

Yıl:2018

Cilt:2:

Sayı:2

Year:2018

Vol:2

No:2

**UYBİSBBD**

ULUSLARARASI YÖNETİM BİLİŞİM SİSTEMLERİ  
VE  
BİLGİSAYAR BİLİMLERİ DERGİSİ

ULUSLARARASI INTERNATIONAL JOURNAL OF  
YÖNETİM MANAGEMENT  
BİLİŞİM SİSTEMLERİ INFORMATION SYSTEMS  
VE AND  
BİLGİSAYAR BİLİMLERİ DERGİSİ COMPUTER SCIENCE

Cilt: 2 • Sayı: 2 • Aralık 2018  
Vol: 2 • No: 2 • December 2018

**ULUSLARARASI YÖNETİM BİLİŞİM SİSTEMLERİ  
VE  
BİLGİSAYAR BİLİMLERİ DERGİSİ**

**INTERNATIONAL JOURNAL OF MANAGEMENT INFORMATION SYSTEMS  
AND  
COMPUTER SCIENCE**

---

Cilt: 2 • Sayı: 2 • Aralık 2018  
Vol: 2 • No: 2 • December 2018

**Dergi Sahibi (Owner)**

Öğr.Gör. Adem KORKMAZ

**Baş Editör (Editor-in-Chief)**

Öğr.Gör. Adem KORKMAZ

**Editörler (Editors)**

Dr. Öğr. Üyesi Mustafa Mikail ÖZÇİLOĞLU

Dr. Öğr. Üyesi Ayşe ÇİÇEK KORKMAZ

Dr. Öğr. Üyesi Hasan Hüseyin ÇAM

Öğr.Gör. Selma BÜYÜKGÖZE

**Yayın Kurulu (Editorial Board)**

Prof. Dr. Florentin SMARANDACHE

Dr. Öğr. Üyesi Mustafa Mikail ÖZÇİLOĞLU

Dr. Öğr. Üyesi Hasan Hüseyin ÇAM

Dr. Öğr. Üyesi Ayşe ÇİÇEK KORKMAZ

Dr. Bogdan PATRUT

Dr. Iulian FURDU

Dr. Sadiq HUSSAIN

Dr. Svitlana ILNYTSKA

**İngilizce Dil Editörleri**

**(English Language Editors)**

Okt. Abdil Celal YAŞAMALI

Okt. Emrah PEKSOY

**Danışma Kurulu (Advisory Board)**

Prof. Dr. Abdulkadir YILDIZ (Kahramanmaraş Sütçü İmam Üniversitesi)

Prof. Dr. Erdem UÇAR (Trakya Üniversitesi)

Prof. Dr. Florentin Smarandache (University of New Mexico)

Prof. Dr. H. Mustafa PAKSOY (Kilis 7 Aralık Üniversitesi)

Prof. Dr. İsmail Rakıp KARAŞ (Karabük Üniversitesi)

Prof. Dr. Kani ARICI (Kilis 7 Aralık Üniversitesi)

Prof. Dr. Nazım ŞEKEROĞLU (Kilis 7 Aralık Üniversitesi)

Prof. Dr. Sadettin PAKSOY (Kilis 7 Aralık Üniversitesi)

Prof. Dr. Sevinç GÜLSEÇEN (İstanbul Üniversitesi)

Prof. Dr. Ülkü BAYKAL (İstanbul Üniversitesi)

Prof. Dr. Yılmaz Kılıçaslan (Adnan Menderes Üniversitesi)

Doç. Dr. Ercan BULUŞ (Tekirdağ Namık Kemal Üniversitesi)

Doç. Dr. Erdiñ UZUN (Tekirdağ Namık Kemal Üniversitesi)

Doç. Dr. İlhan UMUT (Trakya Üniversitesi)

Doç. Dr. Mustafa ŞEKKELİ (Kahramanmaraş Sütçü İmam Üniversitesi)

Doç. Dr. Yusuf Ekrem AKBAŞ Adıyaman Üniversitesi)

**Adres (Address)**

Kilis 7 Aralık Üniversitesi  
Teknik Bilimler Meslek Yüksekokulu  
79100 Kilis / TÜRKİYE

**E-mail :** [ybsbb.info@gmail.com](mailto:ybsbb.info@gmail.com)

**Web :** [dergipark.gov.tr/uybisbbd](http://dergipark.gov.tr/uybisbbd)

## HAKEM KURULU

<b>Prof. Dr. H. Mustafa PAKSOY</b> (Kilis 7 Aralık Üniversitesi)	<b>Dr. Öğr. Üyesi Ö. Fatih KEÇECİOĞLU</b> (Kahramanmaraş Sütçü İmam Üniversitesi)
<b>Prof. Dr. Abdulkadir YILDIZ</b> (Kahramanmaraş Sütçü İmam Üniversitesi)	<b>Dr. Öğr. Üyesi Zülfiye BIKMAZ</b> (Kırklareli Üniversitesi)
<b>Prof. Dr. Sadettin PAKSOY</b> (Kilis 7 Aralık Üniversitesi)	<b>Dr. Öğr. Üyesi Mustafa Oğuz GÖK</b> (Kahramanmaraş Sütçü İmam Üniversitesi)
<b>Prof. Dr. Mustafa AKSU</b> (İstanbul Üniversitesi)	<b>Dr. Öğr. Üyesi Cuma ERCAN</b> (Kilis 7 Aralık Üniversitesi)
<b>Doç. Dr. Yusuf Ekrem AKBAŞ</b> (Adıyaman Üniversitesi)	<b>Dr. Öğr. Üyesi Ebru KÜLEKÇİ AKYAVUZ</b> (Kilis 7 Aralık Üniversitesi)
<b>Doç. Dr. Kemal Delihacıoğlu</b> (Kilis 7 Aralık Üniversitesi)	<b>Dr. Öğr. Üyesi Halil ARSLAN</b> (Cumhuriyet Üniversitesi)
<b>Doç. Dr. İlhan Umut</b> (Trakya Üniversitesi)	<b>Dr. Öğr. Üyesi Muhammet ATALAY</b> (Kırklareli Üniversitesi)
<b>Doç. Dr. İrfan Deli</b> (Kilis 7 Aralık Üniversitesi)	<b>Dr. Öğr. Üyesi Mustafa Mikail ÖZÇİLOĞLU</b> (Kilis 7 Aralık Üniversitesi)
<b>Doç. Dr. Nursal ARICI</b> (Gazi Üniversitesi)	<b>Dr. Öğr. Üyesi Hasan Hüseyin ÇAM</b> (Kilis 7 Aralık Üniversitesi)
<b>Dr. Öğr. Üyesi Hüseyin KOÇARSLAN</b> (Selçuk Üniversitesi)	<b>Dr. Öğr. Üyesi Ayşe ÇİÇEK KORKMAZ</b> (Kilis 7 Aralık Üniversitesi)
<b>Dr. Öğr. Üyesi Bengü HIRLAK</b> (Kilis 7 Aralık Üniversitesi)	<b>Dr. Öğr. Üyesi Hüseyin AKAR</b> (Kilis 7 Aralık Üniversitesi)
<b>Dr. Öğr. Üyesi Yasin ORTAKCI</b> (Karabük Üniversitesi)	<b>Dr. Öğr. Üyesi Emrah AYDEMİR</b> (Ahi Evran Üniversitesi)
<b>Dr. Öğr. Üyesi Mehmet ÖZÇALICI</b> (Kilis 7 Aralık Üniversitesi)	<b>Dr. Murat GEZER</b> (İstanbul Üniversitesi)
<b>Dr. Öğr. Üyesi Yasin ORTAKCI</b> (Karabük Üniversitesi)	<b>Dr. Serra Çelik</b> (İstanbul Üniversitesi)
<b>Dr. Öğr. Üyesi Ramazan ASLAN</b> (Adıyaman Üniversitesi)	<b>Dr. Fatma Öney KOÇOĞLU</b> (İstanbul Üniversitesi)
<b>Dr. Öğr. Üyesi Sibel YAŞAR</b> (Kırklareli Üniversitesi)	<b>Dr. Fatih AYDIN</b> (Kırklareli Üniversitesi)
<b>Dr. Öğr. Üyesi Şebnem ÖZDEMİR</b> (Beykent Üniversitesi)	<b>Dr. Cemal AKTÜRK</b> (Kilis 7 Aralık Üniversitesi)
<b>Dr. Öğr. Üyesi ALİ DURDU</b> (Ankara Sosyal Bilimler Üniversitesi)	<b>Dr. Tarık TALAN</b> (Kilis 7 Aralık Üniversitesi)
<b>Dr. Öğr. Üyesi Hakan ÜSTÜNEL</b> (Kırklareli Üniversitesi)	<b>Dr. Ahmet Çağdaş SEÇKİN</b> (Uşak Üniversitesi)
<b>Dr. Öğr. Üyesi Mehmet BAKIR</b> (Bozok Üniversitesi)	<b>Dr. Emre AKADAL</b> (İstanbul Üniversitesi)
<b>Dr. Nilüfer VATANSEVER TOYLAN</b> (Kırklareli Üniversitesi)	

## YAYIN POLİTİKASI

Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi yılda iki kez Haziran ve Aralık aylarında yayınlanan uluslararası hakemli bir dergidir. Dergide yer alan yazılar kaynak gösterilmeksizin kısmen ya da tamamen iktibas edilemez. Bu dergide yayınlanan çalışmaların bilim ve dil sorumluluğu yazarlarına aittir.

Dergimize gönderilen çalışmalar, alanında uzman iki ayrı hakem tarafından incelendikten sonra uygun görülenler yayınlanmaktadır. Yazım kurallarına ilişkin bilgilere dergimizin web adresinde yer verilmiştir. Bu derginin tüm hakları saklıdır. Önceden yazılı izin almaksızın hiçbir iletişim ve kopyalama sistemi kullanılarak yeniden kopyalanamaz, çoğaltılamaz ve satılamaz.

---

International Journal of Management Information Systems and Computer Science is an international peer-reviewed journal which is published two times a year in June and December. The articles cannot be cited partly or entirely without showing resources. The responsibility about scientific and grammatical issues is belong to authors.

The papers sent to the journal are reviewed by two referees and after their approval, they will be sent to edit before being published. Writing & Publishing Policies can be found in the journal's website. All rights reserved. No part of this publication may be reproduced, stored or introduced into a retrieval system without prior written permission.

## Makaleler / Articles

### **Kobi Sürdürülebilirliğine Göre Bulanık Mantık Yöntemi İle Destek Atamasının Gerçekleştirilmesi**

*Realizing The Support With The Fuzzy Logic Method By Sustainability Of Smes*

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper

**İbrahim ÇİL & Özge GÜLER & Necdet GÜRSOY** **60-75**

### **Doğal Dil İşlemede Derin Öğrenme Uygulamaları Üzerine Bir Literatür Çalışması**

*A Literature Study On Deep Learning Applications In Natural Language Processing*

Makale Türü: Derleme Makalesi / Paper Type: Review Paper

**Doğan KÜÇÜK & Nursal ARICI** **76-86**

### **Teslacrypt Fidyeye Yazılım Virüsünün Tespiti, Teknik Analizi ve Çözümü**

*Detection, Technical Analysis And Solution Of Teslacrypt Ransomware Virus*

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper

**İlker KARA** **87-94**

### **Spam İçerikli E-Postaların Tespiti İçin Bir Metin Madenciliği Uygulaması: Terimlerin Gama İlişki Katsayısına Dayalı Polarizasyonu**

*A Text Mining Application On The Determination Of Spam-Contented E-Mail: Polarization Of Terms Based On The Gama Relationship Coefficient*

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper

**Ahmet YÜCEL & Meltem KESKİN KÖYLÜ** **95-104**

### **Genetik Algoritma Ve Pikselleştirme Yöntemi İle Mayın Tarlası Oyununun Zorluk Seviyesini Belirleme**

*Determining The Difficulty Level Of The Minefield Game With Genetic Algorithm And Pixelization Method*

Makale Türü: Araştırma Makalesi / Paper Type: Research Paper

**Cemal AKTÜRK** **105-113**

# KOBİ SÜRDÜRÜLEBİLİRLİĞİNE GÖRE BULANIK MANTIK YÖNTEMİ İLE DESTEK ATAMASININ GERÇEKLEŞTİRİLMESİ

\*\*\*

## REALIZING THE SUPPORT WITH THE FUZZY LOGIC METHOD BY SUSTAINABILITY OF SMEs

İbrahim ÇİL\*  
Özge GÜLER\*\*  
Necdet GÜRSOY\*\*\*

### Öz

Günümüzde Türkiye ekonomisinde KOBİ'lerin yadsınamaz bir önemi vardır. KOBİ'lerin sürdürülebilir olmaları ve üretimdeki paylarını büyütebilmeleri Türkiye ve diğer pek çok ülkenin ekonomisinde kritik bir noktadır. Sürdürülebilirlik, bir şirketin faaliyetlerinin genel, sosyal, ekonomik ve çevresel performansını artırmak için uzun vadeli bir hedeftir. Bu çalışmanın amacı da maddi destek bekleyen KOBİ'lerin hangi kriterler altında, hangi tutarlarda desteklenmeleri gerektiğini belirlemektir. Bunun için öncelikle, Türkiye'deki KOBİ'lerin ekonomik olarak hayatlarını ikame ettirebilmeleri ve sürdürülebilir olabilmeleri için onları etkileyen kriterler araştırılmış sonrasında da bu kriterler dikkate alınarak KOBİ'lere verilmesi gereken destek miktarına karar verilmesi hedeflenmiştir. Çözüm aşamasında MATLAB programı kullanılarak Bulanık Mantık yöntemi ile uygun destek miktarları belirlenmiştir.

**Anahtar Kelimeler:** Küçük ve Orta Büyüklükteki İşletmeler, Sürdürülebilirlik, Bulanık Mantık

### Abstract

Today, SMEs in Turkey's economy has an undeniable importance. SMEs be sustainable and to increas etheir share in manufacturing is a critical point in the economy of Turkey and many other countries. Sustainability is a long-term objective to improve the overall, social, economic and environmental performance of a company's operations. The purpose of this study is to determine under the which criteria SMEs who are expecting financial support should be supported. To do this, first of all, the criteria that affected them were researched to be economically sustainable of SMEs in Turkey. After, by considering these criteria, it was aimed to decide on the amount of support that should be given to SMEs. In the solution phase, MATLAB program was used to determine the appropriate amount of support using the Fuzzy Logic method.

**Keywords:** Small and Medium-Sized Enterprises, Sustainability, Fuzzy Logic

\* Prof. Dr. Sakarya Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, [icil@sakarya.edu.tr](mailto:icil@sakarya.edu.tr)  
ORCID: 0000-0002-1290-3704

\*\* Sakarya Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, [ozge.guler2@ogr.sakarya.edu.tr](mailto:ozge.guler2@ogr.sakarya.edu.tr)  
ORCID: 0000-0003-1820-2260

\*\*\* Sakarya Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, [necdetgursoy@gmail.com](mailto:necdetgursoy@gmail.com)  
ORCID: 0000-0002-4119-4695

## 1. GİRİŞ

Türkiye ekonomisinde Küçük ve Orta Büyüklükteki İşletmeler (KOBİ) Türkiye'deki toplam girişim sayısının içinde en büyük oranı oluşturmaktadır, bu da KOBİ'lerin önemini çok net bir şekilde ortaya koymaktadır. Bu bilgileri destekleyecek şekilde 28 ülkenin yer aldığı Avrupa Birliğindeki tüm işletmeler içinde de KOBİ'lerin oranı %99,8'dir (Muller, Gagliardi, Caliandro, Unlu Bohn, Klitou, 2014, p. 14). KOBİ'ler, gelişmekte olan ülkelerde ekonomik büyümenin ve iş yaratmanın itici gücünü oluşturmaktadır. Bir imalat KOBİ'sindeki sürdürülebilirlik kavramının operasyonel seviyeye doğru uygulanabilmesi için, mevcut sürdürülebilirlik gelişiminde nelerin değiştiğine dair net bir algıya sahip olmak ve bir imalat KOBİ'sinin rekabet için nasıl kullanılabileceğini anlamak zorunludur (Mohamed Aboelmaged, 2017).

Nastja Tomsic, Stefan Bojnec ve Blaz Simcic (2015) küçük ve orta ölçekteki işletmelerin rekabetçi pazarlarda başarılarını sürdürmek için ekonomik performanslarını sosyal sorumluluk ve çevre korumasıyla dengeli bir şekilde sürdürmeleri gerektiğini belirtmişlerdir. Ayrıca liderliğin kurumsal sürdürülebilirlik üzerinde önemli bir olumlu etkisi olduğu da birçok araştırmada gösterilmiştir. Buna ek olarak, liderlik yenilik süreci için olumlu ve iş birliğine dayalı bir ortam yaratmalı ve insan sermayesine yatırım yapmalıdır. İnsan sermayesi yenilikçiliğe neden olur ve yenilikçilik kurumsal sürdürülebilirlik ve işletmelerin ekonomik performansı üzerinde önemli bir pozitif etkiye sahiptir. Bu bulgular, seçilen kriterler ile kurumsal sürdürülebilirlik arasındaki eş zamanlı nedenselliğe ilişkin özgün bir yapısal eşitlik modelinin geliştirilmesine yardımcı olmuştur. Buradaki amaç, kurumsal sürdürülebilirliği kıstas olarak bir destek planlama sürecinin iyileştirilmesine yönelik tavsiyelerde bulunmaktır. Literatür taraması sonucu destek planlamasında kullanılmak üzere seçilen yedi adet kriter aşağıda açıklamaları ile birlikte verilmiştir.

**Ekonomik Faktörler:** KOBİ'ler büyük ölçüde ekonomik kaynakların ve insan sermayesinin (Narula, 2004) eksikliğinden dolayı büyük işletmelerden farklı olduğu belirtilmelidir (Haskel ve diğerleri,2007). Sürdürülebilirlik kavramı, “bizden sonraki nesillerin ihtiyaçlarını karşılayabilme gücüne zarar vermeden, bugünün ihtiyaçlarını karşılayan gelişme” olarak tanımlanmaktadır (WCED,1987: 54). Bu tanımda ortaya çıkan zorluklar, öznel, mekânsal ve kültürel bütünleşmenin olduğu ekolojik, ekonomik ve sosyal kriterleri içerdiğinden, Charter ve ark.(2008), sürdürülebilirlik düşüncelerinin (çevresel, sosyal ve ekonomik) şirket yapısına, fikir üretmekten araştırma ve geliştirmeye (Ar-Ge) ve pazarlamaya kadar bütünleştirildiği bir süreç olarak tanımlamışlardır (Charter, M.,Gray, C., Clark, T., Woolman, T., 2008). GRI Sürdürülebilirlik Raporlaması Rehberi, finansal ve finansal olmayan bilgilerin raporlanmasını standartlaştırılmış bir çerçevede birleştirerek, şirketin sürdürülebilirlik performansının ekonomik, çevresel ve sosyal boyutlarda tam olarak karakterize edilmesini sağlamıştır (GRI, 2013). Çalışmada ekonomik performansı etkileyen alt faktörler üretim maliyeti, dağıtım maliyetleri, tedarik maliyetleri, finansal kaynak eksikliği olarak ele alınmıştır. Tseng ve diğ. (2009) bir üretim biriminin sürdürülebilirliğini ölçmek için göstergelerin şunları içermesini önermiştir:

1. Doğal çevre üzerinde enerji ve malzeme etkisi,
2. Ekonomik performans,
3. Toplum gelişimi veya sosyal adalet,
4. Sağlık ve güvenlik ortamı
5. Yeşil ürünler.

**Çevresel Faktörler:** Sürdürülebilirlik, geleneksel üretim uygulamalarının yarattığı tehditler ve paydaşların dayattığı düzenlemeler nedeniyle işletmelerin mevcut toplumda ayakta kalmaları için zorunlu bir sorumluluk haline gelmiştir (Barrett, 1994). Bu nedenle, sürdürülebilir kalkınma kavramı yavaş yavaş dikkat çekmiştir (Govindan vd., 2013). Sürdürülebilir üretim süreci, asgari kaynakları kullanan, çevreye olumsuz etkileri en aza indirmiş ve toplum için uygun maliyetle güvenli ürünler yaratılmasını ima etmektedir. Bu nedenle, işletmeler geleneksel işletme modelini

değiştirmelidir. Şirketin stratejik yönünü göz önünde bulundurarak, aynı zamanda sürdürülebilir kalkınma konusunu stratejik analize dahil etmelidirler (Amrina ve Yusof, 2011). Bununla birlikte, Hillary (2004) 'e göre, KOBİ'ler tüm endüstriyel kirliliğin yüzde 70'inden sorumludur. Buna göre, büyük şirketler ve hükümetler, KOBİ firmalarının sürdürülebilirlik inisiyatiflerine katılımını artırmaya yönelik hareketlere dikkat çekiyor (Jenkins,2009).

**Sosyal Faktörler:** Sürdürülebilir kalkınma kavramı sosyal, ekonomik ve çevresel gelişmeyi içermektedir (Jenkins ve Yakovleva, 2006; Seuring ve Müller, 2008; Erol ve diğerleri, 2011; Govindan ve diğerleri, 2016). GRI Sürdürülebilirlik Raporlaması Rehberi, sosyal faktörleri dört "Alt-Kategoriler" olarak ayırmıştır: İşgücü Faaliyetleri ve İnsana Uygun İş, İnsan Hakları, Toplum ve Ürün Sorumluluğu (GRI, 2013). Bu çalışmada da bu kategorilerden yola çıkılarak sosyal faktörler arasında İşçi/Yönetim İlişkileri, İş Sağlığı ve Güvenliği, Çeşitli ve Eşit Fırsat, Kadınlar ve Erkekler için Eşit Ücret, Ayrımcılık yapmama, Örgütlenme Özgürlüğü ve Toplu Pazarlık Hakkı ele alınmıştır.

**Müşteri İlişkileri:** Tseng (2013) çevre sağlığı ve güvenliği, yerel alandan tedarik edilen tedarikçiler, yerel olarak tüketilen ürünler ve yerel topluluklar için istihdam fırsatlarına ek olarak müşteri şikayetleri ve yanıtları gibi daha ekonomik ve toplumsal kriterlerin dikkate alınması gerektiğini öne sürmüştür.

