaman zaman volkanların faaliyete geçmesi ile volkan karakterine göre yoğun dumanlı kötü hava koşulları (hava ve deniz ulaşımını da etkileyen), patlama kaynaklı parçalanmış kaya ve kül yağışları veya akma şeklinde çamur volkanları felaketleri (örneğin Hawaii volkanlarında olduğu gibi) ya da nükleer santral patlamaları, petrol ve kimyasal yangın olayları afetlere sebep olmuştur ve olmaya da devam edecektir. Ancak bir sınıflandırma yapılacak olursa, bu afetler üç grupta toplanabilirler:

Birinci grup afetler: Doğa kaynaklı deprem, heyelan, tsunami, volkanizma, fırtına, sel gibi doğa olaylarından sonra bunlardan kaynaklanan kirlilik ve çevre sorunları. Doğa olayları en sık karşılaşılan olaylardır ve etki alanı afet türüne göre bazen çok büyük bir alanda etkili olabilir. Şehirleri, ülkeleri, bazen bir kıtanın büyük bir bölümündeki yaşamı ve yaşam alanlarını bile etkileyebilir (örneğin; Sumatra (Endonezya) depremi ve tsunamisi, büyük Marmara (Türkiye) depremi).

İkinci grup afetler: Yapay kaynaklı santral patlamaları, kimyasal yangınlar ve sızıntı kaynaklı nükleer, petrol ve benzeri kimyasal olayların yarattığı afetler ve bunlardan kaynaklanan kirlilik ve çevre sorunları. Bu sorunlar aynı zamanda toprak ve su kirliliği kaynaklı besin, su ve tür döngüsünün bozulmasındaki en önemli sebeplerindendir

(örneğin; Çernobil (Ukrayna) ve Fukuşima (Japonya) santrali patlamaları, Meksika körfezi petrol sızıntısı ve yangınları).

Üçüncü grup afetler: Özellikle iç savaşlar, terör, ekonomik kriz ve kötü yönetilme dönemlerinde atık uzaklaştırmada yerel yönetimlerde, sanayilerde, ülke/ülkeler genelinde yaşanan sorunlar nedeniyle kentlerde atık uzaklaştırma düzeninin bozulması, engellenmesi, geçici yapılamaması ya da yeterli arıtma tesisi bulunmaması, sanayilerde katı-sıvı-gaz arıtma yapılmadan atıkların doğaya terk edilmesi ile ortaya çıkan kirlilik ve çevre sorunlarıdır (örneğin; Ortadoğu, Balkanlar ve Afrika'da uzun yıllar süren savaşlarda kullanılan çeşitli savaş malzemeleri kaynaklı gaz ve katı kimyasal atıklar).

Bu sınıflandırmalar gösteriyor ki doğa ya da yapay kaynaklı olaylar yeni bir afete de sebep olabilir. Bu olayların, ani olarak ya da uzun bir süreçte yukarıda anlatılan ve tartışılan nedenlerden dolayı gelişip oluştuğundan sonra özellikle su ve toprak kirliliği başta olmak üzere çeşitli kirlilik türlerine kaynak oluşturabildikleri belirlenmiştir (Bugai vd., 1996; Reynolds, 1997; Karlık ve Kaya, 2001; Rubin ve Hubbard, 2005; Çetiner vd., 2006; Yalçın vd., 2007; Rucker ve Fint; 2007; VanMeir, 2009; Rucker vd, 2009c; Takada, 2014; ADÜTEM, 2015; Özel vd., 2017; Bellonova vd., 2020). Bu çalışmalarda incelendiğinde, doğal ve yapay kaynaklı afetlerden sonra veya insan kaynaklı olarak toprak, yer altı suyu/yüzey suları, maden işleme atık havuzlarıdaki sızıntı kirliliği ve termal sulardaki kirlilik afetlerinin ikincil kirlenme afeti için de kaynak oluşturabileceği gözlenmiştir. Bu nedenle iki kirlilik afetinin de tartışılmasının önemli olduğu düşünülmüştür. Bu anlatılanlara göre kirlilik afeti, ikincil bir afet olarak kabul edilirse, bu afetin kirlenme kaynakları doğal ve yapay kaynak olarak basitçe aşağıdaki gibi gruplandırılabilir.

1. Deprem, volkanizma, tsunami, fırtına, sel, iklim değişikliği gibi çeşitli doğa olayları ile ortaya çıkan ikincil afet olarak kirlilik. Bu olaylar kaynaklı kirlilikte, doğa olayının türüne bağlı olarak kirlilik kaynağı/kaynakları farklı olacaktır. Özellikle bu doğa olayları sonucunda kentsel ortamlarda veya farklı yapılaşma alanlarında (yol, köprü, baraj, tesis) çeşitli yapay atıklar da oluşarak, bunların hepsi toprak, hava, su, koku, görüntü ve gürültü kirlilikleri yaratmaktadır. Böylece ikincil bir afet oluşmakta, bu afet kirliliktir ve bundan ikincil kirlenmeye de kaynak oluşabilir (örneğin büyük Marmara depreminden sonra hasar gören altyapı sistemleri ve sanayi tesislerinden

sızan gaz ve sıvılarla toprak, yer altı suları ve yüzey suları kirliliği, kentsel yapılaşma kaynaklı katı-sıvı-gaz kirlenmelerle deniz kıyısı boyunca deniz kirliliği ve jeotermal kaynaklar yakınında gelişen termal kirlenmeler).

2. Maden ocakları, petrol, kömür, doğalgaz, jeotermal enerji çıkarma/üretim/işleme alanlarında ortaya çıkan organik-inorganik sıvı-katı-gaz kirlenmeler ile ortaya çıkan ikincil afet olarak kirlilik.

Bu kirlilikler çoğunlukla doğal kaynakların çıkarıldığı sahalarında, çıkarılan metalik/metalik olmayan madenin atıkları, ağır metaller, radyoaktif elementler, hidrokarbonlar, sıcak sulardan kaynaklanan kirlenmelerdir. Bunlar doğrudan katı-gaz atıklar veya sıvı ile birlikte yer alan atıklar ve radyoaktif atıklardır. Ya da bu alanlarda oluşan ani bir heyelan, patlama, atmosferik olay, yangın gibi olaylarla etkileşim olması durumunda da bu atıklar muhafaza edildiği ortamlardan veya depolardan sızarak da yer yüzeyinin üstünde ve altında veya koşullar oluştuğu takdirde daha derinlere ilerleyip yayılım göstererek çevre sorunu oluşturacak boyutta kirlilik sebebi olabilirler. Çünkü bu ürünlerin işlendiği riskli işletmelerdeki atıklar doğru yöntemlerle uzaklaştırılmadığı, işletme yer seçimlerinin zamanında doğru yapılamaması, veya sorunlu arıtma tesisleri olması ya da hiç arıtma tesis olmaması gibi kirlilik yaymayacak önlemler alınmadığı takdirde risk oluşturan ikincil afet sebebi kirlilik kaynaklarına dönüşebilmektedirler (örneğin su-toprak-hava-görüntü-gürültü kirliliği sebebi Yatağan (Muğla) ve Tunçbilek (Kütahya) kömür santralleri, termal kirlilik yayan Aydın'daki jeotermal işletmeleri, su-toprak-hava-görüntü kirliliği sebebi Bergama-Ovacık (İzmir) altın madenindeki atık havuzu sızıntıları, su-toprak-görüntü kirliliği sebebi Ergene (Edirne) nehrine deşarj edilen sanayi sıvı atıkları).

3. Kaynağı farklı (örneğin, boru hatları, sanayi ve konut alanı vb. gibi her türlü alanda) büyük yangınlar, kimyasal/nükleer ve diğer kaynaklı patlamalar ve bu patlamalardan sıvı-gaz sızıntılarıyla ortaya çıkan ikincil afet olarak kirlilik.

Bu yangınlar veya patlamalar insan veya doğa kaynaklı bir orman yangını olabileceği gibi konutsal alanlar, işletmeler/çeşitli amaçlar için kullanılan üretim tesisleri ya da boru hatlarındaki çeşitli sebeplerle ortaya çıkan yangınlar veya ani patlama kaynaklı oluşan yangınlar ve sızıntılar da olabilir. Bunun sonucunda farklı kaynaktan/kaynaklardan bir kirlilik afeti oluşabilir ve oluşmaktadır. Bu afet, zamanında yeterli veya başarılı bir müdahale olamaması durumunda yine ikincil bir kirlenme için kaynak da oluşturabilir (örneğin Batman

(Türkiye) bölgesindeki petrol kuyuları çevresinde ve boru hatlarındaki toprak ve yer altı sularına olan sızıntılar, Körfez savaşındaki (Irak) petrol kuyusu yangınları).

4. Toplumsal kargaşa dönemlerinde veya yanlış tarım ve sulama politikaları ile ortaya çıkan ikincil afet olarak kirlilik. Terör, gösteri, iç savaş, doğru atık yönetim sistemlerinin ve tesislerinin olmadığı yerlerde insan yaşamını etkileyen ve yapılaşma alanlarında veya hatlarında ortaya çıkan patlama, yangın, hava yoluyla canlı yaşamını etkileyen gazlar ve belediye hizmetlerinin durması ile evsel/endüstriyel vb. kaynaklı olarak kirlilikler ortaya çıkabilir. Bu afet, zamanında yeterli veya başarılı bir müdahale olamaması durumunda yine ikincil bir kirlenme için de kaynak oluşturabilir. Diğer bir durum ise toprak ve su kaybına sebep olan, tuzlanma, korozyon, kuraklık, toprak ve su kirliliği gibi çevre sorunları ile ortaya çıkan kirliliklerdir. Aşırı pestisit kullanılması, verimli alanlara ya da kullanılabilir ve içilebilir temiz su kaynaklarına kirli su girişi, hava kirliliği gibi sebeplerle, toprak ve su kalitesinin bozulmasına sebep olacak yerel ve bölgesel bir afet oluşması ile ortaya çıkan afet ve bu afet kaynaklı ikincil bir kirlenme oluşabilir (örneğin Atatürk Barajı ile başlayan aşırı sulama ile ve Konya Ovası'nda sulu tarıma geçişle başlayan tarım arazilerinin tuzlanması ile toprak kirliliği).

Dolayısıyla bu olayların olabileceği bölgelerde oluşacak bir veya birden çok doğa olayının veya yapay kaynaklı ya da insan kaynaklı (antropojenik) olayların türüne/türlerine bağlı olarak halkın eğitimi/bilgilendirilmesi bu yönde yapılmalıdır. Başka bir ifadeyle, doğa olayı çoğunlukla bilinen olaylardır ve tüm bilinen önlemlerin alınmasına ve hala geliştirilmekte olan bilimsel önleme yöntemlerine rağmen, patlama, yangın, gaz-sıvı sızıntısı gibi tehlike yaratacak olayların olabileceği işletme veya tesisler de bellidir ve bilinmektedirler. Bu nedenle halk, bu doğa olaylarının veya riskli santral/işletme/tesislerin varlığını, yerini, yaratabileceği tehlikeleri ve acil durumlarda ne yapmaları gerektiğini iyi bilmelidir. Yerel yönetimler ve halk bu konuda her zaman yeterli ve hazırlıklı olmayı sürdürebilmelidir. Böylece günümüzde çevremizde gittikçe artan ve geniş bir alanı etkileyen toprak-su-hava kadar insan ve diğer canlı yaşamını etkileyen afetlerin etkisinin ve çevre sorunlarının azalmasına katkı sağlanabilir.

2.2. Çevresel Etki ve Risk Çerçevesinde Kirlilik ve İkincil Kirlenme Afetleri

Yıllardır artarak devam eden çevresel sorunlar incelendiğinde, olaylar sonrasında toprak, yer altı-yer üstü suları ve atmosfer içindeki kirletici maddelerin ve bitkilerdeki kalıntıların devam ettiği ve çevre ve canlı yaşamında olumsuz etkileri olduğu, hep gözlenmiş ve çeşitli disiplinlerde giderek artan bilimsel araştırmalara konu olmayı sürdürmüştür. Bu nedenle kirlilik olayının ani gelişen ve kısa sürede olumsuz etkileri ortadan kalkan bir sorun olmadığı artık iyi anlaşılmalıdır. Günümüzdeki devam eden çevre sorunlarının kaynağında hep kirlilik yayılımının ve ikincil kirlenme olaylarının devamlılığı olduğu düşüncesine kanıtlar vardır. Çünkü kirletici kaynaklar ve kirleticiler yönetmeliklerle belirlenen uzaklaştırma esalarına uyulmadığı sürece kirlilik yayılımı ve olumsuz etkileri devam edecektir. Buna paralel olarak kirleticiler yayılmasını sürdürürken, yayıldığı alanlarda madde dönüşümleri için gerekli ham madde kaynakları olduğu sürece, ikincil kirlenme için kaynak oluşturmaya da devam edecektir. Bilindiği gibi maddeler katı-sıvı-gaz olsun, toprakta, yeraltı-yerüstü sularında, atmosfer koşullarında, özellikle beslenme zinciri içerisindeki bitki ve hayvan bünyelerinde belli oranlarda bulunmalıdır. Eğer bu oranlar dışında bulunursa, insan ve diğer besin zincirindeki canlıların sağlığı olumsuz etkilenecektir. Çevresel ortamlarda ise doğal arıtma yapmayı engelleyecek oranlarda bu maddelerin su ve toprak ortama deşarj edilmesiyle, toprağın verimsizleşmesine, su veya hava kalitesinin bozulmasına ve ekosistemi bozan türlerin yok olmasına veya göç etmesine gibi sorunlarına sebep olacaktır. Bu nedenle günümüz çevre sorunlarının temeli kirlilik, kirlilik kaynağı olma riski taşıyan tesisler ve ikincil kirlilik kaynaklarıdır. Dolayısıyla bu konuya dikkat çekmenin, bir farkındalık oluşturmanın gerektiği ve önemli olduğu düşünülmüştür.

Doğa olaylarından sonra bazen de yapay kaynaklı olaylar sonucunda ortaya çıkan ve ikincil afet olarak tanımlanan kirlilik olayı, ani olarak büyük ölçekte bir alanı kısa zamanda yada uzun bir süreçte etkileyebilir özellikler de taşımaktadır. Bu nedenle bu olaylardan sonra özellikle yaşam ve yapılaşma alanlarında canlı yaşamının etkileyen çeşitli kirlilik türleri ortaya çıkmakta ve bunlar çeşitli çevre sorunlarına kaynak oluşturmaktadırlar. Bunun sonucunda çoğunlukla toprak, su ve hava ortamlarına etki eden çeşitli kirleticiler özellikle havada, yüzey sularında ve yer altı suyunda ve toprakta kirlilik alanları oluşturmaktadırlar (Şekil 1). Başka bir ifadeyle Şekil 1'de gözlenen kirlilik türleri, 2006-2010 yılları arasında, devam ettirilen izleme çalışmalarıyla birlikte sürekli gözlenmiş çevresel sorunlardandır(Özel,2010).



Şekil 1. Kirlilik afetine de dönüşmüş çevresel sorunlar ve etkileri olan çalışma alanlarından görüntüler (Resimler 2009-2010 arası ve 2014 yılları). Şekil 1.a-b-c’de sızıntı sularıyla kirlenmiş dere suyu, toprak, yüzey suları kirliliği; Şekil 1.d-e-f’de düzensiz depolama alanı toprak, görüntü, gürültü ve hava kirliliği; Şekil 1.g-h-ı’da sıvı-gaz sızıntılı yüzey suyu kirliliği, maden sahasında hava, görüntü, gürültü ve toprak verimsizliği sebebi toprak kirliliği.



Şekil 2. Dünyanın farklı yerlerinde gerçekleşmiş, afete dönüşmüş ve kirlilik olarak ikincil afete de kaynak oluşturmuş önemli doğal ve yapay kaynaklı olaylardan bazı örnekler (www.google.com, 2020).

Şekil 1'deki resimler incelenirse, çevre sorunları çalışılan bazı maden, kentsel ve endüstriyel kaynaklı katı-sıvı-gaz atık alanlarındaki su, toprak, hava, görüntü, gürültü, koku, ağır metal kirliliği ve toprak verimsizleşmesi sorunu gözlenen ve kirlilik afetine de dönüşmüş çevre sorunları ve etkileri olan çalışma alanlarından örnek görüntüler sunulmuştur.

Resimlerde sırasıyla Şekil 1.a-b-c'de sızıntı sularıyla kirlenmiş dere suyu, toprak, yüzey suları kirliliği; Şekil 1.d-e-f'de düzensiz depolama alanlarındaki toprak, görüntü, gürültü ve hava kirliliği; en altta Şekil 1.g-h-ı'da ise sıvı-gaz sızıntılı yüzey suyu kirliliği, sonraki iki resimde ise maden sahalarındaki hava, görüntü, gürültü ve toprak verimsizliği sebebi toprak kirliliklerine örnek görüntüler vardır.

Benzer şekilde Şekil 2 incelenirse, Şekil 2'de sunulan olaylardan sonra ortaya çıkan olası kirlilik türleri ve çevre sorunları için her türlü önlemlere karşın ya da önlemler ihmal edildiği takdirde, doğa ya da yapay kaynaklı afetlerden sonra ikincil afetler de oluşabilmektedir. Örneğin, Çernobil (Ukrayna) ve Fukushima (Japonya) nükleer santral patlamaları, Japonya, Çin, İran, Meksika, Endonezya, Nepal vb. ülkelerdeki büyük depremler, Sumatra Adası ve Japonya depremlerinde büyük bir alanı etkileyen tsunami felaketleri, İzlanda, Hawaii, İtalya ve Pasifik Okyanusu adalarında yanardağ patlamaları ve püskürmeleri, Meksika Körfezi'ndeki büyük çevre olayına dönüşen petrol sızıntısı, Brezilya'daki maden atık baraj setinin ve benzer olarak ülkemizde Kütahya'da siyanür atık baraj setinin çökmesi, Hindistan, Brezilya ve Samsun'daki büyük sel felaketleri, doğal yollarla veya insan kaynaklı olarak Kanada, Amerika, Brezilya (Amazon ormanları bölgesi) ve Avustralya'daki büyük orman yangınları gibi afetler hep olmuştur ve çevresel etkileri devam etmektedir (Bogdevitch, 2002; Szczuciński vd., 2005; Silva ve Weerapperuma, 2006; Yablokov vd., 2009; Kashparova vd., 2012; Chartin vd., 2013; Takada, 2014; Chandrappa ve Kulshrestha, 2015; Bellonova vd., 2020). Bu nedenle bilinen olaylara ait bu örnekler çoğaltılabilir. Dolayısıyla bu çalışmalar ve diğer arşivler incelenirse, çok büyük çevre felaketlerine sebep olan bu tür olayların uzun yıllar geçmesine rağmen, günümüze kadar devam eden etkileri oldukları bilinmektedir ve gözlenmektedir. Bu etkiler, çoğunlukla insan ve canlılar üzerinde etkileri uzun yıllar devam eden genetik bozukluklara bağlı sağlık sorunları ile olay yeri yakın çevresindeki toprakların verimsizleşmesi, bitki örtüsünün çoraklaşması, verimsizleşmesi, türlerin yok olması ve kalan türlerin ise bozulan ekosistem nedeniyle

göç etmesi şeklinde gözlenmiştir. Yine bu olayların yaşandığı çevrede hava olaylarında mevsimsel değişimler gözlenmiş olup, havadaki askıda maddelerin yağışlarla yeryüzüne taşınması ile yüzey suları, topraklar ve insan ve canlı sağlığının olumsuz etkilendiği gözlenmiştir. Topraklarda doğal arıtma için vadoz zonun kalınlığının yeterli olmadığı yerlerde yer altı sularında da kirlenmeler olduğu kanıtlanmıştır. (Reynolds, 1997; Rubin ve Hubbard, 2005; Karlık ve Kaya, 2001; Özel vd., 2017).

Sonuç olarak doğal ve yapay veya insan kaynaklı olaylar nedeniyle afetler yaşandığında bu olaylar sonrasında ortaya çıkan katı-sıvı-gaz kirleticiler,

- kirletici türüne göre veya kirleticinin bir alanı etkileyecek oranda bir alana yayılması durumunda doğal arıtmanın gerçekleşmemesiyle veya böyle bir arıtma ortamı ve koşullarının olmamasıyla ve
- sızma/süzülme, yağış, sel, rüzgar ile taşınarak, sıcaklık ve buharlaşmanın da etkilemesiyle

önemli çevresel sorunlara ve afetlere temel sebepler ve kaynaklar oluşturmaktadırlar. Bu olaylardan sonra, kirleticileri uzaklaştırmak için zamanında müdahale edilemediğinde ya da müdahale çok başarılı ya da yeterince başarılı yapılmadığında, soruna bir çözüm bulunana kadar geçen süreçte, bu kirletici kaynaklar kirleticilerin yayıldığı yeni ortamlarında bu ortamları kirletirken, yeni organik-inorganik dönüşümler geçirebilir ya da başka etkileşimler gösterebilirler. Bu durumda zamanla bu alanlar, yeni bir ikincil kirlilik için de bir kaynak oluşturabilirler. Bu nedenle ikincil afet türü olarak kirlilik afet türü için de bilimsel ve toplumsal çalışmalar yapılması gerektiği düşünülürse, ilk yapılacak olan çalışma, yöreye ya da bölgeye özgü çevresel kirlilik ve her türlü afetler konusunda toplumsal bilinçlenme ve bilimsel modeller üretme olacaktır. Diğer önemli bir durum ise kirlilik yayılımı, yerel veya bölgesel ölçekte olsun ani olarak gerçekleşebileceği gibi, bu yayılmalar zamanla yerel veya bölgesel ölçekte de ortaya çıkabilirler. Ayrıca söz konusu olan ikincil kirlenme gerçekleştiği durumda, etkili olduğu alanda var olan başka bir kirlenme varsa o bölge için ilave kirlilik yükü de meydana gelecektir. Bu nedenle olayın gerçekleştiği alanlarda ve yakın çevresindeki alanlarda risk azalana kadar bir süre bilimsel temellere dayalı izleme çalışmalarının devam etmesi önemlidir. Bu kirlilik yükü alıcı ortamların kapasitesini aşarsa, daha büyük çapta ve uzun süre devam edecek çevresel sorunlarının doğmasına sebep olacaktır. Bu durumda, bu soruna çözüm üreten bir

araştırma ve uygulama planı hazırlanması gerekebileceğinden izleme çalışmalarının önemi büyüktür. Diğer yandan, yeryüzü ve atmosfer ortamında gözlenen kirlilik yayılımları, zamanla yerin altında sığ veya daha derin ortamlarda da tespit edilebilirler. Çok sayıdaki yerbilimsel çalışmalardan elde edilen akademik makaleler bunlara en önemli kanıtlardır (Bugai vd., 1996; Karlık ve Kaya, 2001; Rubin ve Hubbard, 2005; Yalçın vd., 2007; Rucker ve Fint; 2007; Özürkan ve Çekirge, 2007; Frohlich; 2008; VanMeir, 2009; Rucker vd, 2009a, Rucker vd, 2009c; Bugai, 2014; Özel, 2010; Özel vd., 2017, Özel, 2018). Bu çalışmalarda vadoz zon (havalı zon) tanımlanarak, jeolojik tabakalanma ve tabakaların iletkenlik/geçirgenlik/gözeneklik değerleri, kırık/çatlak/boşluk yerleri, yer altı suyu seviyesi, freatik zon özellikleri (tamamen suya doygun zon) ve jeolojik tabakalar ile yer altı suyu sınırları içindeki kirlilik yayılım alanlarının yatay ve düşey sınırlarının haritalanması gibi çalışmalar kirliliğin boyutları ve izleme çalışmaları kapsamındaki kirlilik yayılım durumu hakkında önemli bilgiler sağlamaktadır. Bu nedenle kirlilik önkestim çalışmaları olarak yapılan risk derecelendirme ve haritalama çalışmalarının, yer üstü ve atmosfer için yapılmasının yanı sıra yer bilim araştırma yöntemleri de kullanılarak yer altının sığ ve daha derin ortamları için de yeterince araştırma yapılması ve böylece kirlilik yayılım alanlarının hem yatay hem de düşey yönlerde kesin olarak ortaya konulması gerekmektedir. Sonuç olarak yapılacak çalışmalar iki ayrı grupta toplanabilirler.