**Şirket Yapısı:** Kaplan ve Norton (1996), yalnızca operasyonel performansı ölçmenin tek yöntemi olarak finansal boyutun kullanılmasının, işletmenin gerçek özelliklerini ve genel etkinliğini göstermede kısıtlamalara sahip olduğuna dikkat çekmiştir. Bu nedenle, finansal değerlendirmeye ek olarak, iç süreç, müşteri ve büyüme ve öğrenmenin boyutları eklenmiştir. Journeault (2016), sürdürülebilirliğin üç temel performansına, yani çevresel, sosyal ve ekonomik performansa, paydaş yönetim becerilerini ve yeteneklerini ve iç iş süreçlerini entegre eden bir yaklaşım önermiştir. Bilgi temelli teoriye göre, bilgi, her şirketin gelişmesi ve en önemli rekabet avantajı kaynaklarından biri için çok önemli bir kaynaktır (Spicer&Sadler-Smith, 2006). Şirketteki diğer tüm kaynaklar bilgiye bağlıdır (Chirico, 2008). Daha spesifik olarak, üç önemli bilgi türü vardır:

1. Mevcut işletmenin veya yeni kurulan şirketin yönetim ve geliştirme bilgisi;
2. Ürün ve hizmet geliştirme ve üretim süreçleri bilgisi;
3. Yeni bir şirketin faaliyet göstereceği pazar bilgisi.

Audet ve StJean'e (2007) göre, destek hizmetlerinin kullanımını etkileyebilecek ana faktör, sahiplerinin deneyimleridir.

**Çalışanlarla İlgili Faktörler:** Macdonald, Assimakopoulos ve Anderson (2007), KOBİ'ler için eğitim ve öğretim sağlanmasına odaklanmaktadır. KOBİ çalışanlarına yüksek öğrenim ve eğitimi yenilik için gerekli olan kaynaklarla birleştirmeyi denediler. Henry, Hill ve Leitch (2004), İrlanda'da bir dizi niteliksel ve nicel sonuçların yeni iş yaratma eğitim programlarından kaynaklanabileceğini, Bryan'ın (2006) ise küçük imalat işletmelerinde eğitim ve büyüme arasındaki ilişkiyi araştırdıklarını göstermektedir.

**Dış Çevre Faktörleri:** Birçok ülke KOBİ'lere danışmanlık ve eğitim programları sunmaktadır. Kamu danışmanlık hizmetleri (sübvansiyonlu fiyat) piyasa başarısızlıklarını çözmeyi amaçlar: çarpan etkisi, artan eşitsizlikler ve sosyal problemler, teknoloji yayılma etkileri, sosyal istikrar yoluyla artan talep ve ekonomik büyüme, tek girişimcinin faydalarının çok ötesinde sosyal faydaları teşvik eder (Schwartz, Bar-El, 2004). Berry, Sweeting ve Goto (2006), dış tavsiyelerin KOBİ'lerin büyüme oranı ile pozitif ilişkili olduğunu belirtmiştir. Mason ve Brown (2013), KOBİ'lerin yüksek büyümesi için uygun hükümet politikalarının nasıl oluşturulacağını tartışmaktadır. Politika yapıcılarının KOBİ'leri yanlış yerlerde aradıklarını iddia ettiler. KOBİ'lerin sektör, yaş, büyüklük ve köken açısından heterojen doğası, belirli sektörlerle, teknolojilere veya şirket türlerine destek olmayı imkânsız kılmaktadır. Yüksek düzeydeki bürokrasi ve idari engeller



yeni iş girişimleri üzerinde olumsuz etkilere sahiptir (Patzelt ve Shepherd, 2009). Ayrıca, vergi teşvikleri, yeni iş girişimleri için devlet desteğinin mükemmel biçimleri olabilir. Wiggins ve diğ. (2010), küçük işletmelerin gelişimlerine dolaylı olarak bağlı olan kırsal altyapıya (yol, elektrik, fiziki altyapı, eğitim, güvenli ve sağlıklı su, kırsal kurumların işleyişi gibi kamu mal ve hizmetlerinin sağlanması) katkıda bulunulmasının gerekliliğine değinmişlerdir.

Yukarıda açıklamaları verilen ana kriterlerin ağırlıklarını değerlendirmek ve KOBİ'lerin destek miktarlarının belirlenmesi sağlamak için bir Bulanık Mantık modeli önerilmiştir. Günlük hayatta kendimizi ifade etmek, bir olayı anlatmak veya komut vermek için kullandığımız sözel ifadeler bulanık mantık aracılığıyla kolayca tanımlanabilmektedir (Siler ve Buckley, 2005). Temelini dilsel anlatım ve aralarındaki mantıksal bağlar oluşturmaktadır (Yılmaz ve Ayan, 2013). Lotfy A. Zadeh 1965 yılında bulanık mantık teoreminin temelini oluşturduğunu düşündüğü bulanık kümeleri 0 ile 1 aralığındaki üyelik derecelerine sahip nesnel kümesi olarak tanımladığı "Bulanık Mantık ve Bulanık Kümeler Kuramı" adlı makalesi bu konu ile ilgili yazılan ilk makale olarak literatüre geçmiştir. Bulanık mantık kavramı kullanıldığı ilk yıldan günümüze kadar önemi artarak belirsizliğin matematiksel bir yapısını ortaya çıkarmak ve karar vericilerin sözel değerlendirmelerini daha açık bir şekilde ifade edebilmeleri sağlamak için kullanılmıştır. Başka bir ifade ile bulanık mantık bir konu hakkındaki net çıkarımların yanında, esnek nitelikleri kullanarak gerçek dünyaya dönüştürme aracı olarak dikkate alınır (Amrita Sarkar, 2012). Bulanık mantık, klasik set teorisinin genelleştirilmesi olan bulanık kümelerin matematiksel teorisine dayanmakta ve Boolean mantığının bir uzantısıdır. Bir koşulun doğrulanmasında derece kavramını ortaya koyarak, bir koşulun doğru veya yanlış başka bir durumda olmasını mümkün kılarak mantık için çok değerli bir esneklik sağlar ve bu da yanlışlıklar ve belirsizlikleri hesaba katmayı mümkün kılar.

Bulanık mantığın temeli olan bulanık kümeleme kuramının birçok yöntemi vardır. Klasik kümeleme yönteminde üyelik derecesi 0 veya 1 değerlerini alabilmekte iken, bulanık kümeleme yönteminde üyelik dereceleri 0 ile 1 arasında değerlerden oluşmaktadır (Gülcan, 2012). Üyelik derecesi 1 olduğunda, bulanık tahminin hedefle tamamen uyumlu olduğu; üyelik derecesi 0 olduğunda, bulanık tahminlerin hedefe uymadığı anlaşılmaktadır. Bazı yazarlar, sürdürülebilirlik yönetim süreçlerini etkileyen öznel değerlendirmelerin belirsizliklerini etkin bir şekilde ele almak için bulanık sayıların kullanılabileceğini öne sürmüşlerdir.

## 2. LİTERATÜR TARAMASI

KOBİ'lerin sürdürülebilirlik performansının ölçülmesi ve açıklanması pratiği, paydaşlara karşı hesap verebilirliği sağlar ve şirketin daha sürdürülebilir bir şekilde operasyona geçişi yönetme konusunda şirkete destek verir. Bu nedenle sürdürülebilir olmalarını etkileyen temel faktörleri belirlemek son derece önemlidir (YaoWang, 2016). Üretim yapan küçük ve orta boy işletmelerin (KOBİ'ler) büyümesi, toplam imalat üretiminde önemli bir paya sahiptir. Küreselleşmenin bu çağında, KOBİ'ler aynı zamanda çevresel olarak bilinçli olmakla birlikte, maliyet tasarrufuyla sürdürülebilir kârlılığın sağlanması mecburiyetinde (Nevil S. Gandhi, Shashank J. Thanki, Jitesh J. Thakkar, 2018).

Schreyer (1996) KOBİ'lerin ekonomide oynadıkları anahtar rolü kabul ederek, pek çok hükümet, bilişim hizmetlerinden eğitime ve finansal desteğe kadar çeşitli yardım biçimlerini kasten yürüttüklerini belirtmiştir. Chrisman ve McMullan (2004) kamu kurumları tarafından desteklenen küçük işletmelerin desteklenmeyen işletmelere oranla hem daha fazla büyüyebildiklerini hem de sürdürülebilirlik açısından daha iyi konumda olabildiklerini tespit etmişlerdir.

Türkiye'de de Avrupa'da da genel olarak KOBİ'ler, KOBİ sahiplerinin ekonomik güçlerine göre yönetilmektedirler. Tabii burada KOBİ sahiplerinin ekonomik güçleri içerisinde kendi mal varlıklarının yanı sıra ailelerinden, arkadaşlarından, iş arkadaşlarından aldıkları krediler, paralar da

yer almaktadır. Bunların hepsinin yardımının dışında bile KOBİ'ler yine finansal anlamda sıkıntılar yaşamaktadırlar (Irwin, 1998, s.162). Bu bağlamda KOBİ'lerin desteklenmeleri onlar için kritik bir öneme sahiptir; çünkü KOBİ'ler genel manada maddi olarak çok büyük güçlere sahip değildirler. Maddi güçleri olmamasına rağmen KOBİ'ler büyük şirketlere göre üretim konusunda da hizmet konusunda da daha esneklerdir, problemlere karşı hızlı reaksiyon gösterirler, uzun hiyerarşi süreçleri olmadığı için hızlı çözümlerle sonuçlanan aksiyonlar alırlar ve en önemlisi de büyük şirketler gibi doyuma ulaşmadıkları için sürekli olarak inovasyona değer verirler (Ranga, Miedema ve Jorna, 2008). Del Brio ve Junquera'e(2003) göre KOBİ'lerin inovatif özelliklerinin fazla olduğu fakat bu özelliklerinin etkin hale gelmesi için kaynaklarının da yetersiz olduğunu dile getirenler arasında yer almışlardır. Sürdürülebilirlik tüm üretim şirketlerindeki üretim aktivitelerinin sonuçlarında ortaya çıkabilecek tehditler ve şirket sahiplerinin şirket yönetim biçimleri nedeniyle son yıllarda KOBİ'lerin hayatlarını ikame edebilmeleri için zorunlu bir hal almıştır ve bu öneminden dolayı gün geçtikçe herkes tarafından daha da fazla dikkate alınmaya başlanmıştır (Barrett, 1994).

KOBİ'lerin desteklenmesinin önemini yanında burada konu KOBİ'lerin hangi kriterlere göre desteklenmesi gerektiği olarak ortaya çıkmaktadır. Arena ve Azzone(2012) KOBİ'lerin desteklenmesi için belirlenecek olan kriterlerin üçüncü taraf uzmanlarca yapılmasını öne çıkarmışlardır. Bu metodolojiye benzer şekilde Munoz-Torres ve diğerleride (2012) KOBİ'lerin sürdürülebilirlik kriterlerini belirlemek için literatür araştırmalarının yapılmasını, Arena ve Azzone'nin dediği gibi uzmanlarla görüşülmesini ve hepsini birlikte değerlendirerek ortaya bir model çıkartılması gerektiğini öne sürmüştür. Küresel Raporlama İnisyatifi (GRI-Global Reporting Initiative) en bilinen sürdürülebilirlik etmenlerini belirlemiştir. GRI'ye (2013) göre bu sürdürülebilirlik kategorileri ekonomik, çevresel ve sosyal olarak üçe ayrılmıştır. Lee, Sameen ve Cowling(2015) de sürdürülebilirlik faktörlerini finansal ve finansal olmayan faktörler olarak ikiye ayırmışlardır. Aynı zamanda finansal zorlukların iş büyümesi ve sürdürülebilirliğin önündeki en büyük engel olduğunu da ifade etmişlerdir.

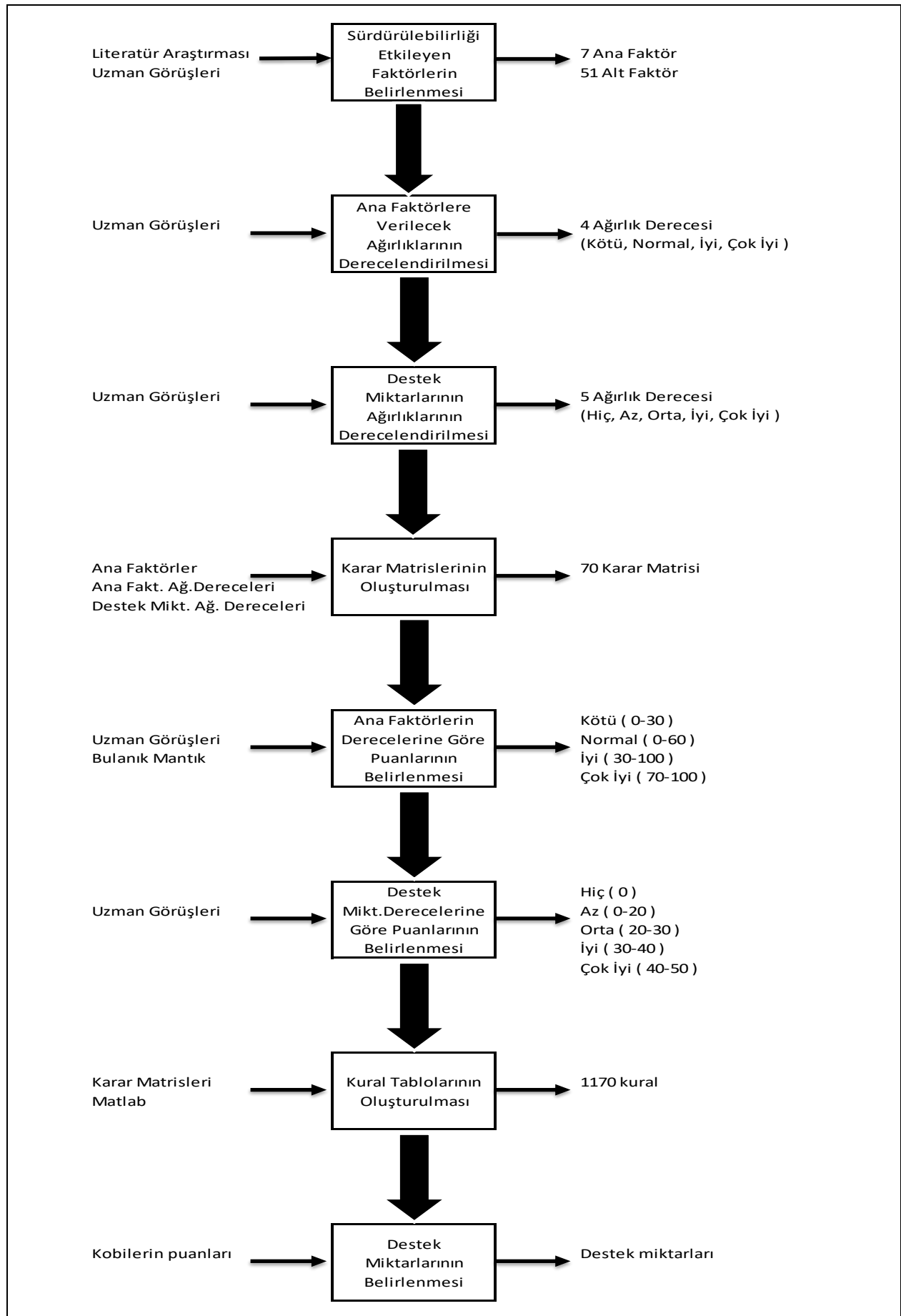
Bu çalışmanın amacı destek talep eden KOBİ'lerin sürdürülebilirlik performansının ölçülmesi ve sürdürülebilirlik kriterlerine göre verilmesi gereken destek miktarının belirlenmesidir. Bu amaç doğrultusunda yapılan literatür taraması sonucunda, dünyanın farklı bölgelerinde yer alan KOBİ'lerin sürdürülebilirlik kriterleri incelenerek Türkiye'deki KOBİ'lerin sürdürülebilirliğini etkileyen faktörler ortaya çıkarılmıştır. Bugüne kadar az sayıda çalışma KOBİ'lerde sürdürülebilirlik değerlendirilmesi konularında nicel yöntemlere başvurmuştur. Sürdürülebilirlik değerlendirmesini stratejik uygulamasına doğru nasıl yönlendireceğini gösteren bir boşluk vardır. Bazı yazarlar, sürdürülebilirlik yönetim süreçlerini etkileyen öznel değerlendirmelerin belirsizliklerini etkin bir şekilde ele almak için bulanık sayıların kullanılabilmesini öne sürmüşlerdir. Bu çalışmada da KOBİ'lerin sürdürülebilirliklerini etkileyen kriterler belirlenmiş ve nitekim, sürdürülebilirlik performansının hem nicel hem de nitel faktörler tarafından karakterize edilmesi nedeniyle Bulanık Mantık yöntemi ile destek miktarları belirlenmiştir. Çalışma sürdürülebilirlik raporunun etkinliği, doğruluğu, eksiksizliği açısından ve her boyuttaki veya sektördeki şirketlere çok yönlülük ve uygulanabilirlik açısından avantajlar sunmaktadır. Dolayısıyla KOBİ'lerin sürdürülebilirlik raporlamasında karşılaştıkları zaman ve mali kısıtlamaların üstesinden gelmede özellikle yararlı olacaktır.

### 3. METODOLOJİ

Bu çalışmada, KOBİ'lerin sürdürülebilirliğini etkileyen kriterler literatür araştırması sonucu derlenmiştir. Literatürden elde edilen birçok alt kriter, belirlenen yedi ana kriter altında incelenmiştir. Ana kriterler ele alınarak MATLAB (MATrix LABoratory) Fuzzy Logic (Bulanık Mantık) toolbox ekranında destek atamalarının gerçekleştirilmesi sağlanmıştır. MATLAB Fuzzy

Logic modülünde Mamdani veya Sugeno olmak üzere iki farklı bulanık mantık yöntemi mevcuttur. Her türlü problemin çözümünde kullanılabilen Mamdani yöntemi kullanılmıştır.

Çalışmada ilk olarak Türkiye'deki KOBİ'lerin sürdürülebilirlik performansının ölçülmesinde etkili faktörler belirlenmiştir ve bunlar yedi ana faktör altında birleştirilmiştir. Sürdürülebilirliğin değerlendirme sürecinde hem nitel hem nicel faktörler kullanılmıştır. Bu özellik, sürdürülebilirlik raporlaması amaçlarına uygundur; çünkü hem nicel hem de nitel bilgiler yoluyla performansın değerlendirilmesine izin vermektedir. Aşağıda **Şekil 1**' de çalışmanın metodolojisi yer almaktadır.



Şekil 1: Sürdürülebilirliği Etkileyen Kriterlerin Belirlenmesi ve Destek Miktarlarının Belirlenmesi Metodolojisi

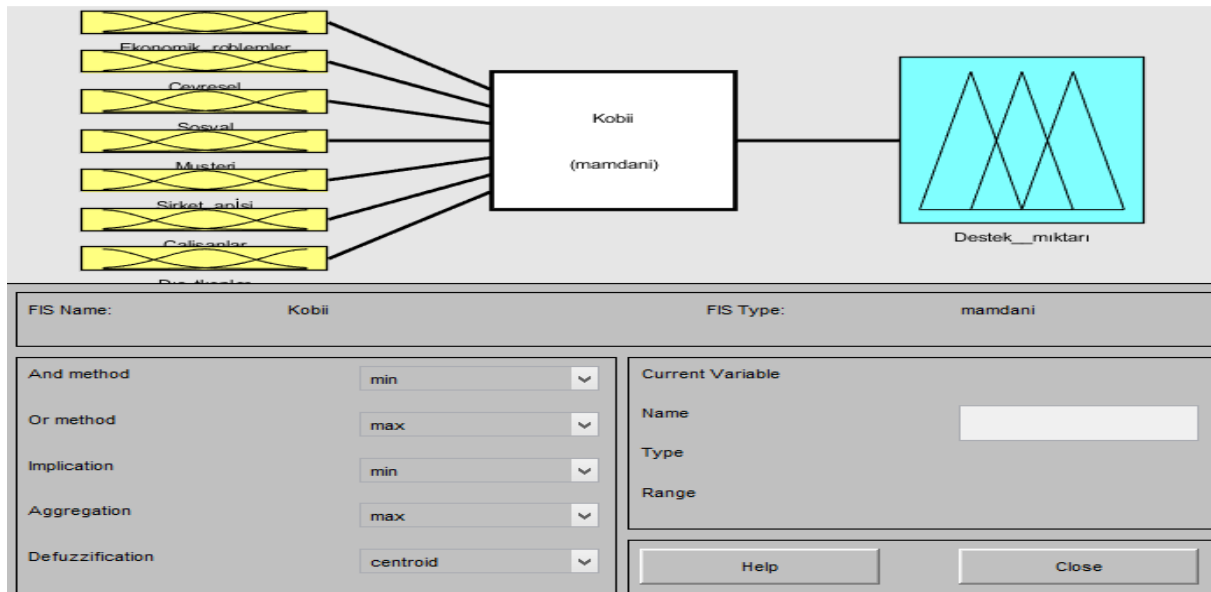
### 3.1.Sürdürülebilirlik Faktörlerinin Belirlenmesi

Sürdürülebilirlik faktörlerinin belirlenerek bunların sınıflandırılması ve bu sınıflandırmalara göre KOBİ destekleri atamalarının yapılması KOBİ'ler için çok kritik bir değerlendirme olup Küresel Raporlama İnişiyatifi'ne (GRI-Global Reporting Initiative) göre üçe ayrılan sürdürülebilirlik faktörleri ülkelerin yapılarına göre farklı şekillerde değerlendirilmektedirler. Bu çalışmada Türkiye'deki genel şirket yapıları dikkate alınarak sürdürülebilirlik ana faktörleri ekonomik, çevresel, sosyal, müşteri, şirket yapısı, çalışan ve dış etkenler olarak yediye ayrılmıştır ve Şekil2'de gösterilmiştir. MATLAB programında Fuzzy Logic modülünde 7 giriş ve 1 çıkışlı bir bulanık mantık denetleyicisi oluşturulmuştur, bu da Şekil 3'de gösterilmiştir.