a. Toplumsal bilinçlenme

Ülke genelinde doğal ve yapay kaynaklı olabilecek bölgesel ve yerel ölçekte kirlilik türlerine göre risk haritaları hazırlanarak bu konuda bölge ve yerel halkların eğitimleri, bu halkı eğitecek kişilerin eğitimleri yapılmalı, eğitilmiş yardımcı halk ekipleri oluşturulmalıdır. Böylece bilinçli bir halk kitlesi oluşturmak olasıdır.

b. Bilimsel modeller üretme

Katı, sıvı, gaz kirleticilerin etkilerini azaltmaya ya da bertaraf etmeye yönelik planlamalar ve yöntemlerdir. Bu çalışmalar atmosfer, yeryüzü ve yeraltı risk haritaları olarak ayrı ayrı planlanmalıdır ve hazırlanmalıdır. Her türlü laboratuvar analizleri için yetişmiş bir laboratuvar ekibinin olduğu yeterli sayıda ve tam donanımlı laboratuvarlar önemli bir gerekliliktir.

Bu konularda ve oluşabilecek tehlikeler için ve kurtarma, temizleme ve tehlikeyi azaltma çalışmalarında, kirlilik kaynağını en kısa sürede bertaraf

etmeye yönelik yerel yönetimlere, kurum ve kuruluşlara, kurtarma ekiplerine yardımcı olabilecek bilgileri halk edinmiş olmalıdır yada halktan seçilecek kişilerle olası afet durumundaki çalışmalara yardımcı olabilecek halk ekiplerinin kurulması da faydalı olacaktır. Bu ekipler kirleticilerden ve zararlı etkilerinden daha bilinçli ve daha kısa sürede haberdar olacağı için ilgili yerlere zamanında bilgi verilecek, uygun olursa yardım edebilecek ve bu durumda kirlilikten etkilenme süresi, kirlilikten etkilenebilecek canlı sayısı ve kirlenme alanı boyutları daha az olacak yönde katkılar sağlanacaktır. Dolayısıyla afetin kaynağı ne olursa olsun, büyük bir afet olduğunda, özellikle etki alanı ve etkilediği yaşam ortamları büyük olduğunda veya bu alanların koruma açısından önem derecesi yüksek olduğunda, kayıpları azaltmaya yönelik halktan yardımcı ekiplerin de oluşturulması önemli bir destek sağlayacaktır. Böylece yeterli bilgilendirme kitapçıkları ve uyarı levhaları hazırlanmalı ve bunların dijital ortamlarda kolayca güncellenebilir ve dağılıbilir özellikte olması da faydalı olacaktır. Yapılan hazırlıklarla riskli ve güvenli alanlar, risklerin ve afetlerin türleri ve etkileri önceden bilindiğinden, olaylar sırasında ve sonrasında toplumda kargaşa yaşanmadan bilinçli bir şekilde sorunlarla daha iyi mücadele edilecektir. Diğer yandan arıtma tesisleri olmayan işletme ve kentlerde de bu eksikliğin acilen giderilmesi önemlidir. Aynı zamanda bölge halkının da önceden bilinçlendirilmiş olması ile çalışmaların daha sağlıklı, hızlı, kaostan/kargaşadan uzak, planlı ve düzenli yapılmasına olanak sağlanacaktır. Belki, kirlilik afeti için de bir acil telefon hattı oluşturulması da faydalı olabilir. Diğer bir öneri ise afetin türü ve kaynağı ne olursa olsun, sebep olacağı kirlilik türleri için de iller bazında veya bölgesel ölçekte acil durum masalarının önceden hazır olması veya afet olayı olduğunda acilen oluşturulması da faydalı olabilir.

3. SONUÇ VE ÖNERİLER

Her türlü kirlilik bir afet türü olarak kabul edilmelidir ve önemlidir. Bu tür bir afette de, toplumsal bilinçlenme ile toplumun olabilecek katkıları sağlanmalı, eğitim ve bilgilendirme sürecinde her bölgenin jeolojik özellikleri dışındaki coğrafik, kültürel ve sosyolojik yapısı ve nüfus oranları da ayrı ayrı dikkate alınmalı ve bilimsel çalışmalar tamamlanmalıdır. Kirleticilere zamanında müdahale edilemeyen süreç, yeni bir ikincil kirlenmenin kaynağı olarak düşünülmelidir. Bilinen afetlere karşı yapılan hazırlıklara, bu makalede ikincil afet türü olarak tanımlanan kirlilik afeti de eklenirse, diğer afetlerde olduğu gibi afet yardım ekiplerini gerektiğinde destekleyecek bilgi ve donanıma sahip ek yardım

gruplarının/toplulukların/kişilerin oluşturulması faydalı olacaktır. Kirlilik nerede/ne zaman/nasıl tespit edilirse edilsin, tüm afetlerden sonra olduğu gibi çok kısa bir süreçte, bütçe-ekip-ekipmanlar-kurtarma planı oluşturulmalı ve kirlilik afetine müdahale çalışmaları zamanında gerçekleştirilmelidir ve ertelenmemelidir. Çünkü bu hazırlık çalışmaları, sürdürülebilir yaşam ortamlarının sağlanmasının yanı sıra yaşam için gerekli kaynakların da sürdürülebilir olmasına, sürdürülebilir bir toplumsal refahı sağlamaya ve çevre korumaya katkılar sağlayacaktır. Bu nedenle bu makale, kirliliğin ve ikincil kirlenmenin de bir afet türü olarak algılanmasına önemli bir katkı sağlayabilir. Bu konuda farkındalık oluşturabileceği düşünülerek hazırlanan bu makalenin önerileri, uygulamaya geçirilirse, bu konu planlanan hedefe ulaşmış olur ve başka çalışmalar için yol gösterici bir kaynağa da dönüşebilir. Böylece insan gücünün ve modern dünya olanaklarının baş edemeyeceği durumlar dışında, konusunda uzman insanlar ve bilinçlenmiş toplumlar, kıyıların, okyanusların, denizlerin, göllerin, nehirlerin, derelerin, yer altı suyu kaynaklarının, toprağın, ormanların, yeşil alanların ve özel koruma gerektiren sit alanlarının, milli parkların, canlı türlerin, endemic türlerin, yaşam ortamlarının ve havanın korunmasını ve sürdürülebilir olmasını sağlayabilirler ve başarılı olabilirler.

Açıklama: Bu yayın ODTÜ Afet Yönetimi Uygulama ve Araştırma Merkezi Türkiye'nin Afet Risk Yönetimi 21. Yuvarlak Masa Toplantısı'nda sözlü olarak sunulan "İkincil bir afet türü: Kirlilik de bir afet türüdür (Özel, 2019)" başlıklı sunumdan geliştirilmiş ve yayına dönüştürülmüştür.

KAYNAKLAR

ADÜTEM, 2015. Jeotermal Enerjinin Çevresel Etkileri. Adnan Menderes Üniversitesi Jeotermal Enerji Araştırma ve Uygulama Merkezi Yayınları, Yayın No:1, 22.

Bellanova, P., Frenken, M., Reicherter, K., Jaffe, B., Szczuciński, W., Schwarzbauer, J., 2020. Anthropogenic pollutants and biomarkers for the identification of 2011 Tohoku-oki tsunami deposits (Japan). *Marine Geology*, Available online 16 January 2020, 106-117. <https://doi.org/10.1016/j.margeo.2020.106117>.

Bogdevitch, I., Sanzharova, N., Prister, B., Tarasiuk, S., 2002. Countermeasures on natural and agricultural areas after Chernobyl accident. Role of GIS in Lifting the Cloud off Chernobyl, Springer Netherlands, 147-158. doi:10.1007/978-94-010-0518-0_12.

Bugai, D., Bugai D., Waters, R., Dzhepo, S., Skalskij A., 1996. Risks from radionuclide migration to groundwater in the Chernobyl 30-km zone. *Health Physics*, 71(1), 9-18. Doi:10.1097/00004032-199607000-00002.

Bugai, D., 2014. Groundwater contamination following the Chernobyl accident: overview of monitoring data, assessment of radiological risks and analysis of remedial measures. Conference: IAEA TM on Groundwater contamination following Fukushima accident, At: VIC, Vienna, Austria. Doi: 10.13140/RG.2.1.1259.6248.

Chandrappa R., Kulshrestha U.C., 2015. Air pollution and disasters. *Sustainable Air Pollution Management*, 325-343.

Chartin C., Evrarda O., Ondab Y., Patinb J., Lefève I., Otléa C., Ayraulta S., Lepagea H., Bontéa P., 2013. Tracking the early dispersion of contaminated sediment along rivers draining the Fukushima radioactive pollution plume. *Anthropocene*, 1, 23-34.

Çetiner, E.G., Ünver, B., Hindistan, MA., 2006. Maden atıkları ile ilgili mevzuat: Avrupa Birliği ve Türkiye. *Madencilik*, 45(1), 23- 34.

Frohlich, R.K., Barosh, P.J., Boving, T., 2008. Investigating changes of electrical characteristics of the saturated zone affected by hazardous organic waste. *J. Appl. Geophys.*, 64, 25-36.

Kashparova, V., Yoschenkoa, V., Levchuka, S., Bugaib, D., VanMeirc, N., Simonuccic, C., Martin-Garind, A., 2012. Radionuclide migration in the experimental polygon of the Red Forest waste site in the Chernobyl zone-Part 1: Characterization of the waste trench, fuel particle transformation processes in soils, biogenic fluxes and effects on biota. *Applied Geochemistry*, 27(7), 1348-1358.

Karlık, G., Kaya, M.A., 2001. Investigation of groundwater contamination using electric and electromagnetic methods at an open waste-disposal site: a case study from Isparta, Turkey. *Environ. Geol.*, 40(6), 725-731.

Özel, S., 2010. Sivas İli Katı Atık Depolama Alanında Sızıntı Suyunun Yeraltındaki Yayılımının Jeofizik Yöntemlerle İncelenmesi adlı M-352 No'lu CÜBAP Projesi, Cumhuriyet Üniversitesi, Sivas.

Özel, S., Yılmaz, A., Candansayar, M.E., 2017. The examination of the spread of the leachates coming out of a solid waste disposal area on the ground with geophysical and geochemical methods (Sivas, Turkey). *Journal of Applied Geophysics*, 138, 40-49.

Özel S., 2018. Bir deponi alanı sızıntı sularının çevresel etkilerinin Jeofizik ve su kimyası yöntemleriyle incelenmesi. *Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 33(2), 113-124.

Özel, S., 2019. İkincil bir afet türü: kirlilik de bir afet türüdür. ODTÜ Afet Yönetimi Uygulama ve Araştırma Merkezi Türkiye'nin Afet Risk Yönetimi 21. Yuvarlak Masa Toplantısı, 22 Şubat 2019, ODTÜ Kültür ve Kongre Merkezi, Ankara, (yayınlanmamış).

Özürkan, G., Cekirge N., 2007. Hydrogeochemical and geophysical investigation of the Istanbul Tuzla-Icmeler spring area for environmental and land use planning purposes. *Environ. Monit. Assess.*, 132, 125-140. DOI 10.1007/s10661-006-9508-y.

Reynolds, J.M., 1997. An introduction to applied and environmental geophysics. WILEY, England.

Rubin, Y., Hubbard, S.S., 2005. Hydrogeophysics. Springer series, Water Sci. Technol. Library, 50, 519.

Rucker, D.F., Fink, J.B., 2007. Inorganic plume delineation using surface high-resolution electrical resistivity at the BC cribs and trenches site, Hanford. *Vadose Zone J.*, 6 (4), 946-958.

Rucker, D.F., Glaser, D.R., Osborne, T., Maehl, W.C., 2009a. Electrical resistivity characterization of a reclaimed gold mine to delineate acid rock drainage pathways. *Mine Water Environ.*, 28, 146-157.

Rucker, D.F., Levitt, M.T., Greenwood, W.J., 2009c. Three-dimensional electrical resistivity model of a nuclear waste disposal site. *J. Appl. Geophys.*, 69 (3-4), 150-164.

Silva, V., Weerapperuma, D., 2006. Impact of Tsunami Disaster on the Water Environment of Water Environment Federation (WEF), Indian.

Szczuciński, W., Niedzielski, P., Rachlewicz, G., Sobczyński, T., Ziola, A., Kowalski, A., Lorenc, S., Siepak, J., 2005. Contamination of tsunami sediments in a coastal zone inundated by the 26 December 2004 tsunami in Thailand. *Environmental Geology*, 49(2), 321-331.

Takada M., Kamada S, Yajima K., Iwaoka K, Enomoto H., Tabe H., Yonehara H., Sugiura N., 2014. Measurement of radiation environment inside residential houses in radioactive contaminated areas due to the Fukushima nuclear accident. *Progress in Nuclear Science and Technology*, 4, 43-46.

www.odtu.edu.tr, 2019. ODTÜ Afet Yönetimi Uygulama ve Araştırma Merkezi Türkiye'nin Afet Risk Yönetimi 21. Yuvarlak Masa Toplantısı. 22 Şubat 2019 ODTÜ Kültür ve Kongre Merkezi, Ankara.

www.google.com, 2020. wikipedia.org, globalresearch.ca, pinterest.com, en.wikipedia.org, tr.earthquake-report.com/, suhakki.org, http://nordiksimit.org/, bbc.com. 31/01/2020 tarihinde elde edilmiştir.

VanMeir, N., Bugai, D., Kashparov, V., 2009. The experimental platform in Chernobyl: An international research polygon in the exclusion zone for soil and groundwater contamination, radioactive particles in the environment. Springer Netherlands, 197-208. Doi:10.1007/978-90-481-2949-2_13.

Yablokov, A.V., Nesterenko, V.B., Nesterenko, A.V., 2009. Chapter III: Consequences of the Chernobyl catastrophe for the environment. *Chernobyl Consequences of the Catastrophe for People and the Environment (First published)*, 1181(1), 1-327, E1-E39, https://nyaspubs.onlinelibrary.wiley.com, https://doi.org/10.1111/j.1749-6632.2009.04830.x.

Meme Kanseri Tespitinde Sınıflandırma ve Sinir Ağları Yöntemlerinin Karşılaştırılması

Elif ÖZTAD*¹ 

*İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, 34295, Türkiye

Araştırma Makalesi, Geliş Tarihi: 24.07.2020, Kabul Tarihi: 17.08.2020

Özet

Meme kanseri, çoğalan ve çoğu zaman tümör adı verilen bir kitle oluşturan bazı hücrelerin bozulmasından kaynaklanır. Tümörler iyi huylu (kansersiz olmayan) veya kötü huylu (kansersiz) olabilmektedir. MRG (manyetik rezonans görüntüleme), mamogram, ultrason ve biyopsi gibi testler yaygın olarak yapılan meme kanserini teşhis etmek için kullanılır. Verilerde iki farklı etiket olduğundan, tahmin iki kategoriye ayrılır (Kötü huylu veya iyi huylu). Makine öğreniminde bu bir sınıflandırma problemi olarak tanımlanır. Bu çalışma, meme kanserinin iyi huylu veya kötü huylu olup olmadığını sınıflandırmayı ve belirli bir süre sonra kötü huylu vakaların nüksünü ve nüksünü öngörmeyi amaçlamaktadır. Kullanılan metodoloji sınıflandırma modelini ve sinir ağları metodunu içerir. Python modülleri, verileri iyi bir şekilde kavramak ve verilerin farklı şekillerde nasıl ele alınacağını düşünmek için verileri tanımak amacıyla harici veri kümelerini içe aktarmak için kullanılmaktadır. Bu amaçla veri setinden makine öğreniminin temel kavramları uygulanır ve sonuçlar veri setine göre değerlendirilir. Bu nedenle, meme kanseri tahmininde yüksek bir doğruluk elde etmek için sinir ağı yöntemleri ve sınıflandırma yöntemleri birbirleriyle karşılaştırılmaktadır.

Anahtar Kelimeler: Meme kanseri, Sinir ağları, Derin öğrenme, Sınıflandırma.

Comparison Between Classification and Neural Network Methods In Breast Cancer Detection

Abstract

Breast cancer is caused by the breakdown of some cells that multiply and often form a mass called a tumor. Tumors can be benign (not cancerous) or malignant (cancerous). Tests such as MRI (magnetic resonance imaging), mammogram, ultrasound, and biopsy are used to diagnose common breast cancer. Since there are two different labels in the data, the estimate is divided into two categories (benign or malignant). In machine learning, this is defined as a classification problem. This study aims to classify whether breast cancer is benign or malignant and predict the relapse and relapse of malignant cases after a certain period of time. The methodology used includes the classification model and neural network methods. Python modules are used to import external datasets in order to grasp the data well and to think about how to handle the data in different ways. For this purpose, basic concepts of machine learning are applied from the dataset and the results are evaluated according to the dataset. Therefore, neural network methods and classification methods are compared with each other to achieve a high accuracy in breast cancer estimation.

Keywords: Breast cancer, Neural networks, Deep learning, Classification.

¹Sorumlu yazar elifoztad@stu.aydin.edu.tr

1. GİRİŐ

Meme kanserinin kadınlarda en sık teřhis edilen kanser türü olduđu bilinmektedir. Sadece kadınları deđil, çok az bir oranda da olsa erkekleri de etkileyebilmektedir. Bu konuda uzmanlardan biri olan Profesör Aghzadi Rajaa'ya göre, her 10 kadından biri hayatı boyunca meme kanseri riski altındadır. Bu da meme kanserini büyük bir halk sađlığı sorunu haline getirmeye yetmektedir. Meme kanseri için birçok etkili tedavi olmasına rađmen, tedavi sonrasında önemli sayıda hastada hastalık yeniden nüksedebilmektedir ve bu durum yeni tedavi stratejilerinin arařtırılmasını sađlamıřtır (Ouadirhi, 2020).

Hızla geliřen teknoloji, meme kanseri için hayati önem taşıyan erken teřhisin yapılmasına katkı sađlamıř, sınıflandırma algoritmaları ve derin öğrenme ile yüksek doğruluk elde edebilen tahminlerin ortaya çıktıđı projeler geliřtirilmiřtir. Veri madenciliđi yöntemlerinden yararlanarak yapılan tespit, uygulanan tedavilerin agresifliđini önemli ölçüde azaltabilmektedir.

Bu çalışmada, meme kanserinde tümörün iyi huylu veya kötü huylu olup olmadıđını sınıflandırmak ve belirli bir süre sonra kötü huylu vakaların nüksünü öngörmek amaçlanmıřtır.

2. KONULAR

2.1. Meme Kanseri Tespitinde Sınıflandırma Algoritmalarının Kullanılması

Sınıflandırma, kategori üyeliđi makine öğrenimi ve istatistiklerinde bilinen gözlemleri (veya örnekleri) içeren bir veri seti temelinde, yeni bir gözlemin hangi kategorilerden hangisine ait olduđunu belirleme sorunudur (Rana, 2015). Bu çalışmada sınıflandırma algoritmaları kullanılmıřtır. Bu algoritmalar ařađıda belirtilmiřtir.

K-En Yakın Komřu: Bu algoritma, basit olması ve düşük hata oranına sahip olması ile bilinmektedir. Öğrenme için mevcut olan en basit sınıflandırma algoritmalarından biridir. Amaç, özellik alanındaki test verilerinin en yakın eřleşmesini aramaktır (Rana, 2015).

Karar Ağaçları: Bu tür modeller insanın akıl yürütmesine çok benzediđi ve anlaşılması kolay olduđu için sınıflandırma modelleri oluşturmak için yaygın olarak kullanılmaktadır (Kotsiantis, 2013).

DVM – Destek Vektör Makineleri (DVM): Nesnelere etiket atamayı örnek olarak öğrenen bir bilgisayar algoritmasıdır. Örneđin, yüzlerce veya binlerce hileli ve hileli olmayan kredi kartı aktivite raporlarını inceleyerek hileli kredi kartı etkinliđini tanımayı öğrenebilir (Noble, 2006).

Her biri iki kategoriden birine ait olarak iřaretlenmiř bir dizi eğitim örneđi göz önüne alındıđında, bir DVM eğitim algoritması bir kategoriye veya diđerine yeni örnekler atayarak bir olasılık dıřı ikili dođrusal sınıflandırıcı haline getirir (Garg, 2018).

Naif Bayes (NB): Naif Bayes Algoritması, her özellik çifti arasında bađımsızlık varsayımı ile Bayes Teoremi'ne dayanmaktadır. Naif Bayes sınıflandırıcılar, belge sınıflandırması ve spam filtreleme gibi birçok koşulda iyi çalışır. Naif Bayes sınıflandırıcıları, daha sofistike yöntemlere kıyasla son derece hızlıdır.

Lojistik Regresyon (LR): Lojistik Regresyon, sınıflandırma için bir makine öğrenme algoritmasıdır. Bu algoritmada, tek bir denemenin olası sonuçlarını tanımlayan olasılıklar bir lojistik fonksiyon kullanılarak modellenmiřtir ve bu amaç için tasarlanmıřtır. Ayrıca birçok bađımsız deđiřkenin tek bir sonuç deđiřkeni üzerindeki etkisini anlamak için en kullanılıřlı olan algoritmadır (Rouse, 2018). Lojistik regresyon, 1 ve 0 deđerlerini alan bir ikili yanıt deđiřkenini modellemek için bir yöntem sađlar (Bewick, Cheek ve Ball, 2005).

Evriřimsel Sinir Ağları (ESA): İngilizce'de Convolutional Neural Network (CNN) olarak geçen bu algoritma, derin öğrenmede kullanılmaktadır. Diđer algoritmalarından farklı olarak, insan beyninde istemsiz olarak gerçekteřen, objeleri en ince ayrıntısına kadar inceleme ve bađıntı kurma özelliđini taklit eder. Örneđin insanlara Medusa Heykeli gösterildiđinde bu heykelin Medusa Heykeli olduđunu bařındaki yılanlardan anlarlar, yılan ve bař objelerini birleřtirip Medusa olduđuna karar verirler. Derin öğrenmede insanların bu özelliđi taklit edilerek, örneđin; basketbol topu, basketbol ayakkabıları giymiř biri ve potanın olduđu fotođraftan, o kiřinin basketbol oynadıđı çıkarımında bulunulabilir. Objelerin birbirleri ile iliřki ve bađlantıları ortaya konur. Bu algoritmanın 5 katmanı mevcuttur. Katmanları kısaca ařađıdaki gibidir.

Evriřimli Katman: Özellikleri saptamak için kullanılmaktadır (Ergin, 2018).