Sürdürülebilirliği Etkileyen Faktörler	
Ana Kriterler	Alt Kriterler
<b>Ekonomik Faktörler</b>	Üretim maliyeti
	Dağıtım maliyetleri
	Tedarik maliyetleri
	Finansal kaynak eksikliği
<b>Çevresel Faktörler</b>	Doğal kaynakları veya çevre koruma prosedürlerini güçlendirmek
	Atık ve İsrar fazlalığı
	Üretimde daha az enerji tüketen ürünler geliştirme
	Geri dönüştürülebilir veya geri dönüştürülebilir atıklara sahip ürünler geliştirme
	İlgili çevre sertifikalarını edinme
<b>Sosyal</b>	Üretilen zararlı maddelerin miktarı
	İşçi / Yönetim İlişkileri
	İş Sağlığı ve Güvenliği
	Çeşitli ve Eşit Fırsat
	Kadınlar ve Erkekler için Eşit Ücret
	Ayrımcılık yapmama
<b>Müşteri İlişkileri</b>	Örgütlenme Özgürlüğü ve Toplu Pazarlık hakkı
	Pazarlama iletişimi
	Müşteri gizliliği
	Müşteri memnuniyeti
	Hizmet kalitesi
	Ürün kalitesi
	Zamanında teslimat
	Müşteri şikayetlerinin sayısının azaltılması
	Müşteri edinme oranının artırılması
	Müşteri sorunlarına cevap hızının artırılması
Müşteri devam oranının artırılması	
Müşterilerin güvenliğini ve sağlığına önem vermek	

<b>Şirket yapısı</b>	<b>Şirketin yönetim ve geliştirme bilgisi</b>
	Üretim süreçlerinin iyileştirilmesi
	Ürün ve hizmet geliştirme ve üretim süreçleri bilgisi
	Pazar bilgisi
	Sahiplerinin deneyimleri
	Firma Lokasyonu
	İşletmenin etkinliği
	İşletme faaliyet alanı
	Uluslararası pazar ve İç piyasa trendlerinden bilgi eksikliği
	Kurumsal inovasyon derecesi
	Mevcut teknolojileri hakkında bilgi
	Yeni ürün veya hizmet geliştirme sıklığı
Kurum içi bilgi paylaşımı	
<b>Çalışanlarla ilgili Faktörler</b>	Çalışanların nitelikleri
	Yetersiz eğitilmiş çalışan sayısının azaltılması
	Çalışan sayısının yeterli olması
	Çalışanların iş tatmin derecesinin artırılması
	Çalışan Eğitim ve Öğretimi
<b>Dış Çevre Faktörleri</b>	Altyapı eksikliği
	Yüksek düzeydeki bürokrasi ve idari engeller
	Uluslararası standartların yokluğu
	Dış danışmanlık kullanımı
	Karmaşık ihracat prosedürleri
	Vergi teşvikleri

Şekil 2: Sürdürülebilirliği Etkileyen Ana Faktörler ve Alt Faktörler



Şekil 3:7 Girişli ve 1 Çıkışlı Bulanık Mantık Denetleyicisi

### 3.2.Faktör ve Destek Atama Ağırlıklarının Belirlenmesi ve Karar Matrislerinin Oluşturulması

Bir uzman aracılığıyla her bir ana faktör birbiri ile ikili ve üçlü gruplar halinde sözel ifadeler ile kıyaslanarak destek miktarına atanacak sözel ifadeler belirlenmiştir ve **Şekil 4**'de gösterilmiştir. Ana faktörlerimizden oluşan 7 giriş “Kötü”, “Normal”, “İyi”, “Çok iyi” olarak 4 bulanık kümeye ayrılmıştır. Destek miktarını ifade eden çıkış ise “Hiç”, “Az”, “Orta”, “İyi”, “Çok iyi” olmak üzere 5 bulanık kümeye ayrılmıştır. Oluşturulmuş olunan karar matrislerinden birkaçı aşağıda verilmiştir:

		Çevresel Problemler			
Ekonomik Problemler		Kötü	Normal	İyi	Çok iyi
	Kötü	Hiç	Hiç	Hiç	Az
	Normal	Az	Orta	Orta	Orta
	İyi	Az	Orta	Orta	İyi
	Çok iyi	Orta	Orta	Çok iyi	Çok iyi

		Çalışanlar			
Müşteri ilişkileri		Kötü	Normal	İyi	Çok iyi
	Kötü	Hiç	Hiç	Az	Az
	Normal	Hiç	Az	Orta	İyi
	İyi	Hiç	Orta	Orta	İyi
	Çok iyi	Hiç	Orta	İyi	Çok iyi

		Dış Etkenler			
Sosyal Etmenler		Kötü	Normal	İyi	Çok iyi
	Kötü	Hiç	Az	Az	Az
	Normal	Az	Az	Orta	Orta
	İyi	Az	Orta	İyi	İyi
	Çok iyi	Az	Orta	İyi	İyi

		Çalışanlar			
Şirket yapısı		Kötü	Normal	İyi	Çok iyi
	Kötü	Hiç	Hiç	Hiç	Hiç
	Normal	Hiç	Az	Orta	İyi
	İyi	Hiç	Orta	Orta	Çok iyi
	Çok iyi	Hiç	İyi	Çok iyi	Çok iyi

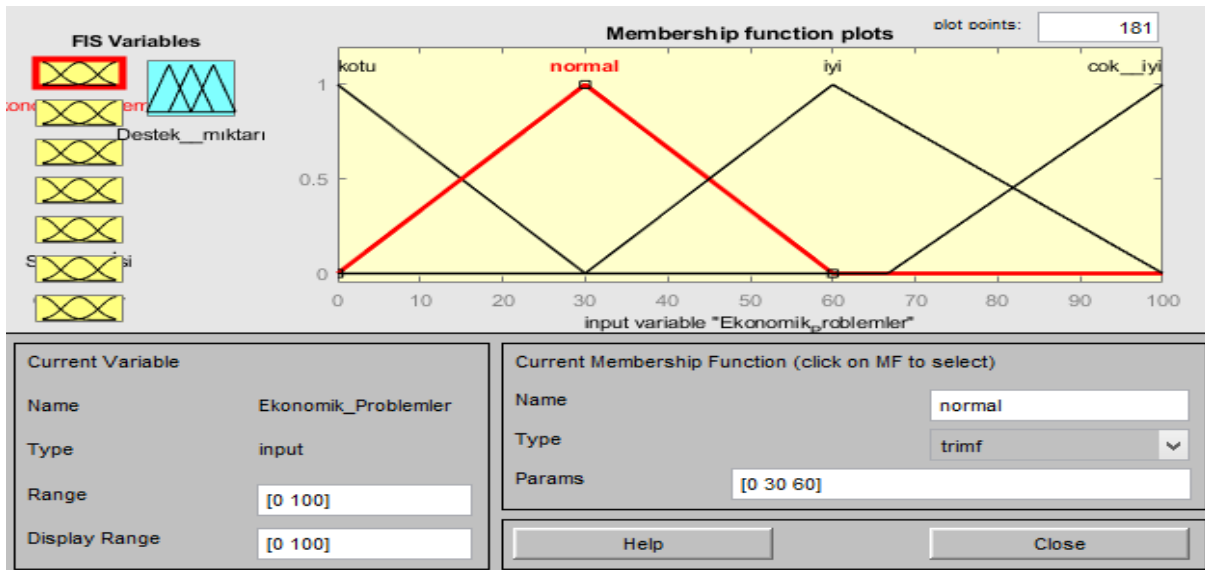
Çevresel	Müşteri	Çalışanlar	Ödeme	İyi	Kötü	Kötü	Hiç
Kötü	Kötü	Kötü	Hiç	İyi	Kötü	Kötü	Hiç
Kötü	Kötü	Normal	Hiç	İyi	Kötü	Normal	Hiç
Kötü	Kötü	İyi	Hiç	İyi	Kötü	İyi	Hiç
Kötü	Kötü	Çok iyi	Hiç	İyi	Kötü	Çok iyi	Hiç
Kötü	Normal	Kötü	Hiç	İyi	Normal	Kötü	Hiç
Kötü	Normal	Normal	Hiç	İyi	Normal	Normal	Orta
Kötü	Normal	Normal	Hiç	İyi	Normal	İyi	İyi
Kötü	Normal	İyi	Hiç	İyi	Normal	Çok iyi	İyi
Kötü	Normal	Çok iyi	Hiç	İyi	İyi	Kötü	Hiç
Kötü	İyi	Kötü	Hiç	İyi	İyi	Normal	İyi
Kötü	İyi	Normal	Hiç	İyi	İyi	İyi	Çok iyi
Kötü	İyi	İyi	Hiç	İyi	İyi	Çok iyi	Çok iyi
Kötü	İyi	Çok iyi	Hiç	İyi	Çok iyi	Kötü	Hiç
Kötü	Çok iyi	Kötü	Hiç	İyi	Çok iyi	Normal	Orta
Kötü	Çok iyi	Normal	Hiç	İyi	Çok iyi	İyi	İyi
Kötü	Çok iyi	İyi	Hiç	İyi	Çok iyi	Çok iyi	Çok iyi
Kötü	Çok iyi	Çok iyi	Hiç	Çok iyi	Kötü	Kötü	Hiç
Normal	Kötü	Kötü	Hiç	Çok iyi	Kötü	Normal	Hiç
Normal	Kötü	Normal	Hiç	Çok iyi	Kötü	İyi	Hiç
Normal	Kötü	İyi	Hiç	Çok iyi	Kötü	Çok iyi	Hiç
Normal	Kötü	Çok iyi	Hiç	Çok iyi	Normal	Kötü	Hiç
Normal	Normal	Kötü	Hiç	Çok iyi	Normal	Normal	Az
Normal	Normal	Normal	Hiç	Çok iyi	Normal	İyi	Orta
Normal	Normal	İyi	Az	Çok iyi	Normal	Çok iyi	Orta
Normal	Normal	Çok iyi	Orta	Çok iyi	İyi	Kötü	Hiç
Normal	İyi	Kötü	Hiç	Çok iyi	İyi	Normal	Az
Normal	İyi	Normal	Az	Çok iyi	İyi	İyi	İyi
Normal	İyi	İyi	Az	Çok iyi	İyi	Çok iyi	Çok iyi
Normal	İyi	Çok iyi	Orta	Çok iyi	Çok iyi	Kötü	Hiç
Normal	Çok iyi	Kötü	Hiç	Çok iyi	Çok iyi	Normal	Orta
Normal	Çok iyi	Normal	Orta	Çok iyi	Çok iyi	İyi	İyi
Normal	Çok iyi	İyi	Orta	Çok iyi	Çok iyi	Çok iyi	Çok iyi

Şekil 4: İkili Ve Üçlü Karar Matrisleri

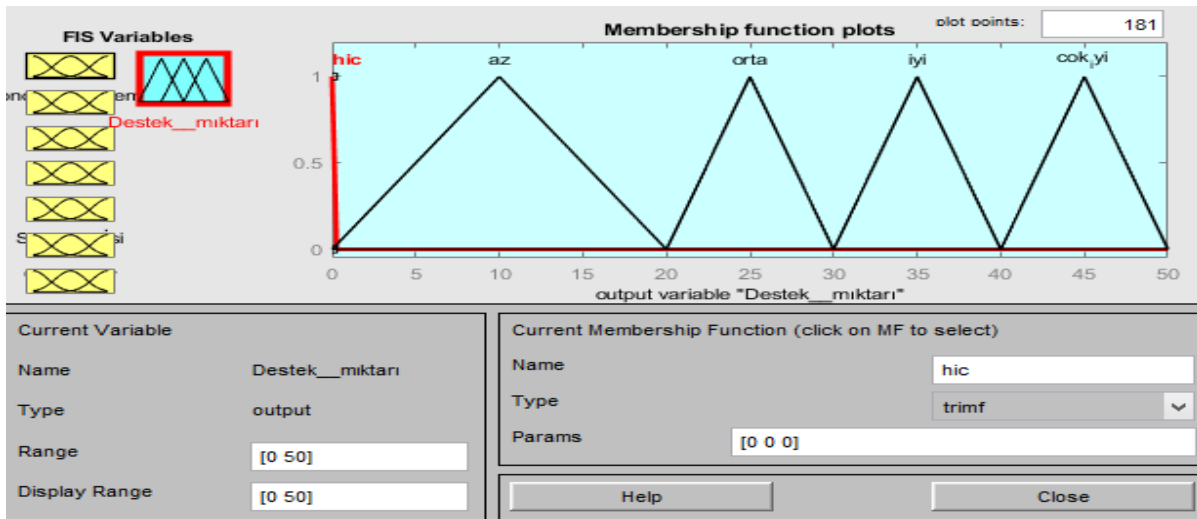
### 3.3.Faktör ve Destek Miktarlarının Puanlarının Belirlenmesi

Her bir ana faktör için üyelik fonksiyon editörü açılmış ve üçgenel 4 adet üyelik fonksiyonu girilmiştir. Ana faktörlerin kullanıcı tarafından 0 ile 100 arasında değerlendirilebilmesi için aralık [0 100] olarak ele alınmıştır. Üçgenel üyelik fonksiyonlarının eğimleri ayarlanmıştır. Her bir derecenin minimum noktası, tepe noktası ve maksimum nokta değerleri elde edilmiştir. MATLAB görüntüsü olarak Şekil 5'te gösterilmiştir. Sürdürülebilirlik değerlendirmesinden sonra KOBİ'lere verilmek istenen destek miktarları da 0 ila 50.000 TL arası olarak ele alınmıştır, bu da Şekil 6'da gösterilmiştir.





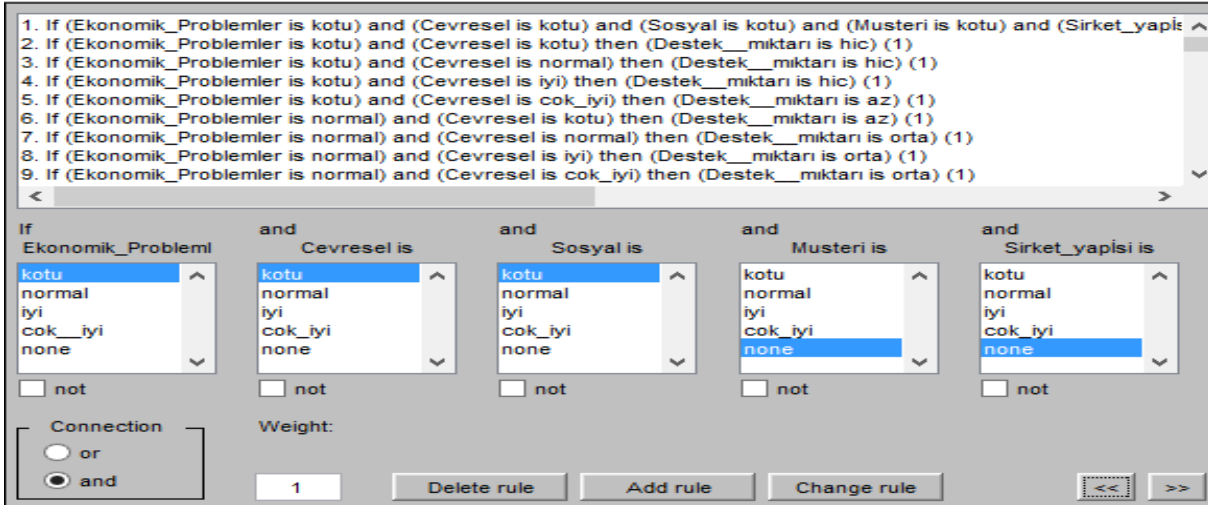
Şekil 5: Faktör Puanlamaları



Şekil 6: Destek Miktarlarının Puanlamaları

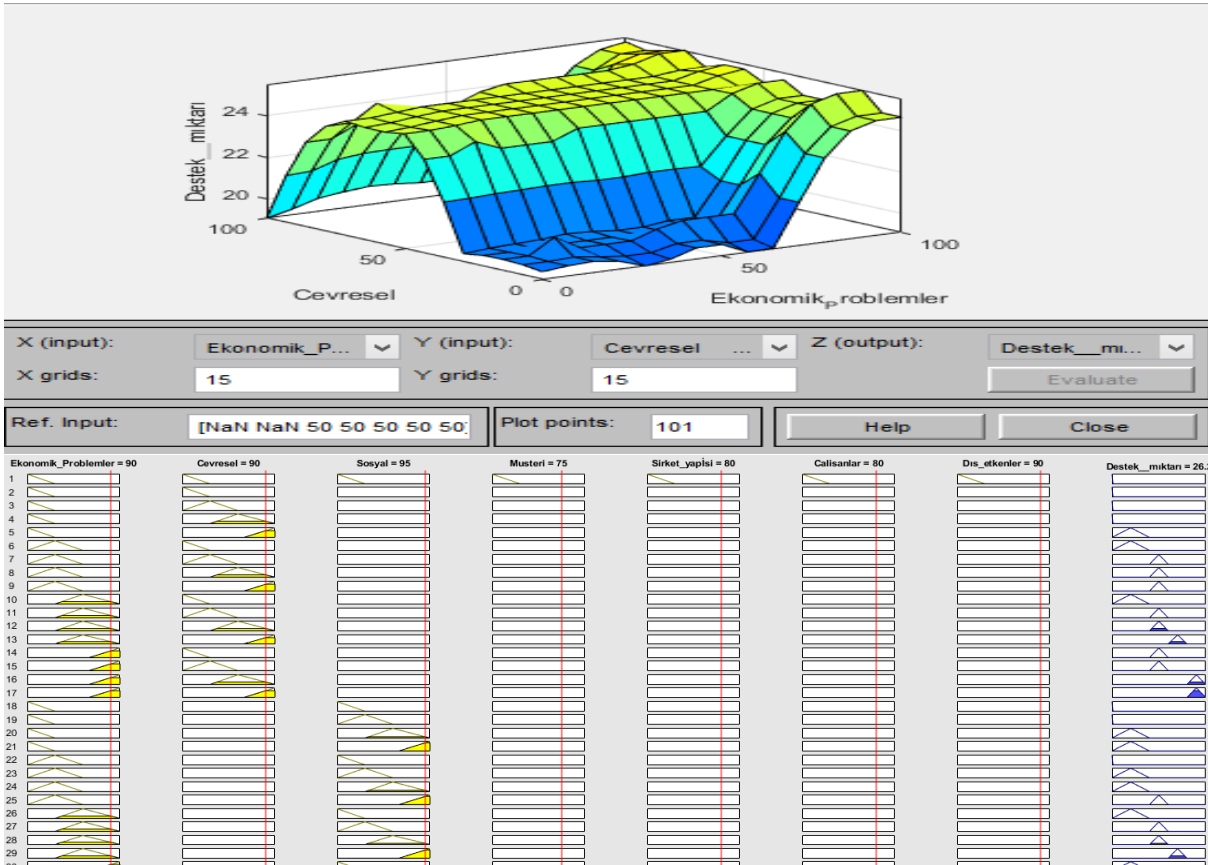
### 3.4.Kural Tablolarının Oluşturulması ve Kobi Puanlarına Göre Destek Miktarlarının Belirlenmesi

Kural tablolarından tüm kurallar MATLAB ortamına aktarılmıştır. Daha doğru sonuçlar elde etmek için toplamda 1170 kural oluşturulmuştur, Şekil 7’de gösterilmiştir.



Şekil 7: Kuralların MATLAB Gösterimi

Sonuç olarak ana faktörlere karşılık destek miktarı şekilsel gösterimleri elde edilmiştir. Aşağıda ekonomik ve çevresel ana faktörlerinden ve destek miktarlarından oluşan üç boyutlu gösterim yer almaktadır, Şekil 8’de gösterilmiştir.



Şekil 8: Destek Miktarının Üç Boyutlu Ve Faktör Bazında MATLAB Gösterimi

#### 4. SONUÇ

KOBİ'lere destek sağlamak ciddi bir sosyal ve profesyonel zorluktur. Bu çalışmanın amacı da Türkiye'deki KOBİ'lere verilebilecek destek miktarlarının planlaması aşamasında kullanılacak bir metot önererek karar verme sürecini hızlandırarak daha verimli sonuçlar elde edilmesini

sağlamaktır. Bu doğrultuda Türkiye'deki KOBİ'lerin sürdürülebilir olmalarını etkileyen tüm kriterler göz önüne alınmış ve sözel değere sahip bu kriterler bulanık mantık yöntemi ile sayısal değerlere dönüştürülmüştür. KOBİ'lerin sürdürülebilir olup olmadıklarına ekonomik, çevresel, sosyal, müşteri, şirket yapısı, çalışan ve dış etkenler ana kriterlerinden aldıkları puanlarla karar verilebilir ve destek miktarları da ana kriterlerin ağırlıklarına ve KOBİ'lerin değerlendirme sonucu aldıkları puanlara göre belirlenebilir hale gelmektedir. Böylelikle maddi destek talebinde bulunan hangi KOBİ'lere ne kadar destek verilmesi gerektiği kararı daha hızlı ve etkin alınmış olacaktır.