Dođrusal Olmayan Katman: Sisteme dođrusal olmayanlıđı (non-linearity) tanıtmaktadır (Ergin, 2018).

```
data = pd.read_csv('data/data.csv', index_col=False,)
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...	r
0	842302	M	17.99	10.38	122.8	1001.0	0.11840	0.27760	0.3001	0.14710	...	r
1	842517	M	20.57	17.77	132.9	1326.0	0.08474	0.07864	0.0869	0.07017	...	r
2	84300903	M	19.69	21.25	130.0	1203.0	0.10960	0.15990	0.1974	0.12790	...	r

Şekil 1. Veri setinden alınan bir kesit

Havuzlama (Altörnekleme) Katmanı: Ağırlık sayısını azaltır ve uygunluğu kontrol etmektedir (Ergin, 2018).

Düzleştirme Katmanı: Klasik Sinir Ağı için verileri hazırlamaktadır (Ergin, 2018).

Tamamen Bağlı Katman: Sınıflandırmada kullanılmaktadır (Ergin, 2018).

2.2. Meme Kanseri Tespitinde Sinir Ağı Yöntemlerini Kullanma

Bilgi teknolojisinde (BT), bir sinir ağı, insan beyindeki nöronların çalışmasından sonra şekillendirilen bir donanım ve yazılım sistemidir. Sinir ağları uyarlanabilir olmaları açısından dikkat çekicidir, yani başlangıç eğitiminden öğrendiklerinde kendilerini değiştirirler ve daha sonraki çalışmalar dünya hakkında daha fazla bilgi sağlar (Chi, Street ve Wolberg, 2007). En temel öğrenme modeli girdi akışlarını ağırlıklandırmaya odaklanmıştır, bu da her bir düğümün öncüllerinden gelen girdinin önemini nasıl ağırlaştırdığıdır (Chi, Street ve Wolberg, 2007). Doğru cevapların alınmasına katkıda bulunan girdiler daha ağırdır (Chi, Street ve Wolberg, 2007).

3. ANALİZ METODU

Bu projedeki temel amaç, belirli bir süre sonra kötü huylu vakaların nüksünü tahmin etmektir. Dolayısıyla, tahakkümün burada önemli bir başlangıç noktası vardır. Makineyi eğitmek için meme kanseri tümörünün veri kümesine ihtiyaç vardır. Bu nedenle, meme kanseri veri kümesi, California Üniversitesi tarafından tutulan makine öğrenimi deposundan alınmıştır.

3.1. Veri Kümesinde Bulunan Bilgi Türlerini Tanımlama

Bu projede, verileri iyi bir şekilde kavramak ve verilerin farklı şekillerde nasıl ele alınacağını düşünmek için verileri tanımak amacıyla harici veri kümelerini içe aktarmak için Python modülleri kullanılmaktadır. Şekil 1'de veri setinden alınan bir kesit gösterilmiştir.

3.2. Keşifsel Veri Analizi

Veri keşfi ve görselleştirme teknikleri Python kütüphanelerinde (Pandas, matplotlib, deniz dibi) kullanılmaktadır. Şekil 2'de kullanılan kütüphanelerin kod görüntüsünden bir kesit gösterilmiştir.

```
import numpy as np
import pandas as pd
from scipy.stats import norm
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.model_selection import cross_val_score
from sklearn.pipeline import make_pipeline
from sklearn.metrics import confusion_matrix
from sklearn import metrics, preprocessing
from sklearn.metrics import classification_report
from sklearn.cross_validation import cross_val_score, KFold
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.pipeline import Pipeline
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression

from sklearn.metrics import accuracy_score
```

Şekil 2. Kullanılan kütüphanelerin kod görüntüsünden bir kesit

SVM algoritma sınıflandırıcısı %70 eğitim datasında kullanıldı.

```
clf = SVC(probability=True)
clf.fit(X_train, y_train)
```

30 özellikli bir NumPy dizisinde X, "M" ve "B" dizilerine "1" ve "0" tamsayıları eklendi. Tümör kötü huylu ise, kötü huylu olduğu anlamına gelen "1" yazılacaktır. Aksi takdirde, "0" olarak yazılacak, yani kötü huylu değil, iyi huylu anlamında.

```
array = data.values
X = array[:,1:31]
y = array[:,0]
#transform the class labels from string "M" and "B" to integers
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
y = le.fit_transform(y)
#Call the transform method of LabelEncoder on two dummy variables
le.transform(['M', 'B'])
#Malignant = 1 (indicates prescence of cancer cells)
#Benign = 0 (indicates abscence)
```

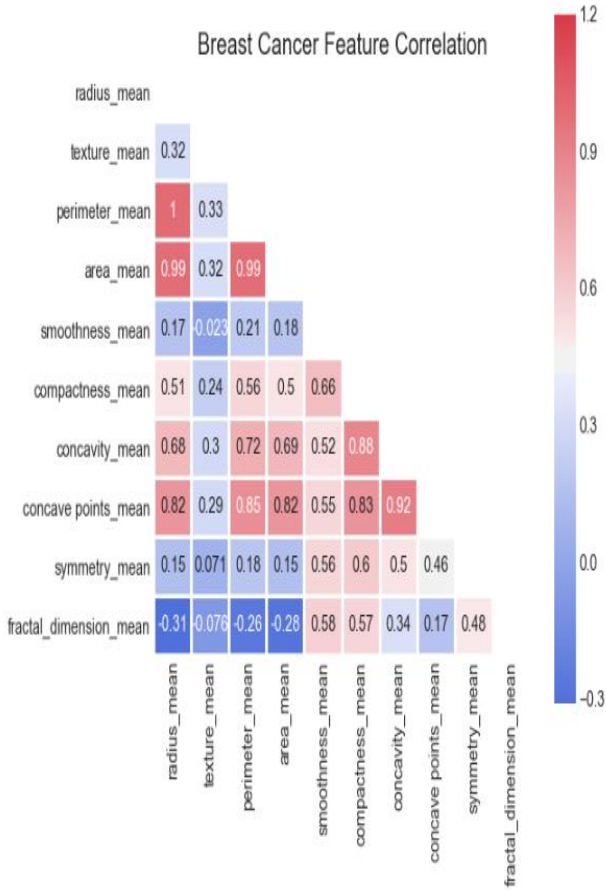
Şekil 3. “M” ve “B” etiketlerinin “1” ve “0” a dönüştürülmesinden bir kesit

Burada string değerleri integer değerlere çevirmek önemlidir çünkü sonuç değeri olarak string kabul edilemez. Şekil 3’te bu etiketlerin “1” ve “0”a dönüştürülmesinden bir kesit gösterilmiştir.

• Doku çekirdeğinin ortalama alanı, yarıçap ve parametrenin ortalama değerleri ile güçlü bir pozitif korelasyona sahiptir.

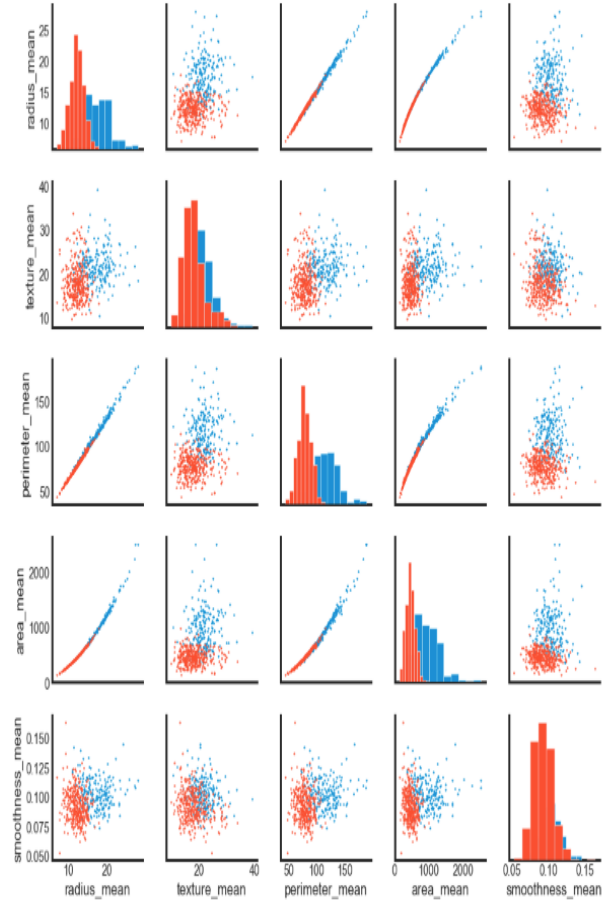
• İçbükeylik ve alan gibi bazı parametreler orta derecede pozitif ilişkilidir.

• Benzer şekilde, fraktal boyut ile yarıçap, doku, parametre ortalama değerleri arasında güçlü bir negatif korelasyon görülmektedir. Aşağıda şekil 5’te kırmızı renk kötü huylu, mavi renk iyi huylu tümörü temsil etmek üzere histogram gösterilmiştir.



Şekil 4. Korelasyon matrisi

Şekil 4’te, 1-0,75 arasındaki ortalama değer parametrelerinde güçlü pozitif ilişkinin olduğu görülmektedir.



Şekil 5. Histogram

- Kanserin sınıflandırılmasında hücre yarıçapı, çevre, alan, kompaktlık, içbükeylik ve içbükey noktaların ortalama deęerleri kullanılabilir. Bu parametrelerin daha büyük deęerleri, kötü huylu tümörler ile bir korelasyon gösterme eğilimindedir.
- Doku, pürüzsüzlük, simetri veya fraktüel boyutun ortalama deęerleri, bir tanının dięerine göre belirli bir tercihini göstermez.
- Histogramların herhangi birinde, daha fazla temizlięi garanti eden fark edilir büyük uç deęerler yoktur.

4. ANALİZ

řekil 6'daki kod görüntüsü sınıflandırma algoritmalarının doęruluk karşılaştırılmasına aittir. řekil 7'deki kod görüntüsü ise CNN algoritmasına aittir.

```
models = []
models.append(('LR', LogisticRegression()))
models.append(('KNN', KNeighborsClassifier()))
#models.append(('DecisionTree',
DecisionTreeClassifier()))
#models.append(('NB', GaussianNB()))
models.append(('SVM', SVC()))
num_folds = 10
num_instances = len(X_train)
seed = 7
scoring = 'accuracy'
num_folds = 10
num_instances = len(X_train)
seed = 7
scoring = 'accuracy'
results = []
names = []
print("COMPARE ACCURACY RESULTS ")
for name, model in models:
    kfold = KFold(n=num_instances, n_folds=num_folds,
random_state=seed)
    cv_results = cross_val_score(model, X_train, y_train,
cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f" % (name, cv_results.mean())
    print(msg)
```

řekil 6. Sınıflandırma algoritmaları kod görüntüsü

```
y_scores = net.predict(test_X)
y_scores = [np.argmax(x) for x in y_scores]
accuracy = accuracy_score(y_true, y_scores)

# inference
print("Network accuracy: {}".format(accuracy))

Network accuracy: 0.767857142857
```

řekil 7. CNN Algoritması kod görüntüsü

5. SONUÇLAR VE TARTIřMALAR

Meme kanserini önleme ve mücadele açısından halkın bu konuda bilgilendirilmesi büyük önem taşımaktadır. Dolayısıyla, medyanın meme kanserine karşı mücadelede etkili bir rolü olduęu aşıkardır. Meme kanserinin erken tespit edilebilmesi iyileşme şansını arttırabileceęi için, erken tespit konusunda literatürde veri madencilięi yöntemlerinden yararlanılarak yapılmış pek çok çalışma mevcuttur. Örneęin Oklahoma Eyalet Üniversitesi'nden arařtırmacıların konuyla ilgili çalışmasında (Delen, Walker ve Kadam (2004) karar ağacı algoritmasının % 93,6 doęruluk verdięini ve tahmin konusunda řuana kadar yapılan çalışmalardaki en yüksek doęruluęa sahip olduęu, yapay sinir ağları ile % 91,2 doęruluk ile ikinci en yüksek doęruluk elde ettikleri, lojistik regresyon ile % 89,2 doęruluk elde ettikleri saptanmıştır. Abdelghani ve Güven'in (2004) yaptıkları çalışmada ise C4.5 Algoritması Naif Bayes Algoritması'nı kullanarak 151.886 kayıttan oluşan veriyi Weka Programı'nı kullanarak analiz etmişlerdir. Doęruluk performansları yüksek sonuçlar vermiştir ancak, verilen veriler için C4.5 Algoritması tarafından üretilen modelin dięer iki teknięe göre çok daha iyi bir performansa sahip olduęunu keşfetmişlerdir.

LR: 0,948450

KNN: 0,932060

DecisionTree: 0,920044

NB: 0,934164

DVM: 0,607918

CNN: 0,767857142857

Burada görüldüğü gibi, Lojistik Regresyon (LR) %94 doğruluk, En Yakın Komşu (KNN) Algoritması %93, Karar Ağacı (DecisionTree) %92, Naif Bayes (NB) %93, Evrimsel Sinir Ağları (CNN) ise %76 doğruluk sağlarken, Destek Vektör Makineleri (SVM) bu veri seti için diğer algoritmalar kadar yüksek bir doğruluk sağlayamamıştır.

Bu çalışmada test edilen sınıflandırma yöntemleri, iyi huylu veya kötü huylu tümör sayısı seçiminde algoritmaların önemini göstermiştir. Çalışma, LR, KNN (En Yakın Komşu), Karar Ağacı ve NB Algoritmaları'nın eğitim sürecinde başarısını göstermiştir. Doğruluk oldukça yüksektir. Ancak DVM Algoritması'nda çok düşüktür. Son olarak, iyi bir veri seti ve o veri setine uygun doğru algoritmayı seçmek bizlere doğruluğu yüksek sonuçlar verecektir.

KAYNAKLAR

Abdelghani, B. ve Guven, E. (2004). Predicting Breast Cancer Survivability using Data Mining Techniques. Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining.

Bewick, V., Cheek, L. ve Ball, J. (2005) Statistics Review 14: Logistic Regression Crit Care 9, 112. doi:10.1186/cc3045.

Chi, L., Street, W. ve Wolberg, H. (2007). Application of Artificial Neural Network Based Survival Analysis on Two Breast Cancer Datasets- AMIA Annu Symp Proc. 2007,130–134.

Delen, D., Walker, G. ve Kadam, A. (2004). Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods. doi:10.1016/j.artmed.2004.07.002.

Ergin, T. (2018). Convolutional Network (ConvNet ya da CNN) Nedir, Nasıl Çalışır. <https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html> erişim adresi.

Garg, R. (2018). 7 Types classification algorithms, *Analytics India Magazine*. <https://www.analyticsindiamag.com/7-types-classification-algorithms> erişim adresi.

Kotsiantis, S. B. (2013). Decision trees: a recent overview, *Artificial Intelligence Review*. vol. 39, 261–283.

Noble, W. S. (2006). What is a support vector machine. *Nature Biotechnology*, 24(12).

Ouardirhi, A. (2020). Meme kanseri: bilgi, farkındalık ve önleme, *Al Bayane*. <http://albayane.press.ma/cancer-du-sein-linformation-la-sensibilisation-et-la-prevention-dabord.html> erişim adresi.

Rana, D. (2015). One class SVM vs SVM classification, *International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064*.

Rouse, M. (2018). Artificial Neural Network (ANN). <https://searchenterpriseai.techtarget.com/definition/neural-network> erişim adresi.

Ag/Aniline blue/n-Si Type Schottky Diode with Organic Dye Interlayers Investigation of Electrical and Photovoltaic Properties

Bahdişen GEZER*¹ , Can GEZER*² 

*¹Uşak University, Engineering Faculty, Department of Electrical-Electronics Engineering, Uşak, 6430, Turkey

*²Sofia Technical University, Engineering Faculty, Department of Aviation Engineering, Sofia, 1113, Bulgaria

Research Article, Received Date: 03.07.2020, Accepted Date: 08.08.2020

Abstract

In this study; Polished part n-Si semiconductor together with a donor concentration of 3×10^{17} /cm, 200 μ m thick, was used. After proper chemical cleaning, an Al ohmic contact was made to the lower face of the semiconductor crystal. After the aniline blue dye was coated with methanol and acetone solution on the upper side, the lower surface of the organic layer was contacted. The electrical and interfacial properties of the manufactured Ag/Aniline blue/n-Si/Al Schottky diode, current-voltage (IV) measurement at room temperature under dark and light, and capacitance (CV) and conductance in the frequency range of 100-900 kHz in 100 kHz steps -Voltage (GV) properties were investigated. Some parameters such as barrier height (ϕ_B) and series resistance (Rs) were obtained from the modified Norde function and Cheung functions using straight feed I-V data. The interface state density (Nss) of the diode was also calculated. In addition, Data obtained as a result of I-V measurements, diode has a good straightening feature.

Keywords: n-type Si, Aniline blue, Schottky diode, Al, Ag.

Organik Boya Ara Katmanlı Ag/Anilin mavisi/n-Si Tipi Schottky Diyot Elektrik ve Fotovoltaik Özelliklerin Araştırılması

Özet

Bu çalışmada, parlatılmış kısım n-Si yarı iletken, 3×10^{17} /cm donör konsantrasyonu ile birlikte, 200 μ m kalınlık kullanılmıştır. Uygun kimyasal temizlemeden sonra, yarı iletken kristalin alt yüzüne bir Al ohm teması yapıldı. Anilin mavisi boyası, üst tarafta metanol ve aseton çözeltisi ile kaplandıktan sonra, organik katmanın alt yüzeyi temas ettirildi. Üretilen Ag/Anilin mavisi/n-Si/Al Schottky diyotunun elektriksel ve arayüzey özellikleri, karanlık ve aydınlıkta oda sıcaklığında akım-voltaj (IV) ölçümü ve 100- frekans aralığında kapasitans (CV) ve iletkenlik 100 kHz'lik adımlarla 900 kHz -Gerilim (GV) özellikleri incelenmiştir. Bariyer yüksekliği (ϕ_B) ve seri direnç (Rs) gibi bazı parametreler, düz besleme I-V verileri kullanılarak değiştirilmiş Norde fonksiyonu ve Cheung fonksiyonlarından elde edildi. Diyotun arayüz durum yoğunluğu (Nss) da hesaplandı. Ayrıca I-V ölçümleri sonucunda elde edilen veriler, diyotun iyi bir doğrultma özelliğine sahiptir.

Anahtar Kelimeler: n-tipi Si, Anilin mavisi, Schottky diyot, Al, Ag.

¹Corresponding Author bahdisengezer@gmail.com, ²cangezer2000@gmail.com

1. INTRODUCTION

The electrical characteristics of the metal-semiconductor (MS) contacts vary depending on the type of interface material used between the metal and the semiconductor. The properties of the material used are important since the interface conditions of metal-semiconductor contacts change many properties of the diode, such as barrier height, ideality factor. Organic semiconductors are considered as the material of the future. Nowadays, many researchers have been working on the use of organic semiconductor materials as interface materials to change the electrical properties of MS structures. Semiconductor materials convert light into current and photodiodes can be obtained as optoelectronic devices. The basis of photodiodes, such as the Schottky diode or n-p joint are semiconductor materials. In recent years, studies focused on the construction of electronic and optoelectronic devices with the formation of organometallic compounds (Sze, 1981). From these applications, dye sensitized solar cells (DSSc) stands out as an alternative to conventional silicon photovoltaic devices. Many solar cells and Schottky diodes, phthalocyanine complexes manufactured using (Rhoderick and William, 1988). Ocak et al. (Ocak et al., 2010) demonstrated the possibility of a new, synthesized Mn hexamide (MnHA) organometallic complex and Schottky diode formation. Thin films of semiconductor organic structure have been useful and very necessary materials for the active operation of electronic and optoelectronic devices. The photovoltaic and electrical properties of MS structures can be changed when an organic layer is added between the metal and the inorganic semiconductor. Obstacle derived from electrical characteristics of Schottky diodes (Ozaydin et al., 2013). The density of the substance providing the interface plays an important role in parameters that determine the properties of the diode, such as height (Φ_b) and the ideality factor (n). Studies in the literature have shown that by increasing the organic thin film on an inorganic semiconductor, the barrier height can be increased or decreased (Oyama et al., 2011).

Çaldıran et al. (Çaldıran et al., 2013) displayed that the Au/Anthracene/n-Si/Al Schottky diode exhibits good rectifier characteristic. They showed the anthracene organic layer increases the influential barrier height of the Au/n-Si/Al diode by creating a physical barrier between Au and n-Si. Orak et al. (Orak and Turut, 2014) investigated the influential of organic layer thickness on electrical and photovoltaic performance by producing organic-inorganic heterojunction. Aydoğan et

al. (Aydoğan and Türüt, 2005) obtained a polyaniline/p-Si/Al metal-insulator-semiconductor (MIS) structure by forming a polyaniline layer on the Si by electrochemical polymerization technique.

Reddy et al. (Reddy et al., 2015); They measured the current voltage properties (I-V) of Ru/Ti/n-InP schottky diodes in the temperature range of 200-400 °C and investigated I-V properties to prediction the Schottky barrier parameters. The results showed that as the temperature increases, the barrier height decreases and the ideality factor increases. These data are consistent with Cheung's $dV/d\ln(I)$ and I . Series resistance values are measured in the range of 16 M Ω - 62 M Ω with enhancement temperature. The nonlinear state at the Richardson constant and Schottky barrier heights was imputed down to the presence of phosphide in the Ru/Ti/n-InP interface.

Güllü et al. (Güllü et al., 2018); They studied the interface, morphological and optical properties of Al/GO/n-InP MIS diode by graphene oxide (GO) intermediate layer between Al/n-InP metal and semiconductor layers. In their study, they measured the barrier height of the reference Al/n-InP diode without GO as 0.43 eV and the barrier height of the Al/GO/n-InP MIS diode with GO interface as 0.85 eV. They investigated that the barrier height of the Al / GO / n-InP diode increased by about 100% relative to the reference Al/n-InP diode.