## KAYNAKÇA

- Aboelmaged M. (2017). "The drivers of sustainable manufacturing practices in Egyptian SMEs and their impact on competitive capabilities: A PLS-SEM model" *Journal of Cleaner Production* 175 207-221, [www.elsevier.com/locate/jclepro](http://www.elsevier.com/locate/jclepro)
- Amrina, E., Yusof, S.M. (2011). "Key performance indicators for sustainable manufacturing evaluation in automotive companies", *IEEE IEEM* 1093-1097.
- Arena, M., Azzone, G. (2012). "A process-based operational framework for sustainability reporting in SMEs", *J. Small Bus. Enterp. Dev.* 19 (4), 669e686.
- Audet, J., St-Jean, E. (2007). "Factors affecting the use of public support services by SME owners: Evidence from a periphery region of Canada", *Journal of Developmental Entrepreneurship*, 12(2), 165-180.
- Barrett, S. (1994). "Strategic environmental policy and international trade", *J. Public Econ.* 54 (3), 325e338.
- Berry, A. J., Sweeting, R., Goto, J. (2006). "The effect of business advisers on the performance of SMEs", *Journal of Small Business and Enterprise Development*, 13(1), 33-47. <http://dx.doi.org/10.1108/14626000610645298>
- Bryan, J. (2006). "Training and performance in small firms", *International Small Business Journal*, 24(6), 635-660. <http://dx.doi.org/10.1177/0266242606069270>
- Charter, M., Gray, C., Clark, T., Woolman, T. (2008). "Review: the Role of Business in Realising Sustainable Consumption and Production".
- Chirico, F. (2008). Knowledge accumulation in family firms. *International Small Business Journal*, 26(4), 433-462. <http://dx.doi.org/10.1177/0266242608091173>
- Chrisman, J. J., McMullan, E. W. (2004). "Outsider assistance as a knowledge resource for new venture survival", *Journal of Small Business Management*, 42(3), 229-244. <http://dx.doi.org/10.1111/j.1540-627X.2004.00109.x>
- Del Brío, J.A., Junquera, B. (2003). "A review of the literature on environmental innovation management in SMEs: implications for public policies", *Technovation* 23 (12), 939e948.
- Erol, I., Sencer, S., Sari, R. (2011) "A new fuzzy multi-criteria framework for measuring sustainability performance of a supply chain", *Ecol. Econ.* 70, 1088-1100.
- Govindan, K., Khodaverdi, R., Jafarian, A. (2013). "A fuzzy multi-criteria approach for measuring sustainability performance of a supplier based on triple bottom line approach", *J. Clean. Prod.* 47, 345-354.
- Govindana, K., Garg, K., Gupta, S., Jha, P.C. (2016). "Effect of product recovery and sustainability enhancing indicators on the location selection of manufacturing facility", *Ecol. Indic.* 67, 517-532.
- GRI-Global Reporting Initiative. (2013). G4, Part2, Implementation Manual available at: [www.globalreporting.org](http://www.globalreporting.org).
- Gülcan B. (2012). "Bulanık Doğrusal Programlama ve Bir Bisküvi İşletmesinde Optimum Ürün Formülü Oluşturma", Yüksek Lisans Tezi, Karamanoğlu Mehmetbey Üniversitesi Sosyal Bilimler Enstitüsü, Karaman.

- Haskel, J.E., Pereira, S.C., Slaughter, M.J. (2007). “Does inward foreign direct investment boost the productivity of domestic firms?”, *Rev. Econ. Stat.* 89 (3), 482-496.
- Henry, C., Hill, M. F., Leitch, M. C. (2004). “The effectiveness of training for new business creation: A longitudinal study”, *International Small Business Journal*, 22(3), 249–271. <http://dx.doi.org/10.1177/0266242604042378>
- Irwin, D. (1998) “Common reasons why businesses fail”, In *Building your business pocketbook* (pp. 159–169), Thorogood Publishing Ltd.
- Jenkins, H. (2009). “A ‘business opportunity’ model of corporate social responsibility for small and medium sized enterprises”, *Bus. ethics A Eur. Rev.* 18, 21-36.
- Jenkins, H., Yakovleva, N. (2006). “Corporate social responsibility in the mining industry: exploring trends in social and environmental disclosure”, *J. Clean. Prod.* 14 (3), 271-284.
- Journeault, M. (2016). “The integrated scorecard in support of corporate sustainability strategies”, *J. Environ. Manag.* 182, 214-229.
- Kaplan, R.S., Norton, D.P. (1996)ç “The Balanced Scorecard. Translating Strategy into Action”, Harvard Business School Press, Boston, Mass.
- Klir, George J., Yuan, Bo. (1995) “Fuzzy Sets and Fuzzy Logic: Theory and Applications”, Paperback, Prentice Hall.
- Lee, N., Sameen, H., Cowling, M. (2015) “Access to finance for innovative SMEs since the financial crisis”, *Research Policy*, 44(2), 370–380. <http://dx.doi.org/10.1016/j.respol.2014.09.008>
- Macdonald, S., Assimakopoulos, D., Anderson, P. (2007). “Education and training for innovation in SMEs: A tale of exploitation”, *International Small Business Journal*, 25(1), 77–95. <http://dx.doi.org/10.1177/0266242607071782>
- Mason, C., Brown, R. (2013). “Creating a good public policy to support high-growth firms”, *Small Business Economics*, 40(2), 211–225. <http://dx.doi.org/10.1007/s11187-011-9369-9>
- Muller, P., Gagliardi, D., Caliandro, C., Unlu Bohn, N., Klitou, K. (2014) “Annual report on European SMEs 2013/2014 – A partial and fragile recovery”, The European Commission, Directorate General for Enterprise and Industry.
- Munoz-Torres, M.J., Escrig-Olmedo, E., Ferrero-Ferrero, I., Fernandez Izquierdo, M.A., Leon-Soriano, R., Rivera-Lirio, J.M. (2012). “Materiality analysis for CSR reporting in Spanish SMEs”, *Int. J. Manag. Knowl. Learn.* (2), 231e250.
- Narula, R. (2004) “R&D collaboration by SMEs: new opportunities and limitations in the face of globalization”, *Technovation* 24 (2), 153-161.
- Nevil S. Gandhi, Shashank J. Thanki, Jitesh J. Thakkar. (2018). “Ranking of drivers for integrated lean-green manufacturing for Indian manufacturing SMEs” *Journal of Cleaner Production* 171, 675-689. [www.elsevier.com/locate/jclepro](http://www.elsevier.com/locate/jclepro)
- Patzelt, H., Shepherd, D. (2009). “Strategic entrepreneurship at universities: Academic entrepreneurs’ assessment of policy programs”, *Entrepreneurship Theory and Practice*, 33(1), 319–340. <http://dx.doi.org/10.1111/j.1540-6520.2008.00291.x>
- Ranga, L. M., Miedema, J., Jorna, R. (2008) “Enhancing the innovative capacity of small firms through triple helix interactions: Challenges and opportunities”, *Technology Analysis and Strategic Management*, 20(6), 697–716. <http://dx.doi.org/10.1080/09537320802426408>
- Schreyer, P. (1996) “SMEs and Employment Creation: Overview of Selected Quantitative Studies in OECD Member Countries, STI Working Papers 1996/4 ed. Directorate for Science”, Technology and Industry, OECD, Paris
- Schwartz, D., Bar-El, R. (2004). “Targeted consultancy services as an instrument for the development of remote SMEs – A Brazilian case”, *International Small Business Journal*, 22(5), 503–521. <http://dx.doi.org/10.1177/0266242604046298>
- Seuring, S., Müller, M. (2008) “From a literature review to a conceptual framework for sustainable supply chain management”, *J. Clean. Prod.* 16 (15), 1699-1710.

- Siler, W., James J.B. (2005). “Fuzzy Expert Systems and Fuzzy Reasoning”, John Wiley Pub. Comp., New Jersey.
- Spicer, D., Sadler-Smith, E. (2006). “Organizational learning in smaller manufacturing firms”, *International Small Business Journal*, 24(2), 133–158.  
<http://dx.doi.org/10.1177/0266242606061836>
- Tomsic, N.,Bojnec, S., Simcic, B. (2015). “Corporate sustainability and economic performance in small and medium-sized enterprises”, *Journal of Cleaner Production*, 108, 603-612
- Tseng, M.L. (2013). “Modeling sustainable production indicators with linguistic preferences”, *J. Clean. Prod.* 40, 46-56.
- Tseng, M.L., Divinagracia, L., Divinagracia, R. (2009). “Evaluating firm’s sustainable production indicators in uncertainty”, *Comput. Ind. Eng.* 57 (4), 1393-1403.
- Wiggins, S., Kirsten, J., Llambi, L. (2010) “The future of small farms”, *World Development*, 38(10), 1341–1348. <http://dx.doi.org/10.1016/j.worlddev.2009.06.013>
- World Commission on Environment. (1987). “El desarrollo sostenible, una guía sobre nuestro futuro común: El informe de la Comisión Mundial sobre el Medio Ambiente y el Desarrollo. Peterson’s”.
- Yao Wang. (2016). “What are the biggest obstacles to the growth of SMEs in developing countries? An empirical evidence from an enterprise survey”, *Borsa Istanbul Review* 16-3 167-176.  
<http://www.elsevier.com/journals/borsa-istanbul-review/2214-8450>
- Yılmaz, A., Ayan, K. (2013) “Cancer Risk Analysis by Fuzzy Logic Approach and Performance Status of The Model”, *Turkish Journal of Electrical Engineering & Computer Sciences*, TÜBİTAK, Vol.21, pp.897-912.
- Zadeh, L.A. (1965). “Fuzzy Sets”, *Information and Control*, 8(3):338–353

## DOĞAL DİL İŞLEMEDE DERİN ÖĞRENME UYGULAMALARI ÜZERİNE BİR LİTERATÜR ÇALIŞMASI

\*\*\*

### A LITERATURE STUDY ON DEEP LEARNING APPLICATIONS IN NATURAL LANGUAGE PROCESSING

Doğan KÜÇÜK\*  
Nursal ARICI\*\*

#### Öz

*Derin öğrenme, yapay zekâ ve makine öğrenmesi alanlarının önemli ve güncel bir konusu haline gelmiştir. Özellikle son yıllarda, farklı derin öğrenme yöntemleri öneren çalışmaların ve mevcut yöntemleri değişik problemler üzerinde uygulayan çalışmaların sayıları hızla artmaktadır. Doğal dil işlemenin çeşitli alt alanlarında da bu yöntemler yaygın olarak kullanılmış ve halen kullanılmaktadır. Bu derleme çalışmasında, ilk olarak derin öğrenme yöntemlerinin bir sınıflandırması sunulmuş, ardından da doğal dil işleme problemlerine derin öğrenme yaklaşımlarının sunulduğu önemli çalışmalar incelenmiştir. Derin öğrenme ve doğal dil işleme problemlerinin çözümü amacıyla derin öğrenme konularıyla ilgili hem teorik çalışmaların hem de pratik uygulamalar içeren çalışmaların sayısının ve yaygınlığının daha da artacağı öngörülmektedir. Bu nedenle çalışmamızın; doğal dil işleme alanında derin öğrenme uygulamaları konusunda önemli bir Türkçe kaynak olacağı düşünülmektedir.*

**Anahtar Kelimeler:** Doğal Dil İşleme, Metin Madenciliği, Derin Öğrenme, Makine Öğrenmesi.

#### Abstract

*Deep learning is an important and recent topic of artificial intelligence and machine learning areas. Especially in recent years, the number of studies proposing different deep learning methods and applying these methods on different problems is increasing. These methods have also been used at various subareas of natural language processing extensively, and are still being used. In this survey paper, firstly, classification of deep learning techniques is presented and then important studies about deep learning approaches for natural language processing problems are discussed. It is expected that the number and prevalence of both theoretical studies and studies with practical applications on deep learning and on deep learning solutions to natural language processing problems are going to increase. Therefore it is considered that our study will be an important Turkish resource on the topic of deep learning applications for natural language processing.*

**Keywords:** Natural Language Processing, Text Mining, Deep Learning, Machine Learning.

\* Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, [dogan.kucuk@gazi.edu.tr](mailto:dogan.kucuk@gazi.edu.tr)

ORCID: 0000-0001-5265-3263

\*\* Doç. Dr. Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, [nursal@gazi.edu.tr](mailto:nursal@gazi.edu.tr)

ORCID: 0000-0002-4505-1341

## 1. GİRİŞ

Derin öğrenme, günümüzde oldukça popüler olan yapay zekâ ve makine öğrenmesi alanlarının önemli güncel bir alt alanıdır. Derin öğrenme metotları, genel olarak yapay sinir ağları (YSA - artificial neural networks) çalışmaları üzerine geliştirilmişlerdir. Ancak bu çalışmalardan farklı olarak daha fazla sayıda gizli nöron ve katman üzerine kuruludurlar (Ravi vd., 2017).

Derin öğrenme, eldeki verinin birden çok soyutlama seviyesinde temsil edilebilmesi için birden çok işlem katmanı bulunan hesaplama modellerini kapsamaktadır (Bengio, 2009; Lecun vd., 2015). Derin ağlar olarak da bilinen derin öğrenme yöntemlerinde, verinin temsili için üst üste olacak şekilde farklı katmanlar mevcuttur. Derin öğrenme yöntemleri ham verilerden etkin bir üst seviye soyutlaması yapmakla, otomatik öznitelik kümeleri oluşturabilmekte, böylelikle normalde çoğunlukla insanlar tarafından belirlenen özniteliklerin otomatik olarak çıkarılıp kullanılması sağlanmaktadır (Ravi vd., 2017).

Bazı derin öğrenme algoritmalarının öğrenme aşaması oldukça uzun sürebilmektedir ve farklı çalışmalarda, derin sinir ağlarının öğrenme süreçlerini kısaltmak için yarı-denetimli öğrenme yaklaşımları önerilmiştir (Lee, 2013). Bazı çalışmalarda ise destek vektör makinaları (support vector machines) gibi derin olmayan ama yaygın kullanılan makine öğrenmesi yöntemlerine derin öğrenmenin yeteneklerini kazandırma yönünde yaklaşımlar sunulmuştur (Cho ve Saul, 2009).

Derin öğrenme yöntemleri; video, ses, metin gibi birçok veri türünün işlenmesinde oldukça başarılı sonuçlar vermiştir (LeCun vd., 2015). Çoğunlukla bir yöntem, doğal dil işleme gibi metin verisi işleme konularında başarılı olabilirken (Collobert ve Weston, 2008), farklı bir yöntem video ve ses verilerinin işlenmesinde daha başarılı sonuçlar verebilmektedir. Bazı çalışmalarda ise, önerilen derin öğrenme yaklaşımları çok modlu (metin, görüntü ve ses gibi farklı modları aynı anda) öğrenmede başarıyla kullanılabilir (Ngiam ve vd., 2011).

Derin öğrenme yöntemlerinin daha ayrıntılı olarak verilebilecek uygulama alanları arasında aşağıdakiler yer almaktadır (Deng ve Yu, 2014):

1. Dil modelleme ve doğal dil işleme (language modeling and natural language processing)
2. Konuşma ve ses işleme (speech and audio processing)
3. Bilgi erişimi (information retrieval)
4. Nesne tanıma ve bilgisayarlı görü (object recognition and computer vision)
5. Çok modlu ve çok görevli öğrenme (multimodal and multitask learning)

Yukarıdaki listede ilk madde olarak yer alan doğal dil işleme konusu, yapay zekâ alanı içerisinde değerlendirilmektedir ve metin madenciliği olarak da adlandırılmaktadır. Doğal dil işleme, doğal dilde oluşturulmuş metinler üzerinde otomatik olarak gerçekleştirilen metin ayrıştırma, metin sınıflandırma, bilgi çıkarımı, duygu analizi gibi birçok önemli konuyu bünyesinde barındırmaktadır ve uygulama alanları da gün geçtikçe artmaktadır.

Bu çalışmamızda; doğal dil işleme problemlerinin çözümü için derin öğrenme yöntemlerini kullanan çalışmalar incelenmiş olup Türkçe bir kaynak olarak araştırmacıların dikkatine sunulmuştur. Çalışmanın kalan bölümleri şu şekilde ilerlemektedir: 2. bölümde derin öğrenme yöntemleri hakkında genel bilgiler verilmiş; 3. bölümde doğal dil işleme alanında yapılmış derin öğrenme çalışmaları incelenmiş, 4. bölümde ise makale özetlenmiş ve sonuçlar verilmiştir.

## 2. DERİN ÖĞRENME YÖNTEMLERİ

Bu bölümde aşağıda listelenen altı derin öğrenme yöntemi (Ravi vd., 2017) incelenmiş ve bunların uygulama alanları hakkında bilgiler verilmiştir:

1. Derin Sinir Ağları (Deep Neural Networks - DNN)
2. Derin Oto-kodlayıcılar (Deep Autoencoders)
3. Derin İnanç Ağları (Deep Belief Networks - DBN)
4. Derin Boltzmann Makinesi (Deep Boltzmann Machine - DBM)
5. Yinelenen Sinir Ağları (Recurrent Neural Networks - RNN)
6. Evrişimsel Sinir Ağları (Convolutional Neural Networks - CNN)

Aşağıda Tablo 1’de bu yöntemlerin niteliksel bir karşılaştırılması (Ravi vd., 2017) çalışmasından yararlanılarak sunulmuştur. Çok katmanlı yapılarından dolayı söz konusu derin öğrenme algoritmalarının tümü için öğrenme süreci hesaplama anlamında maliyetlidir. Algoritmaların farklı alanlarda uygulanmaları sırasında, uygulamanın gereksinimleri, yapılan parametre ayarlamaları ve kullanılan veri kümesinin boyutuna göre performansları farklılık gösterebilmektedir. Dolayısıyla, yöntemlerin karşılaştırması Tablo 1’de sadece niteliksel anlamda sunulmuştur.

**Tablo 1. Tanıtılan Derin Öğrenme Yöntemlerinin Niteliksel Karşılaştırılması**

<b>Derin Öğrenme Yöntemi</b>	<b>Özellikleri</b>
<b>Derin Sinir Ağları</b>	Birçok alanda başarıyla uygulanmaktadır. Ancak öğrenme süreci uygulama alanına bağlı olarak oldukça yavaş olabilmektedir.
<b>Derin Oto-kodlayıcılar</b>	Öğrenme için işaretli veri gerektirmeyen denetimsiz bir derin öğrenme yöntemidir.
<b>Derin İnanç Ağları</b>	Denetimli ve denetimsiz şekilde kullanılabilir. Ancak eğitim süreci oldukça yavaş olabilmektedir.
<b>Derin Boltzmann Makinesi</b>	Derin İnanç Ağları’ndan farkı, sadece en üst iki katman dışında tüm katmanlar arasında da yönsüz bağlantıların olmasıdır. Bu nedenle de Derin İnanç Ağları’na göre hesaplama anlamında daha maliyetlidir.
<b>Yinelenen Sinir Ağları</b>	Sıralı desenleri belirlemede başarılıdır. Bu nedenle özellikle Long Short-Term Memory (LSTM) gibi alt türleri doğal dil işleme gibi alanlarda başarılı sonuçlar vermektedir.
<b>Evrişimsel Sinir Ağları</b>	Genellikle bilgisayarlı görü alanında ve görsel veriler üzerinde kullanılmakta ve büyük boyutta işaretli veri gerektirmesine rağmen başarılı sonuçlar vermektedir. Doğal dil işleme alanındaki problemlere de başarıyla uygulanmıştır.

### 2.1. Derin Sinir Ağları

Literatürde, sinir ağları konusundaki ilk algoritmalarından biri perseptron algoritmasıdır (Rosenblatt, 1958). Bu ağda, bir girdi katmanı mevcuttur ve direk olarak çıktıya bağlıdır. Bu algoritma ile lineer olarak ayrılabilen desenlerin ayrıştırılması amacıyla sınıflandırıcılar geliştirilebilir (Ravi vd., 2017). Daha karmaşık problemler için bu algoritmaya birden fazla gizli katman eklenmiş ve delta kuralı adı verilen öğrenme yöntemiyle her katmanın ağırlığı ayarlanabilmiştir.



Bu tipteki sinir ağlarına daha fazla sayıda gizli katman (ikiden fazla) eklenmesiyle doğrusal olmayan karmaşık ilişkilerin de tespit edilebilmesi sağlanmıştır ve elde edilen bu sinir ağlarına derin sinir ağları (DSA) adı verilmiştir. DSA'lar hem (işaretili veriler varsa) denetimli, hem de kümeleme gibi denetimsiz öğrenme problemleri için kullanılabilir. DSA'lar yaygın olarak sınıflandırma ve regresyon amacıyla kullanılmakta ve başarılı sonuçlar elde edilmesini sağlamaktadırlar. Ancak öğrenme süreçleri oldukça yavaş olabilmektedir (Ravi vd., 2017).

## 2.2. Derin Oto-Kodlayıcılar

Oto-kodlayıcılar, problem çözümü için gerekli öznelik kümesinin veriden otomatik olarak çıkarılması amacıyla ortaya atılmış sinir ağlarıdır. Bir oto-kodlayıcı, girdi vektörüne bir sınıf etiketi vermek yerine onu yeniden oluşturmak için öğretilmektedir (Ravi vd., 2017).

Derin öğrenmenin konusu olan derin oto-kodlayıcılar ise, çok boyutlu veriyi temsil etmek için birden fazla oto-kodlayıcının birbiri üstüne kümelenmesiyle oluşan mimarilerdir (Hinton ve Salakhutdinov, 2006). Literatürde, derin oto-kodlayıcıların birçok farklı türü önerilmiştir.

Derin oto-kodlayıcıların amacı, öznelik kümesini otomatik çıkarmak veya veri boyutu sayısını azaltmaktır (Hinton ve Salakhutdinov, 2006). Bu yöntemde öğrenme için işaretili veri kümesine ihtiyaç yoktur, yani denetimsiz bir öğrenme yöntemidir. Ancak yöntem uygun ağırlıkları bulabilmek için bir ön-öğrenme aşamasına ihtiyaç duymaktadır. Bu ön-öğrenme aşamasında son konfigürasyona uygun yaklaşık ağırlıklar yönteme sağlanmaktadır (Ravi vd., 2017).

## 2.3. Derin İnanç Ağları

Derin inanç ağları (Hinton vd., 2006) ve bir sonraki bölümde tanıtılacak olan derin Boltzmann makineleri, sınırlandırılmış Boltzmann makinesi adı verilen bir algoritmaya (Hinton ve Sejnowski, 1986) dayanmaktadır. SBM algoritması stokastik bir sinir ağı olarak tanımlanmakta ve bu ağlarda Gaussian gibi belirli bir dağılıma sahip stokastik birimler kullanılmaktadır. Öğrenme sürecinde ise Gibbs örnekleme gibi ağırlıkları adım adım ayarlayan yöntemler kullanılmaktadır.

Derin inanç ağları birden fazla SBM'nin bileşkesi olarak algılanabilir (Hinton vd., 2006). Her bir SBM'nin gizli katmanı, bir sonraki SBM'nin görünür katmanına bağlanmıştır ve en üst seviyede yönsüz bağlantılar vardır. Derin inanç ağları hem denetimli hem de denetimsiz öğrenme amacıyla kullanılabilir. Bu ağı başlatmak için katman-katman bir açgözlü algoritma ile öğrenme gerçekleştirilmektedir (Ravi vd., 2017).

## 2.4. Derin Boltzmann Makinesi

Boltzmann makinelerine dayalı bir diğer derin öğrenme yöntemi de derin Boltzmann makineleridir (Salakhutdinov ve Larochelle, 2010). Derin inanç ağlarından farkı; derin Boltzmann makinelerinde ağın tüm katmanları arasında yönsüz bağlantıların olmasıdır. Bir diğer fark ise derin Boltzmann makinelerinde zaman karmaşıklığının daha fazla olması bu nedenle de büyük veri kümelerinde öğrenme sürecinin yavaş olmasıdır (Ravi vd., 2017).

## 2.5. Yinelenen Sinir Ağları

Yinelenen sinir ağları (Williams ve Zipser, 1989), veri akışlarını (stream of data) analiz edebilen gizli katmanlara sahip sinir ağlarıdır ve çıktının bir önceki hesaplamalara bağlı olduğu problemlerin çözümü için çok uygundur (Ravi vd., 2017). Yinelenen sinir ağları bu nedenle özellikle doğal dil işlemenin değişik problemlerinin çözümünde oldukça başarılı olmuştur (Ravi vd., 2017). Yinelenen sinir ağlarında öğrenme sırasında ortaya çıkan bazı problemler nedeniyle, bu ağların uzun kısa-dönem bellek (Long Short-Term Memory - LSTM) (Hochreiter ve Schmidhuber, 1997) gibi farklı sürümleri literatüre girmiştir.

## 2.6. Evrişimsel Sinir Ağları

Evrişimsel sinir ağları, çok boyutlu girdiler için ve özellikle iki boyutlu görsel veriler için önerilmiş bir derin öğrenme yöntemidir (LeCun vd., 1998). Evrişimsel sinir ağları diğer sinir ağlarına göre daha az sayıda nöron bağlantısına sahiptir ve birçok farklı sürümü literatürde mevcuttur. Bu sinir ağları öğrenme aşamasında oldukça büyük boyutta işaretli veriye ihtiyaç duymaktadır (Ravi vd., 2017). Bu derin öğrenme yöntemi doğal dil işleme alanındaki çeşitli problemler üzerinde uygulanmış ve başarılı sonuçlar elde edilmiştir.

## 3. DOĞAL DİL İŞLEME ALANINDAKİ DERİN ÖĞRENME ÇALIŞMALARI

Derin öğrenmenin bir önceki bölümde anlatılan farklı yöntemlerinden özellikle yinelenen sinir ağları ve evrişimsel sinir ağlarının doğal dil işleme problemlerini çözmede de başarılı sonuçlar verdiği görülmüştür. Örneğin; evrişimsel sinir ağlarının doğal dil işlemenin çok farklı problemlerini çözmede kullanılabileceği, çok görevli öğrenme için birleşik bir mimarinin tanıtıldığı bir çalışmada gösterilmiştir (Collobert ve Weston, 2008). Bahsi geçen bu iki yöntemle birlikte; diğer derin öğrenme yöntemlerinin de doğal dil işlemenin çeşitli problemlerinde kullanılabileceği ve bu yöntemlere dayalı başarılı sistemlerin geliştirilebileceği, literatürdeki farklı çalışmalarda ifade edilmiştir (Socher vd., 2012).

Bu bölümün aşağıdaki alt bölümlerinde de doğal dil işleme alanındaki örnek problemler ve bu problemler için derin öğrenme yaklaşımlarının tanıtıldığı çalışmalara yer verilmiştir. Ayrıca, Tablo 2’de bu alt bölümlerde açıklanan problemler ve çözümünde kullanılan derin öğrenme yöntemleri özet şeklinde sunulmuştur.

**Tablo 2. Tanıtılan Doğal Dil İşleme Problemleri ve Kullanılan Derin Öğrenme Yöntemleri**

<i>Doğal Dil İşleme Problemi</i>	<i>Kullanılan Derin Öğrenme Yöntemleri</i>
<b>Metin Sınıflandırma</b>	<ul style="list-style-type: none"><li>• Evrişimsel Sinir Ağları</li><li>• Yinelemeli Evrişimsel Sinir Ağları</li><li>• LSTM ve Evrişimsel Sinir Ağları</li></ul>
<b>Metin Ayrıştırma</b>	<ul style="list-style-type: none"><li>• Evrişimsel Sinir Ağları</li></ul>
<b>Duygu Analizi</b>	<ul style="list-style-type: none"><li>• Derin Oto-Kodlayıcılar</li><li>• Evrişimsel Sinir Ağları</li></ul>
<b>Bilgi Çıkarımı</b>	<ul style="list-style-type: none"><li>• Derin Sinir Ağları</li></ul>
<b>Varlık İsmi Tanıma</b>	<ul style="list-style-type: none"><li>• LSTM ve Evrişimsel Sinir Ağları</li></ul>
<b>Zamansal İlişki Çıkarımı</b>	<ul style="list-style-type: none"><li>• Evrişimsel Sinir Ağları</li></ul>
<b>Olay Çıkarımı</b>	<ul style="list-style-type: none"><li>• Evrişimsel Sinir Ağları</li></ul>
<b>Sözcük Türü Etiketleme</b>	<ul style="list-style-type: none"><li>• Derin Sinir Ağları</li><li>• LSTM</li></ul>
<b>Metin Sıralama</b>	<ul style="list-style-type: none"><li>• Evrişimsel Sinir Ağları</li></ul>
<b>Otomatik Harf Çevirisi</b>	<ul style="list-style-type: none"><li>• Derin İnanç Ağları</li></ul>
<b>Otomatik Soru Cevaplama</b>	<ul style="list-style-type: none"><li>• Evrişimsel Sinir Ağları</li><li>• LSTM</li></ul>

### 3.1. Metin Sınıflandırma

Doğal dil işlemenin uzun zamandır çalışılan ve önemli uygulama alanları olan konularından bir tanesi metin sınıflandırmadır. Derin öğrenme yöntemleri, metin sınıflandırma amacıyla da kullanılmaktadır ve bu konuda çeşitli güncel çalışma mevcuttur.

Örnek bir çalışmada, değişik veri kümeleri üzerinde cümle sınıflandırma problemi (olumlu/olumsuz müşteri değerlendirmeleri gibi) için basit tek seviyeli bir evrimsel sinir ağı yaklaşımı ile oldukça başarılı sonuçlar alınmıştır (Kim, 2014).

Bir diğer örnek çalışmada, metin sınıflandırma amacıyla bir yinelemeli evrimsel sinir ağı altyapısı önerilmiştir. Önerilen yaklaşımda; bağlamsal bilgi için yinelemeli yapı, metnin temsili içinse evrimsel sinir ağı kullanılmıştır (Lai vd., 2015).

Yine metin sınıflandırma konusundaki bir başka çalışmada, karakter seviyesinde evrimsel ağlar kullanılmıştır. Çalışmada, karakter seviyesindeki evrimsel sinir ağlarının metin sınıflandırma etkili bir yöntem olduğu ve bu yöntemin benzer problemlerin çözümünde de kullanılabileceği ifade edilmiştir (Zhang vd., 2015).

Başka bir çalışmada ise sıralı kısa metinlerin sınıflandırılması için yinelemeli sinir ağının bir türü olan LSTM ve evrimsel sinir ağı yöntemleri kullanılarak çeşitli veri kümelerinde test edilerek başarılı sonuçlar elde edilmiştir (Lee ve Dernoncourt, 2016).

### 3.2. Metin Ayırıştırma

Otomatik metin ayırıştırma, doğal dil işlemenin çok uzun yıllardır çalışılan ve çözülmesi amacıyla birçok farklı yöntemin literatüre kazandırıldığı bir alanıdır. Metin ayırıştırma verilen bir metnin gramatik yapısının ortaya çıkarılması amaçlanmaktadır.

Bu konuda gerçekleştirilen bir çalışmada, metin ayırıştırma için evrimsel sinir ağlarına dayalı ve hızlı bir yöntem önerilmiştir. Bu yöntemde evrimsel sinir ağları, grafiklerde yapılandırılmış etiket çıkarımı ile birleştirilmiş ve böylelikle literatürde daha önce mevcut olan grafik dönüştürücü ağ yöntemi kullanılmıştır. Önerilen yöntemin geliştirmeye açık yönleri bulunsa da, metin ayırıştırma için oldukça etkin bir yöntem olduğu ifade edilmiştir (Collobert, 2011).

### 3.3. Duygu Analizi

Duygu analizi, verilen bir metindeki duygu, fikir ve özneliğin otomatik olarak belirlenmesi olarak tanımlanmaktadır (Pang ve Lee, 2008). Yaygın olarak, duygu analizi çalışmalarında verilen metin kullanılarak bu metnin olumlu, olumsuz veya tarafsız şeklinde sınıflandırılması amaçlanmaktadır. Duygu analizi, doğal dil işlemenin uzun süredir çalışılan ve halen de yaygın olarak çalışılmaya devam edilen güncel araştırma konularından biridir. Bu alandaki güncel çalışmalardan bazılarında da derin öğrenme yöntemleri kullanılmaktadır.

Bu konudaki bir çalışmada, duygu analizi yaklaşımlarını alandan bağımsız hale getirebilmek ve böylelikle yeni metin türlerine de adapte edebilmek için oto-kodlayıcı tabanlı bir derin öğrenme yöntemi kullanılmıştır. Kullanılan oto-kodlayıcı yaklaşımıyla denetimsiz şekilde farklı metin türleri için gerekli duygu analizi öznelikleri otomatik olarak belirlenmektedir ve oldukça başarılı sonuçlar elde edilmiştir (Glorot vd., 2011). Bir diğer ilgili çalışmada yine derin oto-kodlayıcılar yarı-denetimli (semi-supervised) bir yaklaşımla eğitilerek duygu analizinde başarılı sonuçlar elde etmişlerdir (Zhai ve Zhongfei, 2016).

Bir diğer örnek çalışmada, evrimsel sinir ağları kullanılarak Twitter üzerinde duygu analizi gerçekleştiren bir yaklaşım anlatılmıştır. Söz konusu çalışmada üç aşamalı ve her aşamada farklı öğrenme yöntemlerinin kullanıldığı bir yaklaşım benimsenmiş, tek seviyeli basit bir evrimsel sinir

ağı kullanılmıştır (Severyn ve Moschitti, 2015a). Çalışma tweet veri kümeleri üzerinde denenmiş ve başarılı sonuçlar elde edilmiştir.

İlgili bir diğer çalışmada, kısa metinlerde (film yorumları ve tweet'ler) duygu analizi için yine evrimsel sinir ağı yaklaşımı benimsenmiştir (dos Santos ve Gatti, 2015). Bu yaklaşımda iki evrimsel katman vardır. Bu katmanlarda sözcüklerden karakter seviyesinde ve cümlelerden cümle seviyesindeki öznitelikler çıkarılmaktadır. Çalışmada önerilen yaklaşım film yorumları ve tweet'lerden oluşan iki ayrı veri kümesi üzerinde denenmiş ve başarılı sonuçlar elde edilmiştir (dos Santos ve Gatti, 2015).

### 3.4. Bilgi Çıkarımı

Bilgi çıkarımı, doğal dil işlemenin önemli araştırma alanlarından bir tanesidir. Bilgi çıkarımı ile doğal dildeki metinlerden; varlıklar, olaylar, tarih ve zaman ifadeleri, kavramlar gibi birçok kayda değer bilgi parçası otomatik şekilde çıkarılmaktadır.

Bilgi çıkarımı için derin öğrenme yöntemi kullanılan bir çalışmada, çıkarılan bilginin türünden bağımsız genel bir yaklaşım önerilmiştir (Qi vd., 2014). Söz konusu çalışmada, bilgi çıkarımı için derin sinir ağları kullanılmış ve özniteliklerin elle oluşturulması yerine özniteliklerin öğrenilmesi sağlanmaktadır. Çalışma, farklı bilgi çıkarımı veri kümelerinde denenmiş ve başarılı sonuçlar elde etmiştir (Qi vd., 2014).

### 3.5. Varlık İsmi Tanıma

Varlık ismi tanıma, doğal dil işlemenin bilgi çıkarımı alanının bir alt problemidir ve uzun süredir çalışılan bir araştırma konusudur. Varlık ismi tanıma sistemleri, genel olarak, verilen bir metindeki kişi, yer ve kurum gibi varlıkların isimlerini otomatik olarak çıkarıp sınıflandırır.

Derin öğrenme yöntemlerinin varlık ismi tanıma da başarılı sonuçlar elde ettiği görülmüştür. Bu konudaki bir çalışmada, varlık ismi tanıma için hibrit ve çift yönlü bir yaklaşım önerilerek hem yinelenen sinir ağlarının bir algoritması olan LSTM, hem de evrimsel sinir ağları kullanılmıştır. Kullanılan bu hibrit derin öğrenme yaklaşımıyla kelime seviyesindeki ve karakter seviyesindeki öznitelikler otomatik olarak belirlenmektedir. Söz konusu yaklaşım, varlık ismi tanıma alanındaki ortak veri kümeleri üzerinde denenmiş ve oldukça başarılı sonuçlar elde edilmiştir (Chiu ve Nichols, 2015).

### 3.6. Zamansal İlişki Çıkarımı

Zamansal ilişkilerin çıkarımı da bilgi çıkarımı konusunun bir alt konusudur ve dolayısıyla doğal dil işlemenin önemli bir araştırma alanıdır. Zamansal ilişkiler arasında şu ve benzeri ifadeler yer almaktadır: “*önce, sonra, hemen önce, hemen sonra, sırasında (esnasında)*”.

Bu konuyla ilgili yapılmış bir çalışmada, zamansal ilişki ifadelerinin çıkarılarak sınıflandırılması amacıyla evrimsel sinir ağları kullanılmıştır. Çalışmada sözlük seviyesindeki ve cümle seviyesindeki öznitelikler kullanılmıştır. Ancak, çalışmada literatürdeki benzer çalışmalardan daha iyi sonuçlar elde edilememiştir ve çeşitli ön işleme (preprocessing) yöntemleri kullanılarak sonuçların iyileştirilebileceği ifade edilmiştir (Do ve Jeong, 2016).

### 3.7. Olay Çıkarımı

Doğal dilde oluşturulmuş metinlerde bahsedilen olayların otomatik çıkarımı da bilgi çıkarımı konusunun bir alt alanıdır ve dolayısıyla doğal dil işlemenin bir problemidir.

Bu alanda yapılmış güncel bir çalışmada, otomatik olay çıkarımı için cümle seviyesindeki özniteliklerin çıkarılması için evrimsel sinir ağlarının kullanıldığı bir yaklaşım tanıtılmıştır. Bu çalışmada ayrıca, kullanılan kelime seviyesindeki öznitelikler içinse farklı bir kelime temsili

yaklaşımı benimsenmiştir. Olay çıkarımı için derin öğrenme yöntemi kullanan bu çalışmanın olay çıkarımı deneylerinde başarılı sonuçlar verdiği gösterilmiştir (Chen vd., 2015).

### 3.8. Sözcük Türü Etiketleme

Sözcük türü etiketleme, doğal dilde verilen bir metindeki kelimelerin türlerinin sınıflandırılması problemidir. Sözcük türü sınıfları isim, fiil, edat, zamir gibi türlerdir. Sözcük türü etiketleme doğal dil işlemenin çok uzun süredir çalışılan bir konudur ve ayrıca daha kapsamlı problemlerin çözümünde de söz konusu sözcük türleri kullanılabilir.

Bu konuda yapılmış bir çalışmada, derin sinir ağları kullanılarak sözcük türü belirleme amacıyla sözcük seviyesinde ve karakter seviyesindeki temsillerin öğrenilmesi sağlanmıştır. Söz konusu sinir ağında, öznitelik çıkarımı için evrişimsel bir katman kullanılmaktadır. İngilizce ve Portekizce üzerinde yapılan deneylerde oldukça başarılı sonuçlar elde edilmiştir (dos Santos ve Zadrozny, 2014).

Sözcük türü etiketleme konusunda yapılmış dilden bağımsız bir çalışmada ise çift yönlü LSTM ağları kullanılmıştır. 22 farklı dil üzerinde yapılan testler sonucunda söz konusu derin öğrenme metoduyla başarılı sonuçlar elde edildiği gözlenmiştir (Plank vd., 2016).

### 3.9. Metin Sıralama

Metin sıralama, bilgi erişiminin kapsamında bulunan önemli bir araştırma konusudur. Bu sıralama işlemi için de derin öğrenme yöntemini kullanan çalışmalar mevcuttur.

Bu yöndeki bir çalışmada, evrişimsel derin öğrenme yöntemi kullanılarak kısa metinlerin sıralanması konusu incelenmiştir. Bu sıralama işlemindeki ana konu, metinlerin sıralanması için kullanılacak benzerlik işlevinin öğrenilmesidir. Söz konusu çalışmada kullanılan derin öğrenme yöntemiyle, elle öznitelik çıkarma işlemi veya dış kaynaklar kullanılmasına gerek kalmadığı ifade edilmiştir. Önerilen yaklaşım, kısa metinlerin sıralanması deneylerinde başarılı sonuçlara ulaşmıştır (Severyn ve Moschitti, 2015b).

### 3.10. Otomatik Harf Çevirisi

Otomatik çeviri alanında kullanılan otomatik harf çevirisi alanında da derin öğrenme yöntemlerinin kullanıldığı görülmektedir. Otomatik harf çevirisi ile kelimelerin (özel isimler gibi) bir alfabedeki harflerden diğer alfabedeki harflere çevrilmesi sağlanmaktadır.

Bu konuda yapılmış bir çalışmada; derin inanç ağları yöntemi, Arapça-İngilizce arası harf çevirisinde kullanılmıştır. Bu konuda; daha önceki çalışmalarda, sonlu-durum makinaları ve ifade-tabanlı yöntemler kullanılmıştır. Derin inanç ağları, önceki çalışmalardaki bu yöntemlerden daha başarılı sonuçlar vermese de mevcut diğer yöntemlerle beraber kullanılabilirdiği ifade edilmiştir (Deselaers vd., 2009).

### 3.11. Otomatik Soru Cevaplama

Otomatik soru cevaplama sistemleri, doğal dilde ifade edilmiş bir sorunun cevabını doğal dildeki büyük veri kümeleri içerisinde bularak yine doğal dilde otomatik olarak cevaplamayı amaçlayan sistemlerdir.

Derin öğrenme yöntemleri soru cevaplama sistemlerinde de kullanılmıştır. Örneğin, Yu vd. (2014) tarafından gerçekleştirilen çalışmada, soru cevaplama işleminin bir alt işlemi olan cevap cümlesi seçimi için evrişimsel sinir ağları kullanılmış ve karşılaştırılan mevcut başarılı sistemlere yakın sonuçlar elde edilmiştir.

İlgili bir diğer çalışmada yine soru cevaplama probleminin çözümünde kullanılmak üzere cevap seçimi için çift yönlü LSTM algoritması ile evrişimsel sinir ağlarını kullanılmıştır. Önerilen

birleşik yöntem, karşılaştırmada kullanılan diğer algoritmalara göre başarılı sonuçlar elde etmiştir(Tan vd., 2015).

#### 4. SONUÇ

Derin öğrenme, son dönemde oldukça popüler olmuş önemli bir araştırma alanıdır. Sınır ağlarına dayalı olan derin öğrenme konusunda literatürde birçok farklı yöntem önerilmiştir ve bunların bazıları özellikle bazı araştırma alanlarında oldukça başarılı sonuçlar elde edilmesini sağlamıştır. Doğal dil işleme alanı da söz konusu önemli araştırma alanlarından biridir.

Bu çalışmamızda, doğal dil işleme için derin öğrenme yöntemleri konusunda yapılmış çalışmaların derlenmesi amaçlanmıştır. Çalışmamızda ilk olarak, derin öğrenme konusunda temel bilgiler ile literatürde mevcut bulunan altı önemli derin öğrenme yöntemi hakkında detaylı bilgiler ve bunların belli başlı uygulama alanları verilmiştir. Ardından da doğal dil işleme için derin öğrenme yöntemlerini tanıtan çalışmalar hakkında bilgiler sunulmuş, yöntemlerin başarımları ve yetersiz kaldığı durumlar açıklanmıştır. Söz konusu derin öğrenme çalışmaları, ilgili doğal dil işleme problemlerine göre mümkün olduğunca sınıflandırılmış, ayrı alt başlıklar altında verilmiştir. Çalışmamızın, Türkçe doğal dil işleme alanında derin öğrenme yöntemlerini uygulayacak araştırmacılara Türkçe bir kaynak olarak ışık tutması ve yol gösterici olması beklenmektedir.

#### KAYNAKÇA

- Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning* 2(1), 1-127.
- Chen, Y., Xu, L., Liu, K., Zeng, D., & Zhao, J. (2015). "Event extraction via dynamic multi-pooling convolutional neural networks". *Annual Meeting of the Association for Computational Linguistics and International Joint Conference on Natural Language Processing*, 167-176.
- Chiu, J. P., & Nichols, E. (2015). "Named entity recognition with bidirectional LSTM-CNNs". *arXiv preprint arXiv:1511.08308*.
- Cho, Y., & Saul, L. K. (2009). "Kernel methods for deep learning". *Advances in Neural Information Processing Systems*, 342-350.
- Collobert, R. (2011). "Deep learning for efficient discriminative parsing". *International Conference on Artificial Intelligence and Statistics*, 224-232.
- Collobert, R., & Weston, J. (2008). "A unified architecture for natural language processing: Deep neural networks with multitask learning". *International Conference on Machine Learning (ICML)*, 160-167.
- Conneau, A., Schwenk, H., Barrault, L., & Lecun, Y. (2016). "Very deep convolutional networks for natural language processing". *arXiv preprint arXiv:1606.01781*.
- Deng, L., & Yu, D. (2014). "Deep learning: methods and applications". *Foundations and Trends in Signal Processing*, 7(3-4), 197-387.
- Deselaers, T., Hasan, S., Bender, O., & Ney, H. (2009). "A deep learning approach to machine transliteration". *International Workshop on Statistical Machine Translation*, 233-241.
- Do, H. W., & Jeong, Y. S. (2016). "Temporal relation classification with deep neural network". *International Conference on Big Data and Smart Computing (BigComp)*, 454-457.