Yakuphanoglu et al., (Yakupoğlu et al., 2011), obtained the metal structure n-type Si with organic dye. Electronic parameters and photovoltaic properties of Au/Methylene blue/n-Si diodes were measured by current parameters (I-V) and capacitance-conductivity frequency (C-G-f) techniques. They proved the diode exhibited non- optimum attitude due to the presence of organic layer and series resistance. Also; reported that the diode had a photovoltaic attitude with a maximal open circuit voltage of 230 mV and a short circuit current of 20.8 μ A in a solar simulator (100 mW/cm²).

In this work, we researched the electrical and photovoltaic properties of Ag/Aniline blue/n-Si/Al structure less work in the literature. At every stage of this study, the sample was prepared using the ultrasound-assisted technique. Also, after the metal/organic/semiconductor/metal structure is obtained, electrical power in the dark and solar simulator under lighting conditions of 100 mW/cm² photovoltaic parameters were obtained.

2. MATERIALS AND METHODS

2.1. Cleaning and Obtaining of Sample

A 200µm thick n-Si crystal with a donor concentration of $3 \times 10^{15} / \text{cm}^3$ and polished on both sides was used. Since both sides of the sample were polished, no mechanical cleaning of the surfaces was required. To remove organic and inorganic impurities on the crystal and Acetone to remove surface roughness and washed in methanol in an ultrasonic-assisted bath for 10 minutes (Kem, 2008; Gezer, 2018). Flowing in 15 MΩ deionized water for 30 minutes then dried with nitrogen gas (Tatar and Urgen, 2013). Al to be used for the ohmic contact has already been cleaned. Later 10^{-5} torr pressure on wafer placed in vacuum unit thermally Al evaporated to a surface. Sample contact taken from vacuum medium is homogeneous and evaporation to penetrate into the semiconductor 5 min at 400 °C in N₂ medium annealing in the annealing furnace was obtained ohmic contact. 1 cm x 1 cm of wafer with omic contact on one side a piece was cut. On the other surface of the n-type Si wafer, the solution formed by adding 0.0015 M alcohol to 3 µL of aniline blue organic dyestuff was dropped. It was allowed to dry at room temperature under N₂ atmosphere for 75 minutes. Then direct current supply (DC) magnetron sputter to make Schottky contact Ag target plate was placed into the system. With the DC Sputter technique, Schottky contacts by a diameter of approximately 1.5 mm were made on the other surface of the Si wafer. As a result, Ag/aniline blue/n-Si/Al (metal/interface/semiconductor/metal) samples were obtained.

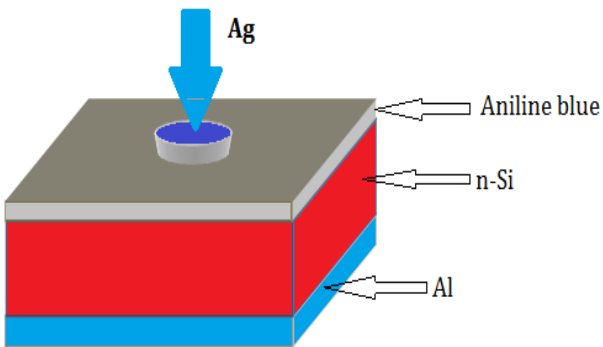


Figure 1. Schematic representation of Ag/Aniline blue/n-Si/Al Schottky diode

The electrical and photovoltaic properties of the obtained Schottky diode (SD) in the dark and in the solar simulator under 100mW/cm² light were investigated. Figure 1 shows the Ag/Aniline blue/n-Si/Al semiconductor Schottky diode obtained using the organic interface material. It is given. Aniline blue organic dyestuff molecular the structure is given in Figure 2.

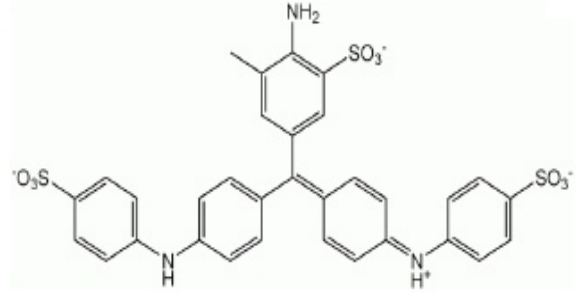


Figure 2. Molecular structure of aniline blue

2.2. Current-Voltage Characteristics

Experimental I-V measurements can be investigated by thermionic emission theory in calculating the ideality factor of the diode (Norde, 1979).

$$I = I_0 \left[\exp \left(\frac{eV_d}{nkT} \right) - 1 \right] \quad (1)$$

After taking ln of both sides in equation (1) according to V if the differential is taken and regulated;

$$n = \frac{q}{kT} \left(\frac{dV}{d \ln I} \right) \quad (2)$$

The ideal factor (n) is a dimensionless parameter that shows deviation of the diode from ideal properties. For an ideal diode, this factor should be approximately equal to 1. The linear part of the lnI-V graph gives the saturation current I₀ at the point at which the fit line intersects the vertical axis at V = 0. (1) the saturation current I₀;

$$I_0 = AA^* T^2 \exp \left(- \frac{q\Phi_B}{kT} \right) \quad (3)$$

is defined as. Here, the diode field A is defined as the A* Richardson constant, the temperature as T Kelvin, k the Boltzmann constant, the charge of q electrons, and the barrier height at B zero voltage. When logarithm of both sides of equation 3 is taken and solved with respect to Φ_B;

$$e\Phi_B = kT \ln \left(\frac{AA^* T^2}{I_0} \right) \quad (4)$$

obstacle height is achieved.

2.3. Obstacle Height and Series with Norde Model Calculation of Resistance

Norde proposed a new method for calculating obstacle height and calculating series resistance (Bohlin, 1986). The modified Norde function can be written as follows.

$$F(V) = \frac{V_0}{\gamma} - \frac{kT}{q} \ln \left(\frac{I(V)}{AA^*T^2} \right) \quad (5)$$

The current-voltage characteristics of Schottky diodes are shown using the function (McLeon, 1986; Zafer, 2006). Where A is the area of the diode, A* is the modified Richardson constant, q is the electronic charge. Considering the minimum point of the function F (V), the barrier height value using the F (V₀) value corresponding to the minimum V₀ voltage;

$$\Phi_B = F(V_0) + \frac{V_0}{\gamma} - \frac{q}{kT} \quad (6)$$

obtained in the form. I₀ corresponding to series resistance value V₀ using the value;

$$R_s = kT \left(\frac{\gamma - n}{qI_0} \right) \quad (7)$$

obtained in the form. Where Y is the first integer greater than n. I₀ is the current value where V potential is minimum (Güllü et al., 2010). This model, which is used by Norde, is valid for ideal cases and cases where the series resistance is small, and it is shown that the series resistance and obstacle height values can be calculated by using the generalized Norde model for non-ideal cases (Tunç and Gökçen, 2012).

2.4. Investigation of Photovoltaic Properties

In order to compare the solar cells produced and characterized in laboratories around the world, all measurements should be made under standard test conditions. According to these conditions, the temperature of the measured solar cell should be 25 °C, the measured radiation intensity should be 100 mW/cm². Air mass AM is the ratio of the sun's rays to the atmosphere when the sun is in the θ° position while the sun is in the atmosphere when the sun is in the Zenith position. Photovoltaic systems generally vary in current, depending on the voltage applied at different radiation intensities, including darkness. They are characterized by measuring the density (Balkanski and Wallis, 2000).

2.5. Measurement Devices Used in the Study

Keithley 2410 programmable constant current source was used for current-voltage (I-V) measurements. The

device in which current-voltage (I-V) measurements are made is shown in Figure 3.

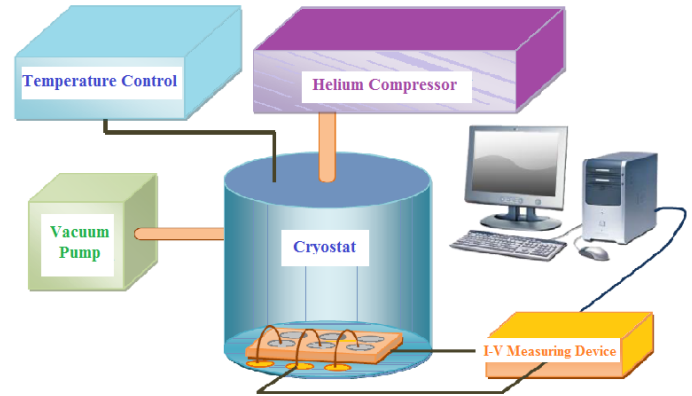


Figure 3. Diagram of the experimental setup used for current-voltage (I-V) measurements

3. RESULTS AND DISCUSSION

3.1. Current-Voltage Characteristics

The values obtained from current-voltage (I-V) measurements and other fixed data are written in (2) and (4) expressions and ideality factors and obstacle height are calculated. In Figure 4, a half-logarithmic current-voltage graph and fit line are given together at room temperature and darkness. For the Ag/Aniline blue/n-Si/Al diode, (2) and (4), the ideality factor was 3,433 and the barrier height was calculated as 0,574 eV (Altındal et al., 2014).

The lower the series resistances of the metal-semiconductor rectifier contacts, the linear the current-voltage characteristics, and the higher the diode quality. Ag/Aniline blue/n-Si/Al Schottky diode series resistance at room temperature and obstacle height values were calculated using Norde Functions.

Figure 4 shows the I-V graphs of Ag/n-Si/Al diodes and Ag/Aniline blue/n-Si/Al diodes used in the study. The n ideality factor calculated for the reference Ag/n-Si/Al diode was n = 1.17 and the barrier height $\Phi_B = 0.51$ eV (Çetinkaya et al., 2013). When Cheung functions were used, the ideality factor of the studied Ag/Aniline blue/n-Si diode was measured as n = 3,461 and barrier barrier = 0.586 eV. The barrier barrier according to Norde functions was measured as $\Phi_B=0.769$ eV

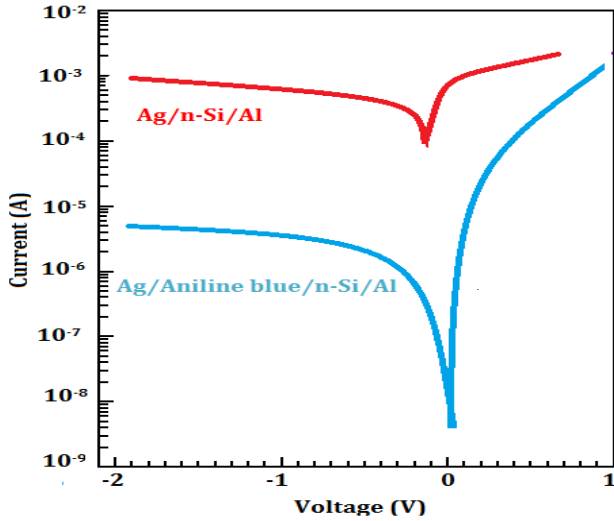


Figure 4. I-V graphs of the Ag/n-Si/Al reference and the organic interface Ag/Aniline blue/n-Si/Al diode without organic interface at room temperature

3.2. Photovoltaic Measurements

Photovoltaic measurements to AM1.5 air mass filter in the solar simulator with 100 mW/cm². Figure 5 shows I-V graphs of Ag/Aniline blue/n-Si/Al diode under dark and light conditions. It is investigated that the diode feedback voltage and current value under light increase by 10 times. There has been an increase in the current of electrons and void pairs resulting from light. It is seen that light does not have a significant effect on current in forward bias value and as the voltage value increases, it is reached in darkness and linearity in light environment. This situation is realized by series resistance effect in accordance with the literature. From the I-V graph data, it is concluded that the studied Ag/Aniline blue/n-Si/Al diode has a rectifying behavior.

3.3. Temperature-dependent Current-Voltage (I-V) Characteristics

Direct forward current-voltage (I-V) characteristics of prepared metal/organic/semiconductor (Ag/Aniline blue/n-Si/Al) Schottky diodes were examined according to thermo-ionic emission theory (TE) (Gökçen and Alli, 2014). According to this theory, current expression,

$$I = I_0 e^{\left[\frac{qV}{nkT} - 1\right]} \quad (8)$$

Here I_0 is the saturation current, which is found in the semi-logarithmic $\ln I - V$ graph by fitting the linear part of the curve to the current axis at zero voltage ($V = 0$) and

$$I_0 = AA^* T^2 e^{\left(\frac{-q\Phi_{B0}}{kT}\right)} \quad (9)$$

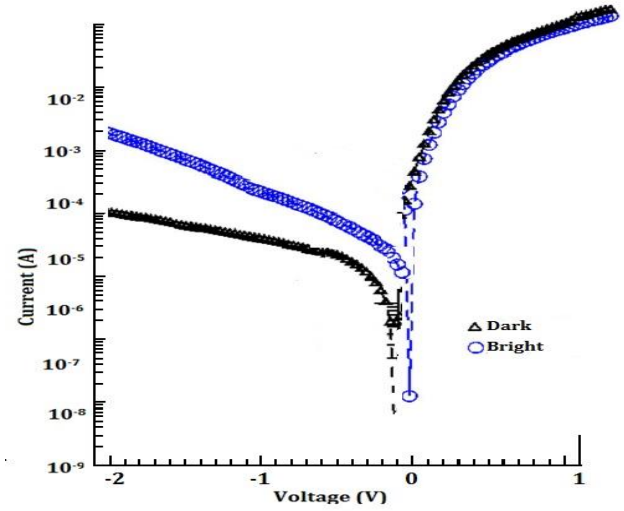


Figure 5. Light and dark medium I-V graphics of Ag/Aniline blue/n-Si/Al diode with organic interface at room temperature (light power 100 mW/cm²)

given with equality. Here, the charge of the electron q , the correct supply voltage applied V , the area of the diode A , the k Boltzmann constant, the effective Richardson constant for A^* n-type GaAs ($A^* = 8.16 \text{ A cm}^{-2} \text{ K}^{-2}$), barrier Φ_{B0} obstacle height at zero feed, The ideality factor of n diode and T is the temperature in Kelvin. Φ_{B0} and n values obtained from the semi-logarithmic correct feed $\ln I - V$ curves are very tightly dependent on the temperature. While the value of Φ_{B0} increases with increasing temperature, n value decreases. Temperature variation of zero feed barrier height (Φ_{B0}) and diode ideality factor (n) are given in Figure 6, respectively. These values for the lowest (70 K) and highest (350 K) temperatures for the obstacle height with the ideal factor, $n = 7.28-2.69$, and $\Phi_{B0} = 0.182-0.801 \text{ eV}$ were obtained respectively. The value of the ideality factor is expected to be in the ideal state ($n = 1$) (Güllü and Turuüt, 2008). However, in practice, this situation is almost never encountered. Because the obstacle height depends on the applied voltage. This dependency leads to a potential drop on the insulating layer, thus the change of the I-V characteristic and the ideality factor to be greater than 1. If the ideal factor is higher than 1, it means that it is difficult to obtain an ideal diode in practice. This increase in the ideality factor can be explained by inhomogeneities in organic film thickness, interface conditions, and uneven distribution of interface loads (Kılıçoğlu et al., 2007). In addition, image load reduction and production-recombination processes are among the reasons explaining the greatness of the ideal factor (Ocak et al., 2015).

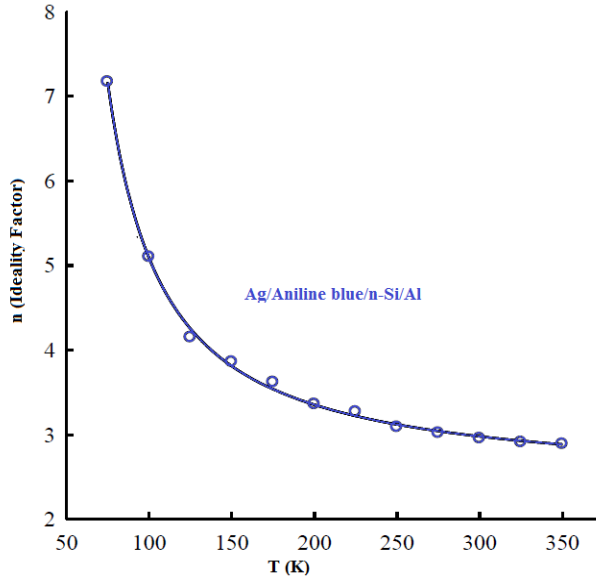


Figure 6. Temperature change graph of the ideal factor of Schottky diode Ag/Aniline blue/n-Si/Al

The fact that the ideal factor is higher than normal means that the current-carrying properties cannot be explained only by the thermo-ionic emission model. In this case, it can be mentioned that there is a secondary mechanism at the interface. Although the organic-inorganic interface is formed in a sharply inactive structure, the organic layer between the metal and the semiconductor appears to cause an important modification in the interface conditions. Therefore, the change in obstacle height can be expressed by the interface dipole through organic layer passivation (Orak et al., 2017).

4. CONCLUSIONS

After obtaining Ag/Aniline blue/n-Si/Al Schottky diode, electrical and photovoltaic properties were investigated in darkness at room temperature (25 °C) and under 100 mW/cm² light in solar simulator. Calculated ideality factor and barrier height values Özeydin et al. (Özeydin et al., 2013; Yakuphanoglu, 2010) calculated in Cu (II) complex/n-Si heterojunction structures. According to the parameters we produce diode has better straightening feature. From the current-voltage (I-V) graph obtained by dark and light measurements, it was seen that the current intensity was 10 times higher than the dark environment with the effect of light at the feedback voltage. This situation is attributed to the increase of electron and hole pairs in the light Aniline blue paint layer which is placed between the metal-semiconductor surface. I-V measurements obtained from the Cheung function $dV/d(\ln I) - I$ serial resistance

value $R_s = 151,854 \Omega$, the ideal factor was measured. When Cheung functions were used, the ideality factor of the studied Ag/Aniline blue/n-Si diode was measured as $n = 3,461$ and barrier height = 0.586 eV. These values are in accordance with the literature.

The value of the current was found to be dependent on the Light state, which is attributed to the increase in load intensity at the interface by the light effect. The Ag/Aniline blue/n-Si/Al diode shows photovoltaic properties. It was observed that the light status did not cause an important change in the parameters in the forward feed. It was seen that the current value increased with the effect of light in reverse feed. It was evaluated that the increase in current caused by the increase in electron and hole pairs due to light effect. When the organic interface is used, it is seen that the flow value decreases by 2 order (100 times) in dark environment and reverse feed. This was attributed to increasing the resistance effect of the organic interface.

Capacitance (C) and Conductance (G) measurements were withstood to be mightily interdependent on the pre-voltage and frequency for the Ag/Aniline blue/n-Si/Al structure. At each frequency, the measured capacitance was found to decrease with increasing frequency due to the continuous distribution of interface states in the 100-900 kHz frequency range (Gezer, 2018). Conductance values rising with rising voltage for each frequency value. It has been found that aniline blue organic compound can be used in semiconductor technology.

REFERENCES

- Altındal, S., Tunc, T., Tecimer, H. and Yucedag I. (2014). Electrical and photovoltaic properties of Au/(Ni, Zn)-doped PVA/n-Si structures in dark and under 250 W illumination level. *Materials Science in Semiconductor Processing*, 28, 48-53.
- Aydoğan, Ş., Sağlam, M. and Türüt, A. (2005). On the barrier inhomogeneities of polyaniline/p-Si/Al structure at low temperature. *Appl. Surf. Sci.*, 250(1), 43-49.
- Balkanski, M., and Wallis, R. F. (2000). *Semiconductor Physics and Applications*, Oxford Univ. Press, New York.
- Bohlin, K. E. (1986). Generalized Norde plot including determination of the ideality factor. *Journal of Applied Physics*, 60(3), 1223.

- Çaldıran, Z., Deniz, A. R., Aydoğan, Ş., Yesildag, A. and Ekinci, D. (2013). The barrier height enhancement of the Au/n-Si/Al Schottky barrier diode by electrochemically formed an organic Anthracene layer on n-Si. *Superlattices and Microstructures*, 56, 45–54.
- Çetinkaya, H.G., Tecimer, H., Uslu, H. and Altındal, S. (2013). Photovoltaic characteristics of Au/PVA (Bi-doped)/n-Si Schottky barrier diodes (SBDs) at various temperatures. *Curr. Appl. Phys.*, 13, 1150-1156.
- Gezer, B. (2018). Studies on an Ultrasonic Synthesis, Characterization, and Thermodynamic Analysis of New Metal Nanocatalysts Applied Directly to Alcohol Fuel Cells. *Arabian Journal for Science and Engineering*, 43, 6203–6209
- Gökçen, M. (2015). Illumination Effects on Electrical Characteristics of Au/Bi₄Ti₃O₁₂/n-Si Structures. *Journal of Nanoelectronics and Optoelectronics*, 10(3), 309-313.
- Gökçen, M. and Alli, A. (2014). Investigation of electrical and photovoltaic properties of Au/poly(propylene glycol)-b-polystyrene/n-Si diode at various illumination intensities. *Philosophical Magazine*, 94(9), 925-932.
- Güllü, O. and Turut, A. (2008). Photovoltaic and electronic properties of quercetin/p-InP solar cells. *Sol. Ener. Mater. Sol. Cell*, 92(10), 1205–1210.
- Güllü, Ö., Asubay, S., Aydoğan, Ş., and Türüt, A. (2010). Electrical characterization of the Al/new fuchsin/n-Si organic-modified device, *Physica E*, 42, 1411-1416.
- Güllü, Ö., Cankaya, M., and Reddy, V. R. (2019). Barrier enhancement of Al/n-InP schottky diodes by grafene oxide thin layer. *Indian Journal of Physics*, 93(4), 467-474.
- Kern, W. (2008). Overview and evolution of silicon wafer cleaning technology Handbook of Silicon Wafer . *Cleaning Technology*, 2.
- Kılıçoğlu, T., Aydın, M. E. and Ocak Y. S. (2007). The determination of the interface state density distribution of the Al/methyl red/p-Si Schottky barrier diode by using a capacitance method. *Phy. B. Cond. Matter*, 388 (1), 244–248.
- McLeon, A.B. (1986). Limitations to the Norde I-V plot. *Semicond. Science Tech.*, 1,177-179.
- Norde, H. (1979). A modified forward I-V plot for Schottky diodes with high series resistance. *Journal of Applied Physics*, 50, 5052.
- Ocak, Y.S., Ebeoğlu, M. A., Topal, G. and Kılıçoğlu T. (2010). Temperature dependent electrical characteristics of an organic–inorganic heterojunction obtained from a novel organometal Mn complex. *Physica B*, 405, 2329-2333.
- Orak, İ., Toprak, M. and Turut, A. (2014). Illumination impact on the electrical characterizations of an Al/Azure A/p-Si heterojunction. *Physica Scripta*, 89, 1158-1165.
- Orak, I., Turut, A. and Toprak, M. (2015). The comparison of electrical characterizations and photovoltaic performance of Al/p-Si and Al/Azure C/p-Si junctions devices. *Synthetic Metals*, 200, 66-73.
- Orak, I., Kocyigit, A. and Turut, A. (2017). The surface morphology properties and respond illumination impact of ZnO/n-Si photodiode by prepared atomic layer deposition technique. *Journal of Alloys and Compounds*, 691, 873-879.
- Oyama, N., Takanashi, Y., Kaneko, S., Momiyama K., Suzuki, K. and Hirose, F. (2011). Pentacene/n—Si heterojunction diodes and photovoltaic devices investigated by I–V and C–V measurements. *Micro. Eng.*, 88 (9), 2959–2963.
- Özaydin, C., Akkılıç, K., İlhan, S., Ruzgar, S., Gullu, O. and Temel, H. (2013). Characterization of an Au/n-Si photovoltaic structure with an organic thin film. *Materials Science in Semiconductor Processing*, 16(4), 1125-1130.
- Reddy, R., Reddy, V., Padmasuvarna, R., and Narasappa, T. (2015). Ru/Ti schottky contacts on n-type In-P (100): Temperature Dependence of Current-Voltage (I-V) characteristics. *Procedia Materials Science*, 10, 666 – 672.
- Rhoderick, E.H. and William, R.H. (1988). Metal-Semiconductor Contacts. 2nd ed. Clarendon, Oxford.
- Sze, S.M. (1981). Physics of semiconductor devices. 2nd ed. New York Wiley.
- Tatar, B., Demiroğlu, D. and Urgan, M. (2013). Structure and photovoltaic properties of Ag/p-CuPc/a-Si/c-Si/Ag organic-inorganic hybrid heterojunction fabricated by chemical spray pyrolysis technique. *Microelectronic Engineering*, 108, 150-157.

Tunc, T., and Gokcen, M. (2012). Preparation and Electrical Characteristic Of Au/n-Si (110) Structure With PVA-Nickel Acetate Composite Film Interfacial Layer. *Journal Of Composite Materials*, 46, 2843-2850.

Yakuphanoglu, F. (2010). Interface control and photovoltaic properties of n-type silicon/metal junction by organic dye. *Journal Alloys Comp.*, 494(2), 451-455.

Yakuphanoglu, F., Ocak, Y. S., Kılıcoglu, T. and Farooq, W. A. (2011). Interface control and photovoltaic properties of n-type silicon/metal junction by organic dye. *Micro. Eng.*, 88, 2951-2944.

Zafer, C. (2006). Organik Boya Esaslı Nanokristal Yapılı İnce Film Güneş Pili Üretimi. *Ege Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi*.