- Dos Santos, C. N., & Gatti, M. (2014). "Deep convolutional neural networks for sentiment analysis of short texts". *International Conference on Computational Linguistics (COLING)*, 69-78.
- Dos Santos, C. N., & Zadrozny, B. (2014). "Learning character-level representations for part-of-speech tagging". *International Conference on Machine Learning (ICML)*, 1818-1826.
- Glorot, X., Bordes, A., & Bengio, Y. (2011). "Domain adaptation for large-scale sentiment classification: A deep learning approach". *International Conference on Machine Learning (ICML)*, 513-520.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). "Reducing the dimensionality of data with neural networks". *Science*, 313(5786), 504-507.
- Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). "A fast learning algorithm for deep belief nets". *Neural Computation*, 18(7), 1527-1554.
- Hinton, G. E., & Sejnowski, T. J. (1986). "Learning and relearning in boltzmann machines". *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1(282-317), 2.
- Hochreiter, S., & Schmidhuber, J. (1997). "Long short-term memory". *Neural Computation*, 9(8), 1735-1780.
- Kim, Y. (2014). "Convolutional neural networks for sentence classification". *arXiv preprint arXiv:1408.5882*.
- Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). "Recurrent convolutional neural networks for text classification". *AAAI Conference on Artificial Intelligence*, 2267-2273.
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning". *Nature* 521(7553), 436.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*, 86(11), 2278-2324.
- Lee, D. H. (2013). "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks". *Workshop on Challenges in Representation Learning, ICML (3)*, 2
- Lee, J. Y., & Deroncourt, F. (2016). "Sequential short-text classification with recurrent and convolutional neural networks". *arXiv preprint arXiv:1603.03827*.
- Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). "Multimodal deep learning". *International Conference on Machine Learning*, 689-696.
- Pang, B., & Lee, L. (2008). "Opinion mining and sentiment analysis". *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Plank, B., Søgaard, A., & Goldberg, Y. (2016). "Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss". *arXiv preprint arXiv:1604.05529*.
- Qi, Y., Das, S. G., Collobert, R., & Weston, J. (2014). "Deep learning for character-based information extraction". *European Conference on Information Retrieval*, 668-674.
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). "Deep learning for health informatics". *IEEE journal of Biomedical and Health Informatics* 21(1), 4-21.
- Rosenblatt, F. (1958). "The perceptron: a probabilistic model for information storage and organization in the brain". *Psychological Review*, 65(6), 386.

- Salakhutdinov, R., & Larochelle, H. (2010). “Efficient learning of deep Boltzmann machines”. International Conference on Artificial Intelligence and Statistics.
- Severyn, A., & Moschitti, A. (2015) a. “Twitter sentiment analysis with deep convolutional neural networks”. International ACM SIGIR Conference on Research and Development in Information Retrieval, 959-962.
- Severyn, A., & Moschitti, A. (2015) b. “Learning to rank short text pairs with convolutional deep neural networks”. International ACM SIGIR Conference on Research and Development in Information Retrieval, 373-382.
- Socher, R., Bengio, Y., & Manning, C. D. (2012). “Deep learning for NLP (without magic)”. Tutorial Abstracts of ACL 2012, 5.
- Tan, M., Santos, C. D., Xiang, B., & Zhou, B. (2015). “LSTM-based deep learning models for non-factoid answer selection”. arXiv preprint arXiv:1511.04108.
- Williams, R. J., & Zipser, D. (1989). “A learning algorithm for continually running fully recurrent neural networks”. Neural Computation, 1(2), 270-280.
- Yu, L., Hermann, K. M., Blunsom, P., & Pulman, S. (2014). “Deep learning for answer sentence selection”. arXiv preprint arXiv:1412.1632.
- Zhai, S., & Zhang M. Z. (2016). “Semisupervised autoencoder for sentiment analysis”. AAAI Conference on Artificial Intelligence, 1394-1400.
- Zhang, X., Zhao, J., & Lecun, Y. (2015). “Character-level convolutional networks for text classification”. Advances in Neural Information Processing Systems, 649-657.



# TESLACRYPT FİDYE YAZILIM VİRÜSÜNÜN TESPİTİ, TEKNİK ANALİZİ VE ÇÖZÜMÜ

\*\*\*

## DETECTION, TECHNICAL ANALYSIS AND SOLUTION OF TESLACRYPT RANSOMWARE VIRUS

İlker KARA\*

### Öz

Bilişim teknolojilerinde yaşanan hızlı gelişmeler internet kullanıcılarının hayatında pek çok şeyi kolaylaştırmışken, kötü niyetli kişilerinde amaçlarına daha hızlı ulaşması için yeni olanaklar sağlamaktadır. İlk tasarım amacından tamamen uzaklaşan zararlı yazılımlar, günümüzde profesyonel suçlular tarafından siber terörizmden fidye istemeye kadar geniş bir uygulama alanı için tasarlanmaktadır. Bu suçlular çok çeşitli yöntemler ve taktikler geliştirerek amaçlarına kolaylıkla ulaşmakta, bu duruma maruz kalma olasılığı kullanıcıların korkulu rüyası haline gelmektedir. Son günlerde dünya genelinde TeslaCrypt olarak adlandırılan yeni nesil fidye yazılım siber saldırı vakaları görülmeye başlamıştır. TeslaCrypt, kullanıcılara e-mail yoluyla ulaşarak ekinde bulunan zararlı yazılımın çalıştırılması ile sistemdeki birçok dosyayı şifrelemektedir. Kullanıcının şifrelenmiş dosyalara erişim sağlayabilmesi için fidye göndermesini istemektedir. TeslaCrypt'nin sebep olduğu bu durum için çalışmalar devam etmekle birlikte hala kesin çözüm bulunamamıştır. Bu çalışma, TeslaCrypt tehditinin tespiti, hedef sisteme sızma, sistemdeki dosya-dizin hareketlerinin teknik analizini ve çözümünü içermektedir. İncelemeler; hem statik hem de dinamik yöntemlerle gerçekleştirilmiştir. Çalışma sonucunda fidye yazılımının neden olduğu şifrelerin kırılabilir olduğu gösterilmiştir.

**Anahtar Kelimeler:** Zararlı Yazılım Analizi, Fidye Yazılımı, Kriptokitleyici.

### Abstract

Although the rapid developments in information technologies have facilitated numerous things in the lives of Internet users, these developments also allow malicious people to reach their goals faster. Malicious software that completely drift away from their initial design goal are now being designed by professional criminals for a wide range of applications from cyber terrorism to ransom demands. These criminals reach their goals easily by developing a variety of methods and tactics, and the possibility of being exposed to this situation becomes the worst nightmare for the users. Recently, a new generation of Ransomware, known as TeslaCrypt, has begun to be seen worldwide. TeslaCrypt reaches users through e-mail and encrypts many files in the system after execution of its payload found in the e-mail attachment. It demands ransom to allow access to encrypted files of the user. Although there are continuing works to find a solution to this problem caused by TeslaCrypt, there is still no definitive solution. This study discusses the detection of TeslaCrypt threat, and technical analysis on its infiltration into the target system and file-directory actions in the system and solution. The analysis has been performed by both static and dynamic methods. As a result of the study, it was shown that the passwords caused by the ransomware virus broke the password.

**Keyword:** Malware analysis, Ransomware, Cryptolocker.

\* Öğr.Gör.Dr.Hacettepe Üniversitesi, Bilgisayar Mühendisliği, Bilişim Enstitüsü, [karaikab@gmail.com](mailto:karaikab@gmail.com)  
ORCID: 0000-0003-3700-4825

## 1.GİRİŞ

Tüm dünyada olduğu gibi, ülkemizde de son günlerde fidye yazılım saldırıları hızla artmaya başlamıştır (Shen vd., 2018:1; Feizollah vd., 2017:65). Her geçen gün daha fazla kullanıcıyı etkileyen bu zararlı yazılımlar, kullanıcıların zaman ve maddi kayba uğramasına neden olmaktadır (Rieck vd., 2011:640). Saldırganlar, fidye yazılım yoluyla kullanıcıların sisteminde bulunan dosyaları şifrelemekte daha sonra ise şifrelerin çözümü için fidye talep etmektedir. Hedef sisteme sızmak için en yaygın kullandıkları yöntem, kullanıcılara e-postalar göndermek ve e-posta ekinde bulunan zararlı yazılımı sistemlerine kurmak için yönlendirmektir (Bassett vd., 2006:1; Kara., 2015:87; Rieck vd., 2011:641).

Bu yöntemde, kullanıcıya ortalama yöntemini kullanılarak tasarlanmış bir e-posta içeriği gönderilmektedir. İlk bakışta resmi bir kurumdan gelen bir belge gibi görünen içerik, kullanıcının dikkatini çekmek ve onu korkutmak için özel olarak tasarlanmıştır. Burada amaç kullanıcının e-posta ekinde bulunan .exe uzantılı zararlı yazılımı aktif hale getirmesini sağlamaktır. Kullanıcı bu dosyayı çalıştırdığı andan itibaren zararlı yazılım hedef sistemde aktif olmaya başlamakta ve sistemin alt katmanına yerleşip, dosyayı şifrelemektedir (Yaqoob vd., 2017:444).

Hedef sistemde aktif hale gelen zararlı yazılım ilk olarak kendini rastgele isimlerle C:/Windows/ dizininin içine kopyalamaktadır (Luo ve Liao, 2007:196). Sistem her açıldığında otomatik olarak aktif hale gelebilmek için sistemin kayıt defterinde bir anahtar girdisi oluşturmaktadır (Bhardwaj vd., 2016:2). Bu aşamalardan sonra zararlı yazılım hedefteki tüm dosyaları şifrelemektedir (Richardson ve North, 2017:11).

Son günlerde ülkemizde sıklıkla görülmeye başlanan TeslaCrypt fidye yazılımı oldukça tehlikeli yeni nesil fidye yazılımdır (Garg vd., 2018:245). TeslaCrypt fidye yazılımlarının ilk sürümleri özellikle oyun dosyaları üzerinde etkili olmuş, oyun kartları ve oyuncu profilleri gibi dosyaları hedef almıştır (Salz vd., 2003:12). Güncellenen sürümlerde şifreleme kapasitesi daha da artmıştır. TeslaCrypt fidye yazılım, hedef sistemdeki dosyaları şifreledikten sonra kullanıcıya, dosyaların şifrelendiğini bildiren bir mesaj göndermekte, şifrelerin çözümü içinse bitcoin eşdeğer miktarda 500\$ veya € ödeme istemektedir (Villeneuve, 2015:13). Bu davranışları ile en çok bilinen fidye yazılım olan Cryptolocker (kripto kitleyicilere) ile çok benzemektedir (Zheng vd., 2017:313).

## 2. TESLACRYPT NASIL BULAŞIR?

TeslaCrypt fidye yazılımının hedef sistemlere sızma yöntemlerini genel olarak şöyle sıralanabilir;

- e-posta içeriğinde bulunan .exe uzantısı ile,
- Ücretsiz paylaşım sitelerinden (Torrent'ler gibi) indirilen oyun, film, müzik dosyaları ile,
- Uzak Masaüstü erişimi (RDP) ya da diğer sistem açıklıklarını kullanarak,
- Ücretsiz çevrimiçi film, dizi izleme bloklarını kullanabilmek için flash/java gibi programların eklentilerinin indirilmesiyle,
- Resmi olduğu düşünülen, siber suçlular tarafından modifiye edilerek değiştirilen kaynaklardan indirilen uygulamalar ile,
- Sosyal paylaşım blogları, haber ve ilan sitelerinin ziyaretleri sırasında (istemsiz), hedef sisteme sızmakatadır.

TeslaCrypt fidye yazılımı, AES (Advanced Encryption Standard) şifreleme yönteminde 256-bit standart bir algoritma kullanmaktadır. AES şifrelemesi, Amerikan hükümeti tarafından uluslararası alanda da defacto şifreleme standardı olarak kullanılmaktadır (Aurangzeb, 2017:2).

TeslaCrypt fidye yazılım tehdidinden kurtulmak ve bu yazılım ile etkin bir mücadele geliştirmek için öncelikle tehditin detaylı tanımlanması gerekmektedir. TeslaCrypt'ların hedef

sisteme sızma ve dosyaları şifreleme işlemi sırasında kullandığı yöntemlere yönelik uygulamalı çalışmalar, problemin çözümünde önemli katkılar sağlayabilir.

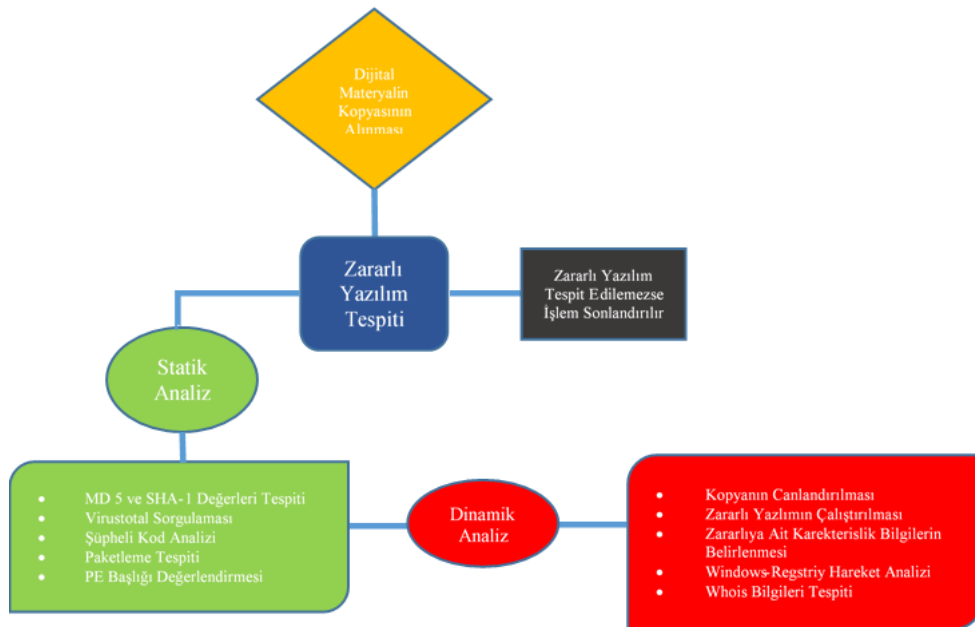
Bu çalışmada, yeni nesil TeslaCrypt fidye yazılımının hedef sistemde tespiti edilmesi, sisteme sızması ve dosyaları şifreleme davranışları detaylı olarak incelenmiş olup çalışma sonunda şifrelenen dosyalar başarılı şekilde kurtarılmıştır.

### 3. METOT ve BULGULAR

Zararlı yazılım analizleri için standart olarak uygulanan bir metod bulunmamaktadır. Bununla beraber ilk olarak zararlı yazılım çalıştırmadan yapılan statik analiz, ikinci olarak kontrollü bir ortamda zararlı yazılımı çalıştırarak yapılan davranış hareketlerinin (dosya-dizin hareketleri) analiz edildiği dinamik analiz ve son olarak ise zararlı yazılımın sahip olduğu mimarisinin analizi için kod analizi yapılmaktadır. Zararlı yazılımın bulunduğu kurban bilgisayarda (canlı sistemde) kontrolün zor olması, zararın daha büyük boyutlara ulaşması gibi kötü senaryolar nedeniyle prensip olarak inceleme yapılamamaktadır. Bu nedenle kurban bilgisayar kopyası alınarak farklı bir ortamda (iş bilgisayarında) analizler yapılmaktadır.

TeslaCrypt fidye yazılım saldırısına maruz kalmış bilgisayarın, FTK Imager (free version) programı kullanarak kopyası alınmıştır. İncelenen TeslaCrypt fidye yazılımı çalıştırdıktan sonra kullanıcı verilerine hızlı bir şekilde saldıracağından, inceleme yapılan iş makinesi sanal makine modunda çalıştırılmıştır. TeslaCrypt fidye yazılımının karakteristik davranış analizi için yapılan incelemeler “AccessData Forensic Toolkit v6.2.1.10 (FTK)”, “Process Explorer” “Cuckoo” aracılığıyla gerçekleştirilmiştir.

Zararlı yazılım analizleri için standart olarak uygulanan bir metod olmamakla beraber genel eğilim basitten karmaşığa doğru ilerlemektedir. İlk olarak zararlı yazılım çalıştırılmadan elde edilebilecek tüm bilgilere ulaşmak daha sonra zararlı yazılımın kontrollü bir ortamda çalıştırılmasıyla davranış hareketlerini incelemek son olarak ise zararlı yazılımın kod mimarisinin incelenmesi doğru bir yöntem olacaktır. Bu çalışmada, analiz adımları için bir model önerilmiş ve önerilen model algoritması uygulanmıştır (Şekil 1).



Şekil 1. Önerilen zararlı yazılım analiz model algoritması.

Önerilen model algoritması doğrultusunda incelemelerde ilk olarak statik analiz yapılmıştır. Statik analiz sonucunda, TeslaCrypt fidye yazılımının bulunduğu kurban bilgisayara ait bilgiler FTK ve Process Explorer programları kullanılarak elde edilmiş olup Tablo 1 ve Tablo 2’de verilmiştir.

**Tablo 1.** Cihaz Bilgileri

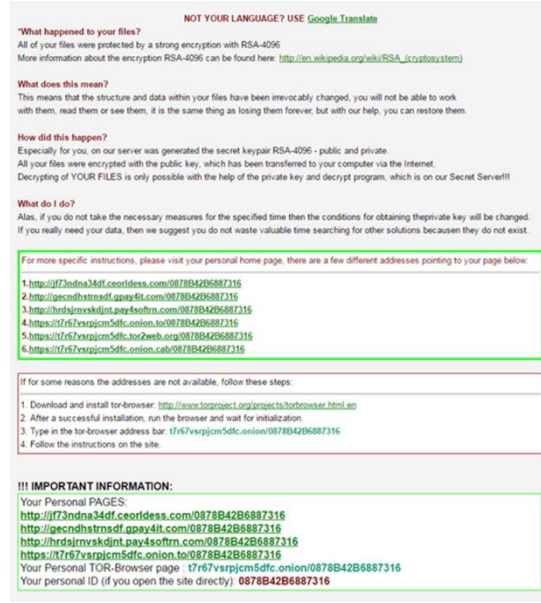
Açıklama	Physical Disk, 1.953.525.168 Sectors 931,5 GB
Toplam Kapasite	1.000.204.886.016 Bytes (931,5 GB)
Toplam Sektör	1.953.525.168
MD5 Değeri	5ebd653ee743d236356914173e8c6174
MD5 Doğrulama	5ebd653ee743d236356914173e8c6174
SHA1 Değeri	6f316c8a441fd83a6ee04caffbd87d1a3c715ab0
SHA1 Doğrulama	6f316c8a441fd83a6ee04caffbd87d1a3c715ab0

**Tablo 2.** İşletim Sistemi Bilgileri

Ürün Adı	Windows 7 Ultimate
Kayıtlı Sahip	user
Sistem Kökü	C:\Windows\
Güncel Veriyonu	6.1
CSD Versiyonu	Service Pack 1
Kurulum Tarihi	27.11.2015 11:55:39 UTC
Son Kapanma Tarihi	23.12.2015 10:57:10 UTC

Tablo 1 ve 2’de TeslaCrypt fidye yazılımının bulunduğu kurban bilgisayara ait bilgiler zararlı yazılım analizleri için gereklidir. Bu bilgilerden yolla çıkarak analiz için kullanılacak programlarının ilgili sürüm ve özellikleri seçilmektedir. Ayrıca buradan edinilen bilgiler zararlı yazılımın olası karakterislik hareketlerinin tahmin edilmesi sağlamaktadır.

Önerilen model algoritması doğrultusunda incelemelerin ikinci aşaması olarak dinamik analiz yapılmıştır. Dinamik analiz için alınan kopya çalıştırıldığında saldırganın ait olduğu anlaşılacak mesaj kurban bilgisayarın ekranında görülmüştür (Şekil 2). Mesajdan da anlaşılacağı gibi saldırının nedeni “TeslaCrypt” fidye yazılımı olduğu tespit edilmiştir. Hedef sistemde aktif hale gelen TeslaCrypt fidye yazılıma ait ilk iz “IMAGE.E01/Partition 2/NONAME [VSC]/[Current]/[root]/Users/user/Desktop/” dizini altında bulunan “**Howto\_Restore\_FILES.HTM**” isimli dosya içerisinde Şekil 2’de verilen nota ulaşılmıştır. Bu aşamadan sonra incelemeler “Howto\_Restore\_FILES.HTM” dosyası üzerine yoğunlaştırılmıştır.



Şekil 2. Kurban bilgisayarda TeslaCrypt virüsünün olduğu bildiren saldırıya ait mesaj.

Tablo 3. FTK ve Cuckoo programları kullanılarak elde edilen "Howto\_Restore\_FILES.HTM" dosyasına ait teknik bilgileri.

Dosya Bilgileri	
Dosya Adı	Howto_Restore_FILES.HTM
Oluşturma Zamanı	20.12.2015 13:57:55 (2015-12-20 11:57:55 UTC)
Erişim Zamanı	20.12.2015 13:57:55 (2015-12-20 11:57:55 UTC)
Değiştirme Zamanı	20.12.2015 13:57:55 (2015-12-20 11:57:55 UTC)
Dosya Boyutu (Byte)	10.631 bytes (10,38 KB)
MD5 Hash Değeri	83e36d32d8e26e55515437c4492d6270
Dosya Yolu	IMAGE.E01/Partition 2/NONAME [VSC]/[Current]/[root]/Users/user/Desktop/

TeslaCrypt fidye yazılımının hedef sistemdeki hangi dosyaları şifrelediğini tespit etmek için yapılan incelemelerde .vzv uzantısına sahip çok sayıda şifreli dosyanın varlığı tespit edilmiştir (Şekil 2).

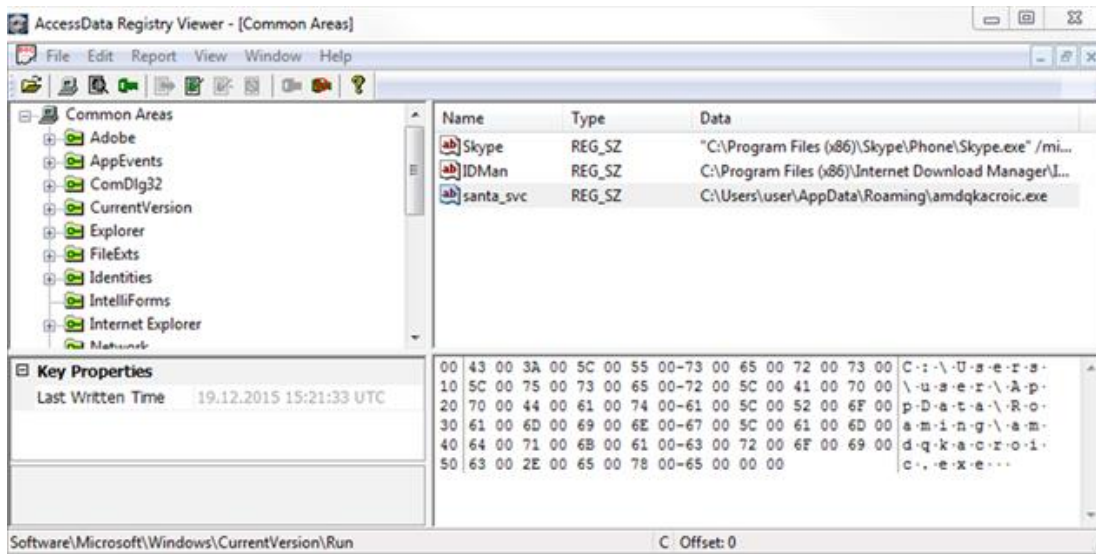
<input checked="" type="checkbox"/>	Name	Modified
<input type="checkbox"/>	app.css.vzv	19.12.2015 17:21:34 (2015-12-19 15:21:34 UTC)
<input type="checkbox"/>	ai.js.vzv	19.12.2015 17:21:34 (2015-12-19 15:21:34 UTC)
<input type="checkbox"/>	id.js.vzv	19.12.2015 17:21:34 (2015-12-19 15:21:34 UTC)
<input type="checkbox"/>	ps.js.vzv	19.12.2015 17:21:35 (2015-12-19 15:21:35 UTC)
<input type="checkbox"/>	require.js.vzv	19.12.2015 17:21:35 (2015-12-19 15:21:35 UTC)
<input type="checkbox"/>	main.js.vzv	19.12.2015 17:21:35 (2015-12-19 15:21:35 UTC)
<input type="checkbox"/>	main.css.vzv	19.12.2015 17:21:36 (2015-12-19 15:21:36 UTC)
<input type="checkbox"/>	main.css.vzv	19.12.2015 17:21:36 (2015-12-19 15:21:36 UTC)
<input type="checkbox"/>	main.css.vzv	19.12.2015 17:21:36 (2015-12-19 15:21:36 UTC)
<input type="checkbox"/>	main.css.vzv	19.12.2015 17:21:36 (2015-12-19 15:21:36 UTC)
<input type="checkbox"/>	main.css.vzv	19.12.2015 17:21:36 (2015-12-19 15:21:36 UTC)

Şekil 3. Kurban bilgisayardaki dosyaların şifrelenmiş olduğunu gösteren ekran görüntüsü.

Kopya üzerinde yapılan incelemelerde dosya şifreleme işlemlerinin “19.12.2015 17:21:34 (2015-12-19 15:21:34 UTC)” zaman diliminde başladığının tespit edilmesi üzerine, bu tarihte çalıştırılan dosyalar üzerinde incelemeler gerçekleştirilmiştir. Yapılan incelemelerde olay günü "IMAGE.E01/Partition2/NONAME[VSC]/[Current]/[root]/Users/user/NTUSER.DAT»Software»Microsoft»Windows»CurrentVersion»Run»" altında bulunan registry kaydından “amdqkacroic.exe” (TeslaCrypt) isimli zararlı yazılım dosyalarının şifrelendiği tespit edilmiştir (Şekil 3).

“amdqkacroic.exe” isimli zararlı yazılım araştırıldığında söz konusu zararlı yazılımın dosyaları şifreledikten sonra kendisini sildiği görülmüştür. Bu aşamadan sonra TeslaCrypt fidye yazılımının sebep olduğu şifrelenmiş dosyaların kurtarılması yönünde araştırmalar gerçekleştirilmiştir. Şifre kırmak için birçok yöntem olmakla beraber inceleme konusu TeslaCrypt fidye yazılımının çok güçlü bir şifreleme standartını kullandığı bilinmektedir. Bu nedenle brute-force (her olası harfleri, numaraları ve karakterleri doğru şifreyi bulana kadar denemek) yöntemini kullanmak mantıklı değildir. Çözümlemiş fidye yazılımların şifrelerini içeren kelime listesi (word list) yöntemini kullanmak bazı durumlarda başarılı sonuçlar vermektedir.

Bu çalışmada şifrelenmiş dosyaları kurtarma işlemleri için “AccessData Registry Viewer” adlı program kullanılmıştır. AccessData Registry Viewer programı, windows işletim sistemi için AccessData tarafından geliştirilmiş ve FTK programı ile ilişkili olarak çalıştırılmaktadır. Önceden hazırlanmış word list ile yapılan analizler sonucunda şifrelenmiş dosyalar başarılı şekilde kurtarılmıştır (Şekil 4).



Şekil 4. “amdqkacroic.exe” isimli zararlı yazılımın çalıştırıldığına ait ekran kaydı.

#### 4. SONUÇLAR VE DEĞERLENDİRME

Her geçen gün yeni tasarımlar ile piyasaya sürülen zararlı yazılımlar tüm kullanıcılar için giderek büyüyen bir tehlike haline gelmiştir. Fidye yazılımlar hedef sisteme sızdıktan sonra sistemde bulunan dosyaları şifrelemekte, dosyalara tekrar erişim sağlanmak için gerekli anahtar alma karşılığında ise saldırganlara fidye ödenmesini talep etmektedir.

TeslaCrypt, yeni nesil fidye yazılım türlerinden birisidir. Sanal ortamda her gün yayılmaya devam etmekte ve birçok kullanıcının başına bela olmaktadır. Saldırganlar, TeslaCrypt fidye yazılımı ile daha ziyade küçük işletmeleri ve şirketleri hedef almaktadır. Son günlerde artan saldırılar sonucunda, TeslaCrypt fidye yazılımından etkilenen kullanıcı sayısında ciddi artışlar

görülmüştür. Bu nedenle gerçek bir olaydan alınan TeslaCrypt fidye yazılım saldırısı detaylı olarak incelenmiştir. Bu alanda yapılan çalışmalara teknik analiz boyutunun gerekli olduğu sunulmuştur. Tespiti yapılan TeslaCrypt fidye yazılımının şifrelediği dosyalar üzerinde yapılan incelemeler tamamlandıktan sonra, şifrelenen dosyalar başarılı şekilde kurtarılmıştır.

İşletmelerin, kişilerin aleyhlerine giderek sıklaşan ve giderek daha tehlikeli hale gelen kötü niyetli saldırılara karşı, güvenlik tehditlerini ve zaafları ortadan kaldıran kapsamlı bir çözüm gereksinimleri vardır. Fidye yazılımlar tarafından şifrelenmiş dosyalar erişimi engellendiği için temel olarak onarılamaz kabul edilmektedir. Bu nedenle fidye yazılımların saldırılarına maruz kalmamak için bir takım tedbirler alınması gereklidir. Fidye yazılım saldırılara karşı alınması gereken önlemler şu şekilde önerilebilir;

- i. Fidye yazılımlar en yaygın sızma yöntemi e-posta yolunu kullanmaktadır. Bu nedenle kaynağından emin olunmayan bu gibi sahte e-postaların içindeki sahte faturaları açmamak.
- ii. Kullanıcı e-mail sistemlerine AntiSpam ürünlerini kullanmak,
- iii. Kullanıcılara fatura, haciz, kargo bilgileri gibi e-faturanın zip formatında veya exe formatında olmayacağını farkındalığının oluşturmak
- iv. Herkes tarafından kabul gören güncel bir Internet Security yazılımı ile sistemleri koruma altına almak,
- v. Uzaktan erişim portunu değiştirmek (mümkünse port numarası uzun karakterlerden oluşturmak), uzaktan erişen kullanıcılara kısıtlı hesaplar ve güçlü şifreler oluşturmak ve kullanılmıyacaksa uzaktan erişime sisteme kapatmak,
- vi. Daha önce tespit edilen fidye yazılım alanlarına girişleri bloke etmek,
- vii. Kullanıcının önemli verilerini başka bir depolama sisteminde günlük yedek almak (En az 30 günlük “geri yükleme noktaları” oluşturmak).

Bu ve benzeri önlemleri alarak kesin olarak fidye yazılımlardan korunmamakla birlikte kullanıcılar için farkındalık oluşturması açısından önemlidir. Araştırmanın yalnızca siber güvenlik endüstrisi üzerinde değil, yakın gelecekte yapılacak araştırmalarda ve birçok bireysel internet kullanıcısının üzerinde de olumlu bir etki yaratmasını beklenmektedir.

## KAYNAKÇA

- Aloul, F. A. (2012). The need for effective information security awareness. *Journal of Advances in Information Technology*, 3(3), 176-183.
- Aurangzeb, S., Aleem, M., Iqbal, M. A., & Islam, M. A. (2017). Ransomware: A Survey and Trends. *Journal of Information Assurance & Security*, 6(2).
- Bassett, R., Bass, L., & O'Brien, P. (2006). Computer forensics: An essential ingredient for cyber security. *Journal of Information Science & Technology*, 3(1).
- Bhardwaj, A., Avasthi, V., Sastry, H., & Subrahmanyam, G. V. B. (2016). Ransomware digital extortion: a rising new age threat. *Indian Journal of Science and Technology*, 9(14), 1-5.
- Feizollah, A., Anuar, N. B., Salleh, R., Suarez-Tangil, G., & Furnell, S. (2017). Androdialysis: Analysis of android intent effectiveness in malware detection. *computers & security*, 65, 121-134.
- Garg, D., Thakral, A., Nalwa, T., & Choudhury, T. (2018). A Past Examination and Future Expectation: Ransomware. In *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 243-247). IEEE.

- Kara, İ., (2015). Türkiye de Zararlı Yazılımlarla Mücadelenin Uygulama Ve Hukuki Boyutunun Değerlendirilmesi. Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi 52: 87-98.
- Luo, X., & Liao, Q. (2007). Awareness education as the key to ransomware prevention. Information Systems Security, 16(4), 195-202.
- Salz, J., Balakrishnan, H., & Snoeren, A. C. (2003). TESLA: A Transparent, Extensible Session-Layer Architecture for End-to-end Network Services. In USENIX Symposium on Internet Technologies and Systems.
- Shen, J., Gong, S., & Bao, W. (2018). Analysis of Network Security in Daily Life. Information and Computer Security, 1(1).
- Richardson, R., & North, M. (2017). Ransomware: Evolution, mitigation and prevention. International Management Review, 13(1), 10-21.
- Rieck, K., Trinius, P., Willems, C., & Holz, T. (2011). Automatic analysis of malware behavior using machine learning. Journal of Computer Security, 19(4), 639-668.
- Villeneuve, N. (2015). TeslaCrypt: Following the Money Trail and Learning the Human Costs of Ransomware. Threat Research Blog.
- Yaqoob, I., Ahmed, E., ur Rehman, M. H., Ahmed, A. I. A., Al-garadi, M. A., Imran, M., & Guizani, M. (2017). The rise of ransomware and emerging security challenges in the Internet of Things. Computer Networks, 129, 444-458.
- Zheng, B., Zhu, L., Shen, M., Du, X., Yang, J., Gao, F., ... & Yin, S. (2017). Malicious Bitcoin Transaction Tracing Using Incidence Relation Clustering. In International Conference on Mobile Networks and Management (pp. 313-323). Springer, Cham.



# SPAM İÇERİKLİ E-POSTALARIN TESPİTİ İÇİN BİR METİN MADENCİLİĞİ UYGULAMASI: TERİMLERİN GAMA İLİŞKİ KATSAYISINA DAYALI POLARİZASYONU

\*\*\*

## A TEXT MINING APPLICATION ON THE DETERMINATION OF SPAM- CONTENTED E-MAIL: POLARIZATION OF TERMS BASED ON THE GAMA RELATIONSHIP COEFFICIENT

Ahmet YÜCEL\*

Meltem KESKİN KÖYLÜ\*\*

### Öz

Teknolojinin gelişimi, iletişimin düzey ve şeklini de değiştirmiştir. İki nokta arası kapalı devre iletişim (telefon, mektup, telgraf, vb.) modellerinin yerini daha çok, tek noktadan tüm dünyaya açılan (Facebook, Twitter, Instagram, vb.) iletişim modelleri almıştır. Bu durum iletişimin sınırlarını kişisel olarak belirlememizi imkânsız hale getirirken, gizlemesi mümkün olmayan (E-mail, Whatsapp numarası, vb.) birçok kişisel iletişim yolunu da dünyaya açık hale getirmektedir. Basit bir e-mail yoluyla, bilgisayarda kayıtlı özel verilerin istenmeyen kişilerin eline geçmesi gibi, mevcut durum birçok risk taşımaktadır. Buna engel olmak amacıyla birçok virüs yazılımı geliştirilmekte ve elektronik ortamda karşılaşılan riskli unsurların tespitinde yardımcı olmaktadır. Ancak bazı riskli unsurlar virüs formatından uzak, normal bir metin olarak karşımıza çıkmaktadır. Bu tarz durumlarda ilgili metnin içerik olarak incelenip, riskli olup olmadığına karar vermek gerekmektedir. Bu çalışmada, istenen ve istenmeyen içeriğe sahip e-postaların bir metin madenciliği algoritması ile tespit edilip sınıflandırılması işlemi yapılmaktadır. Bu amaçla, gama ilişkisi katsayısına dayalı kompozit bir polarite değişkeni oluşturulmuş ve bu değişken üzerine genelleştirilmiş lineer modeller kurulmuştur. Modellerin sınıflandırma başarısı ortalama ise yaklaşık % 81,2'dir.

**Anahtar Kelimeler:** Metin Madenciliği, Veri Madenciliği, Genelleştirilmiş Lineer Model, Polarite, Gama İlişki Katsayısı, Sınıflandırma, İletişim, İstenmeyen İçerik.

### Abstract

The development of technology has also changed the level and form of communication. Two-ended closed-circuit communication (telephone, letter, telegraph, etc.) models have been replaced by communication models that are originated from a single point and opens to the world (Facebook, Twitter, Instagram, etc.). While this makes it impossible for us to determine the limits of communication personally, it also makes a lot of personal communication paths that cannot be hidden (E-mail, Whatsapp number, etc.). The current situation carries many risks, such as by a simple e-mail, that private data stored on the computer gets into the hands of undesirable people. In order to prevent this, many virus software is being developed and it helps to detect the risky elements encountered in electronic environment. However, some risky elements appear as a normal text rather than a virus format. In such cases it is necessary to examine the relevant text as content and decide whether it is risky or not. In this study, e-mails with spam and ham content are determined and classified by a text mining algorithm. For this purpose, a composite polarity variable based on the gamma relationship coefficient was created and generalized linear models were established on this variable. The average classification success of the models is approximately 81.2%.

**Keywords:** Text Mining, Data Mining, Generalized Linear Model, Polarity, Gamma Relationship Coefficient, Classification, Communication, Spam Content.

\* Dr. Öğr. Üyesi Ankara Yıldırım Beyazıt Üniversitesi, Şereflikoçhisar Uygulamalı Bilimler Fakültesi, Bankacılık ve Finans Bölümü, [ayucel@ybu.edu.tr](mailto:ayucel@ybu.edu.tr)  
ORCID: 0000-0002-2364-9449

\*\* Dr. Öğr. Üyesi Ankara Yıldırım Beyazıt Üniversitesi, Şereflikoçhisar Uygulamalı Bilimler Fakültesi, Uluslararası Ticaret ve Lojistik Bölümü, [mkeskinkoylu@ybu.edu.tr](mailto:mkeskinkoylu@ybu.edu.tr)  
ORCID: 0000-0002-8536-4940

## 1. GİRİŞ

En sık kullanılan iletişim aracının ilk sırasında yer alan e-postalar, kişisel iletişimden, ticarete, eğitimden sağlığa hayatın her alanında kullanılmaktadır. Elektronik iletişim araçları kullanıcıları, internet bağlantısı sayesinde mekân ve zamandan bağımsız olarak iletişim kanalları kullanılabilir. Bu durum kullanıcılarına pek çok avantaj sağlarken dezavantajlarla da karşı karşıya kalmasına neden olabilmektedir.

Haberleşmede pek çok sosyal iletişim araçları kullanmakla birlikte yaygın olarak kullanılan bir diğer kişisel iletişim aracı e-postalardır. Bu postalar hızlı, maliyetsiz ve pek çok kişiye aynı anda ulaşabilmektedir. Bu nedenledir ki dünyada en yaygın kullanılan iletişim araçlarından biridir. İnternet kullanıcısının sahip olduğu e-posta adresleri, kişiye özel görülmesine rağmen dünyaya açık ve kolayca ulaşılabilir bilgilerdir. Yaygın olarak kullanılan e-posta ağları, her türlü saldırıya açık olabilmektedir. Kötü amaçlı kişiler, kötü yazılımlı içerikleri kullanma yoluyla bilgisayar kullanıcısının güvenliğini tehdit edebilmektedir. Bu kötü yazılımlara “spam” denmektedir. Bilgisayar kullanıcıları için temel problemlerden birini spam oluşturmaktadır. Kapsamları ve nitelikleri sürekli gelişen spam’ları belirlemek için yöntemler geliştirilmektedir. Ancak geliştirilen yöntemlerde zaman zaman yetersiz kalabilmektedir. Kimi kötücül yazılımlar virüs içerikleri taşımadığı için tehlikesiz gibi algılanabilmektedirler. Bu aşamada normal metin içeriği barındıran yazılımın içeriğinin araştırılması önemli olmaktadır. Bu nedenle kullanıcıların e-posta adreslerine düşen postaların istenmeyen içerikler barındırıp barındırmadığı da postanın bizzat kendisinin istenmeyen mi yoksa istenen e-posta olduğunun tespiti önemli olmaktadır.

Bu çalışmada, metin madenciliği algoritması aracılığı ile önceden öngörülemeyen bilgiye ulaşmanın ve elde edilen sonuçları karar aşamasında kullanmanın çok aşamalı süreç yönetilmiştir. Çalışmada, toplamı oluşturan 5574 adet e-mail verileri “Public Domain Dedication” lisansına sahip olup paylaşım platformu olan Kaggle.com’dan sağlanmıştır. Elde edilen veriler incelendikten sonra 740’ı spam ve 740’ı spam olmayan e-mail olmak üzere 1480 e-mail değerlendirilmiştir. Gama ilişki katsayısı ile değişken oluşturulmuş ve elde edilen değişkenler üzerinde genelleştirilmiş lineer model kullanılmıştır. Gama parametresine dayalı polar değişkenin modellerde sağladığı başarı oranı en düşük %79 ve en yüksek %84 arasında oluşmaktadır. Çalışmada kurgulanan modelin sınıflandırma başarısı ortalama %81,2 ile en düşük ve en yüksek değer arasında yer almayı başarmıştır.

## 2. LİTERATÜR VE İLGİLİ ÇALIŞMALAR

Sosyal ağları kullanarak haberleşme dünyanın her yerinde yaygın olarak kullanılan yöntemdir. İlk elektronik haberleşme sistemi olarak görülen Elektronik Enformasyon Takas Sistemi New Jersey Teknoloji Enstitüsünde grubun kendi arasında e-posta gönderebilmesi ile 1978 yılında başlamıştır (Hambrick, 2012). Günümüzde ise tüm bilgisayar kullanıcılarının e-posta hizmeti ücretsiz olarak hizmet sağlayıcılarından alarak kullanılmaktadır. İşletmeler pazarlama stratejileri dâhilinde müşterilerinin e-posta adreslerini sahip olmak için gerekli çabaları göstermektedirler. Böylece elde edilen e-posta adresleri diğer pek çok işletme ile de paylaşılması mümkün olabilmektedir. Sonuç olarak e-posta kullanıcısı istenmeyen e-postalara maruz kalabilmektedir. Bu postaların içeriklerinde güvenliği tehdit edebilecek içeriklerde barındırma olasılıkları yüksektir. İşte böyle postalara “istenmeyen posta” (spam, junk e-mail) denilmektedir (Richardson, (2017)’den aktaran, Şahin, 2018: 1). Cobb (2003) ise istenmeyen postaların ekonomiye getirdiği yükleri spam ekonomisi adlı çalışmasında örnekleri ile açıklamıştır. Renukavd (2017) tarihli çalışmalarında istenmeyen postaların sınıflama algoritmaları üzerinde durmuşlar ve 1990’lü yıllardan başlayarak istenmeyen postaların hızla yayıldığını ve kullanıcılara büyük zararları olduğunu vurgulamışlardır.

Spam belirlemek için Integro, Pajek, ReDites, SybilRank gibi pek çok harici virüs yazılım yöntemleri bulunmaktadır. Ayrıca spam virüs tespit etmek amacı ile yapılmış çalışmalara literatürde rastlanmaktadır. Bunlardan biri Çıtlak, 2018 yılında yaptığı çalışmadır. Çalışmada, spam hesap tespit modelinin veri kümesinde ve canlı sosyal ağlar üzerindeki sorgulanmıştır. Sosyal ağlara

yönelik öğrenmeye dayalı bir spam hesap tespit modeli ile spam hesapların Twitter üzerinde tespit edilmesine yönelik çalışma yapılmış. Spam hesaplamaları çalışmada, Link analizi, makine öğrenmesi ve metin analizi yöntemleri kullanılmıştır (Çıtlak, 2018). Zararlı yazılımları erken uyarı yöntemlerinden biri olan “Honeypot” ağı ile uyarı sistemi 2004 yılında Dagon vd. leri tarafından araştırılmıştır (Dagon vd.,2004). Bu çalışmadan sonrada pek çok farklı yöntemle spam belirlemede sistemleri geliştirilmiştir.

Metin Madenciliği, verilerin açık parçası olmayan büyük çaplı verilerin içinden özel unsurları açığa çıkararak gelecekle ilgili tahminleri bilgisayar programları aracılığı ile açığa çıkarma sürecidir (Kalikov, 2006). Sarıkoz, yaptığı çalışmasında Veri Madenciliğinin temellerinin 1950’li yıllardaki yapay zekâ ve makine öğrenimi çalışmalarına dayandığını ama asıl pratikte kullanımının 1990’lı yıllara dayandığını vurgulamıştır (Sarıkoz, 2010. 4-5).Yirmi birinci yüzyıla gelindiğinde sanayiden finansa, sağlıktan eğitime kadar geniş çalışma alanı bulan Veri Madenciliği hem yurt içi hem de yurt dışı pek çok çalışmada kullanılmıştır. Verin madenciliği bir dizi süreçten oluşur. Bu süreçler; problemi belirleme, verileri hazırlama, madencilik algoritmalarını uygulama, modeli kullanma ve sonuçları izleme değerlendirme aşamalarını kapsar (Shearer, 2000). Veri madenciliği, bankacılık, pazarlama, sigortacılık ve sağlık gibi sektörler başta olmak üzere pek çok alanda kullanılmaktadır (Akman, 2010). Sarıkoz, spam filtrelemek için veri madenciliğinin kullanılması ile ilgili çalışmasında, kullanıcılara ve e-posta hizmeti servis sağlayıcılarına spam mesajlardan korunmak için görevler düşüğüne çalışmasında yer vermiştir. Yapay sinir ağları ve kümeleme metotları ile veri setleri üzerinde deneme yaparak ‘spam’ ve ‘spam olmayan’ e-postaların belirlemiştir (Sarıkoz, 2010).

### 3. YÖNTEM VE UYGULAMA

#### 3.1. Gama ( $\gamma$ ) İlişki katsayısı

Gama katsayısı, her iki değişken de sıralama (*ordinal*) düzeyinde ölçüldüğünde, derlenen istatistiksel verileri tanımlamak için kullanılır. Katsayının hesaplanması için derlenen veriler ilk önce bir çapraz tablo formatında gösterilmesi gerekir. Katsayı hesabı aşağıdaki formüle göre yapılır.

$$\gamma = \frac{U-T}{U+T} \quad (1)$$

burada U değeri, çapraz tablodaki her gözlemsel frekansın, frekansın kendisinin sağ alt kısmındaki gözlemlerdeki frekansların toplamı ile çarpılmasıyla elde edilir. Benzer şekilde T değeri, çapraz tablodaki her bir gözlemsel frekansın, frekansın kendisinin sol alt kısmındaki gözlemlerdeki frekansların toplamı ile çarpılmasıyla elde edilir.

Gama parametresi -1 ile 1 arasında değişir. -1 ve +1 her biri mükemmel ilişkileri gösterir ve 0 hiçbir ilişki olmadığı anlamına gelir. Parametre değeri 0'dan uzaklaştıkça, ilişkinin gücü artar. Parametre -1 ile 0 arasındaysa, ilgili değişkenler zıt yönlü bir ilişkiye sahiptir ve benzer şekilde parametre 0 ile +1 arasındaysa, değişkenler aynı yönlü bir ilişkiye sahiptir denir. Katsayıya dayanan hipotezi test etmek için klasik z testi uygulanır.

$$z = \frac{\gamma - E(\gamma)}{\sqrt{\frac{n(1-\gamma^2)}{U+T}}} \sim N(0,1) \quad (2)$$

Burada  $E(\gamma) = 0$ (başlangıç hipotezine bağlı olarak) beklenen değerdir ve parametrenin varyansı  $V(\gamma) = \frac{n(1-\gamma^2)}{U+T}$  ‘dır. Ayrıca belirlenen anlamlılık düzeyi  $\alpha = 0.05$  ve istatistiksel olarak anlamlı ilişkilerin ilgili tablolarda kırmızı ile vurgulandığı görülmektedir (Ünver vd., 2016: 288-289).

### 3.2. Polarite ve Matematiksel İfadesi

Polarite, kelime anlamı olarak zıt kutuplu olma halidir. Veri madenciliği alanında ise polarite, veri içinde yer alan iki farklı eğilim olarak ifade edilebilir. Bu çalışmada polaritenin birinci eğilimini ‘spam’, ikinci eğilimini ise ‘not spam’ olarak isimlendiriyoruz. Polaritenin matematiksel hesabı kontrol grubuna dayalı olarak yapılmaktadır ve elde edilen matematiksel değerler test grubu üzerinde uygulanmaktadır.

Buna göre, kontrol grubunda her bir terim ile bağımlı değişken arasındaki ilişki düzeyi gama katsayısı olarak hesaplanır. Burada,  $m$  satır (e-mail/doküman) sayısı ve  $n$  sütun (terim/değişken) sayısı olmak üzere,

$$G_{n \times 1} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} \quad (3)$$

öyle ki,  $g_i$ . ( $i = 1, 2, \dots, n$ ) terimin bağımlı değişkenle olan gama katsayısıdır.

$$A_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (4)$$

ilgili metin veriden elde edilmiş terim-doküman ikili (*binary*) frekans dağılım tablosu olsun. Buna göre,

$$A_{m \times n} \cdot G_{n \times 1} = \begin{bmatrix} a_{11}\gamma_1 + a_{12}\gamma_2 + \dots + a_{1n}\gamma_n \\ a_{21}\gamma_1 + a_{22}\gamma_2 + \dots + a_{2n}\gamma_n \\ \vdots \\ a_{m1}\gamma_1 + a_{m2}\gamma_2 + \dots + a_{mn}\gamma_n \end{bmatrix}_{m \times 1} \quad (5)$$

$A_{m \times n} \cdot G_{n \times 1}$  matris çarpımı  $m \times 1$  boyutunda yeni bir matris oluşturuyor öyle ki  $j$ .sattır, $j$ .dokümanın polariteye dayalı toplam gama değerini vermektedir. Bu durumda,

$$P_j = \frac{1}{w_j} \left( \sum_{i=1}^n \sum_{j=1}^m a_{ij} g_i \right) \quad (6)$$

öyle ki  $w_j$ . ( $j = 1, 2, \dots, m$ ) dokümanda seçilen terim sayısıdır.

### 3.3. Genelleştirilmiş Lineer Model (GLM)

$Y$  bağımlı değişken ve  $X_1, X_2, \dots, X_n$  rasgele değişkenler olmak üzere,  $P(Y | X_1, X_2, \dots, X_n)$  şartlı olasılık dağılımına bağlı genelleştirilmiş lineer model, şu şekilde ifade edilebilir:

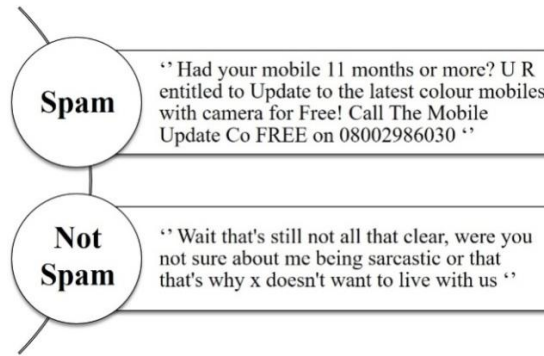
- 1-  $X_i$ ’lerin  $Y$  üzerindeki etkileri  $g$  fonksiyonu ile ifade edilir öyleki  $gX_i$ ’lerden oluşan lineer bir kombinasyondur.
- 2- Model bir link fonksiyonu üzerine kurulur.
- 3-  $l: g \rightarrow E(Y) = \mu$  tersinir bir fonksiyon olmak üzere,

$$g = \beta_0 + \sum_{i=1}^n \beta_i X_i \quad (7)$$

ifadesi bir genelleştirilmiş lineer modeldir öyle ki  $g = l(\mu)$  link fonksiyonudur (Levy, 2012: 107-108).

### 3.4. Veri

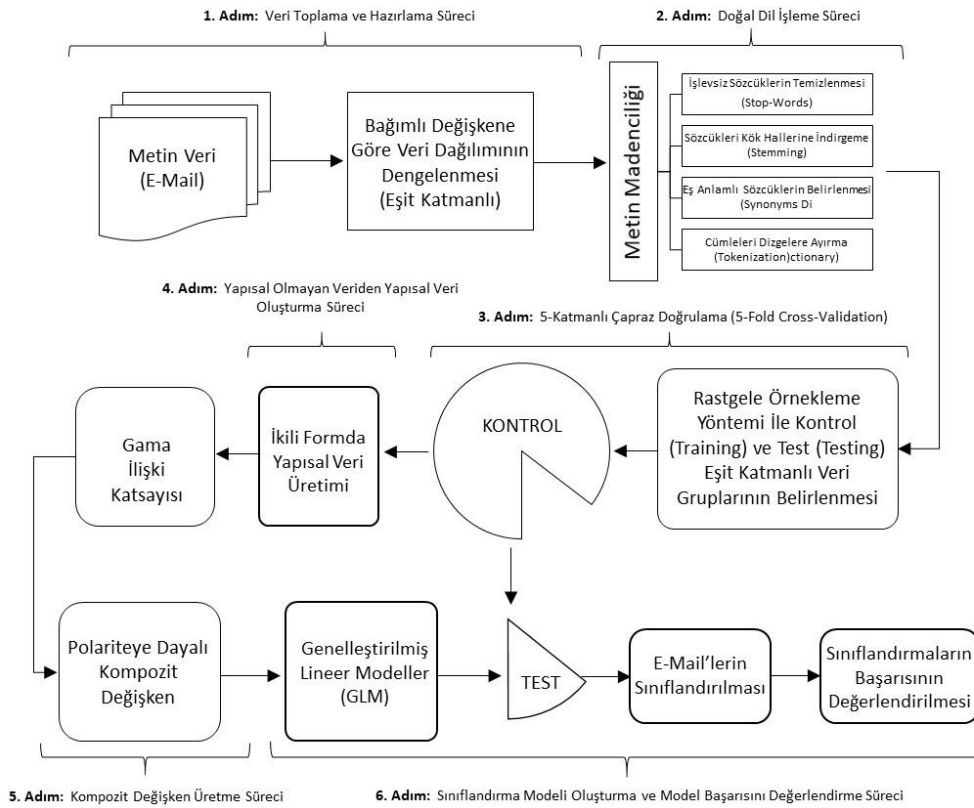
Çalışma için gerekli olan veri, dünya çapında bir veri paylaşım platformu olan Kaggle.com'dan temin edilmiştir. Veri CC0 1.0 Universal (CC0 1.0) "Public Domain Dedication" lisansına sahiptir. Verinin orijinal adı 'Spam\_Messages' (İstenmeyen Mesajlar) olup, orijinal halinde %13 (747 adet) 'spam' ve %87 (4827 adet) 'not spam (ham)' olmak üzere, toplam 5574 e-mail dokümanı bulunmaktadır. Veri temizleme ve dengeleme işleminden sonra, 1480 e-mail (740 'spam' ve 740 'not spam') incelemeye alınmıştır. Veri içinde yer alan e-maillerden ikisine örnek olarak Şekil 1'de paylaşılmıştır.



Şekil 1: Veride yer alan, istenen (not spam) ve istenmeyen (spam) e-postalara örnekler

### 3.5. Yöntem

Çalışmanın, veri toplamadan model oluşturmaya kadar yöntem ve uygulamalara dair genel adımlar Şekil 2'da ifade edilmiştir.



Şekil 2: Yöntem ve uygulama adımları

İfade edilen yöntem genel olarak 6 adımdan oluşmaktadır. Birinci adımda, veri depolandığı kaynaktan, web tarayıcı (web-crawling) bir algoritma yazılımı ile doğrudan veya bir veri tabanından hazır paket olarak elde edilir. Daha sonra, veri içinde yer alan, çalışmayla alakasız

kısımlar (link veya dosya uzantıları, veri kaynağına ait özel isimler, vb.) veriden temizlenir ve oluşturulacak olan modelin başarısının doğru bir şekilde değerlendirilebilmesi için verinin kategorik dağılımı bağımlı değişkenin kategorilerine göre (her kategoriye eşit sayıda doküman olacak biçimde) dengelenir. İkinci adımda, metin madenciliğinin en önemli adımı olan doğal dil işleme süreci uygulanır. Bu amaçla, veriye her hangi bir şekilde anlamsal katkı sağlamayan işlevsiz terimler (stop-words) bir liste halinde programa (StatSoft Statistica 12) yükleniyor ve metinlerden çıkarılır. Bunun yanı sıra eş anlamlı (synonyms) terimlerde listelenerek programa yüklenir. Bu noktada, online olarak hizmet veren ‘Cambridge English Dictionary’ (<https://dictionary.cambridge.org/dictionary/essential-american-english>) online sözlük aracından (Amerikan İngilizcesi kategorinde) faydalanılmıştır. Ayrıca Statistica 12 İngilizce için kelimeleri köklerine indirgeyen (stemming) ve cümleleri dizgelerine ayıran (tokenization) algoritmaları saptanmıştır. Bu algoritmalar yardımıyla metinler içinde yer alan kelime ve kelime grupları yapısal olarak ifade edilecek forma hazır hale getirilmiştir. Üçüncü adımda, veri kontrol ve test gruplarına ayrılmıştır. Burada veri rastgele ve dengeli olarak beş parçaya (%20) ayrılmıştır (5-katmanlı çapraz-doğrulama (5-fold cross-validation) öyle ki sırayla her bir parça test ve geriye kalan dört parça ise kontrol grubu olarak işlev görmektedir. Dördüncü adımda mevcut kontrol veriden ikili (binary) formda terim-doküman frekans matrisi elde edilir. Elde edilen matrizen beşinci adımda, bağımlı değişken ve terimler arasında yer alan gama ilişki düzeyleri, ilgili kısımda verilen yöntem ile hesaplanır ve daha sonra, ilgili kısımda ifade edildiği gibi polariteye dayalı kompozit değişken oluşturulmuştur. Altıncı ve son adımda, elde edilen kompozit değişkenle genelleştirilmiş lineer bir model kurularak test veri üzerine uygulanmış ve test veri üzerine uygulanan sınıflama modelinin doğruluk oranı hesaplanmıştır. Modelin doğruluk oranı, doğru sınıflama sayısının toplam birim sayısına oranıyla elde edilmiştir. Modellerin tanımsal kodlaması Tablo 1’de ve Tahmin Modeli İşaretleme Dili GLM algoritması Tablo 2’de paylaşılmıştır.

**Tablo 1:** GLM tanımsal kodlama

```
Distribution : BINOMIAL
Link function: LOGIT
Response variable: "SPAM_Category (0:NO,1:YES)"

Codes of dependentvariable
0: Primary code (1)
1: Secondary code (0)
Design Effects:
Continuous effects: "Gamma_Polarite1"
Categorical effects:

Model specification:
GLZ;
RESPONSE = "SPAM_Category(0:NO,1:YES)" (0 1);
GROUPS = none;
COVARIATE = "Gamma_Polarite1";
DESIGN = "Gamma_Polarite1";
INTERCEPT = include;
PARAM = sigma;
SDELTA = 7;
SURFACE = none;
MIXTURE = none;
SAMPLE = "Testing_Set1" ( 102.);
COUNTV = none;
MBUILD = all;
CONVERGE = 7;
MAXITER = 100;
INITIALS = none;
OFFSET = none;
OUTPUT = none;
```

**Tablo 2:** StatSoftStatistica GLM algoritması (StatSoftStatistic PMML (Predictive Model Markup Language))

```
<?xmlversion="1.0" encoding="windows-1254" ?>
<PMML version="2.0">
<Header copyright="STATISTICA Data Miner, Copyright (c) StatSoft, Inc., www.statsoft.com."/>
<Data Dictionary number Of Fields="2">
  <DataField name="SPAM_Category(0:NO,1:YES)" optype="categorical">
    <Value value="0" Numeric Value="0"/>
    <Value value="1" Numeric Value="1"/>
  </DataField>
  <DataField name="Gamma_Polarite1" optype="continuous"/>
</Data Dictionary>
<GeneralizedLinearModel
  Function Name="classification"
  Model Name="Generalized linearregression"
  Model Type="general Linear"
  Target VariableName="SPAM_Category(0:NO,1:YES)">
<Extension name="Distribution" value="binomial"/>
<Extension name="LinkFunction" value="logit"/>
<Parameter List>
  <Parameter name="p1" label="Intercept"/>
  <Parameter name="p2" label="Gamma_Polarite1"/>
</Parameter List>
<Factor List>
</Factor List>
<Covariate List>
  <Predictor name="Gamma_Polarite1"/>
</Covariate List>
<PPMatrix>
  <PPCellvalue="1" predictor Name="Gamma_Polarite1" parameter Name="p2"/>
</PPMatrix>
<Extension name="Correct Dummy Code" value="1"/>
<Extension name="Incorrect Dummy Code" value="-1"/>
<ParamMatrix>
<PCelltarget Category="0" parameter Name="p1" beta="2.39698409648957e+000"/>
<PCelltarget Category="0" parameter Name="p2" beta="-6.95747291468540e+000"/>
</Param Matrix>
  </Generalized Linear Model>
</PMML>
```

#### 4. BULGULAR

Veride toplam 1480 doküman (e-mail) bulunduğu için, her bir test veri (%20) için toplam 296 doküman, rastgele ve doküman kategorilerine göre ('spam'/'not spam'(0: hayır, 1: evet)) eşit (148 adet) şekilde dağıtılmıştır. Doğrusal modeller (GLM) binomial dağılıma sahip bir veri üzerine uygulanmıştır ve logit olasılık fonksiyonu kullanılmıştır. Her bir katman için Tablo 3'de GLM'lere ait 'Olabilirlik Oranı Tip 1Test' ve Tablo 4'de, 'Parametre Tahmin' sonuçları paylaşılmaktadır. Sonuçlar modellerin tamamının  $\alpha = 0.05$  anlamlılık düzeyinde, istatistiksel olarak anlamlı sonuç verdiğini göstermektedir.

Tablo 5'de modellerin sınıflamada doğru ve yanlış tahmin sayıları ve modellerin doğruluk oranları verilmektedir. Ek olarak, Şekil 3'de modellerin doğruluk oranları sütun grafiğiyle gösterilmiştir. Tabloda yer alan 'Eksik' sütunu kontrol veriden test veriye geçişteki bilgi kaybının sayısal bir ifadesidir. Yöntem kısmında da ifade edildiği üzere, modeller oluşturulmadan önce, veri %80 kontrol ve %20 test grubu olarak iki parçaya ayrılmıştır. Daha sonra kontrol grubunda yer alan metinlerden oluşturulan sözlük ve terimlere ait polar oran ile test grubunda yer alan metinler yapısal hale getirilmiştir. Bu süreçte, kontrol grubu içinde yer alan ancak test grupta bulunmayan veya tam tersi, kontrol grupta bulunmayıp test grupta bulunan bilgiler polariteye dayalı kompozit değişken hesabına dâhil edilmemiştir. Bu durum, test grubunda yer alan bazı dokümanların mevcut bilgiyle anlamlandırılmasını imkânsız kılıyor. Modellerin doğruluk oranları, 'eksik' tahminlerin de yanlış kabul edilerek hesaplanmıştır.

Gama parametresine dayalı polar değişkenin modellerde sağladığı başarı oranı en düşük %79 ve en yüksek %84 arasında değişmektedir. Sonuçlar, modellerin tamamıyla otomatik süreçte üretilen değişkenlere dayandığı dikkate alındığında, oldukça başarılı bir noktadadır.

**Tablo 3:** Olabilirlik Oranı Tip 1Test (LikelihoodType 1 Test)

Modeller		Log-Likelihood	Chi-Square	P-Value
Katman 1	Intercept1	-742,752		
	Gamma_Polarite1	-269,536	946,431	0,0001
Katman 2	Intercept2	-748,114		
	Gamma_Polarite2	-284,523	927,181	0,0001
Katman 3	Intercept3	-750,898		
	Gamma_Polarite3	-281,020	939,756	0,0001
Katman 4	Intercept4	-746,238		
	Gamma_Polarite4	-282,198	928,079	0,0001
Katman 5	Intercept5	-745,010		
	Gamma_Polarite5	-296,196	897,628	0,0001

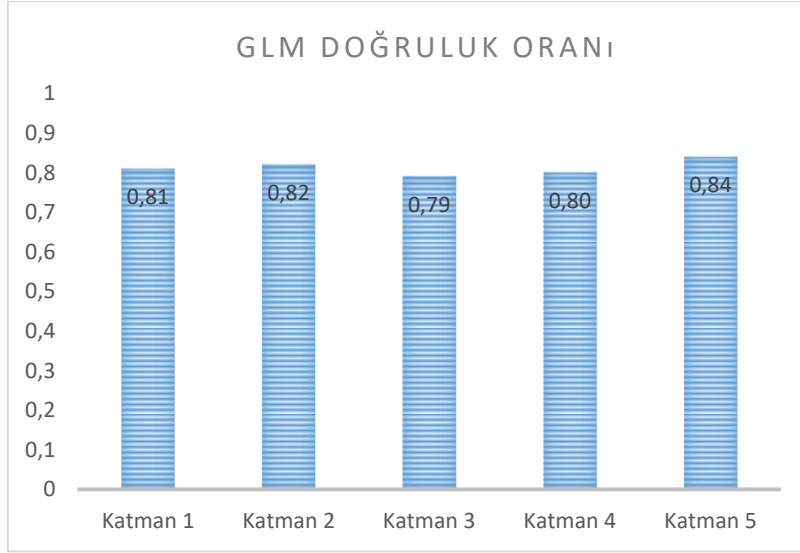
**Tablo 4:** Parametre Tahminleri

Modeller		Estimate	Standard Error	Wald Stat.	P-Value
Katman 1	Intercept1	2,396	0,202	140,143	0,0001
	Gamma_Polarite1	-6,957	0,406	292,306	0,0001
Katman 2	Intercept2	2,300	0,196	136,858	0,0001
	Gamma_Polarite2	-6,532	0,379	296,147	0,0001
Katman 3	Intercept3	2,387	0,201	140,465	0,0001
	Gamma_Polarite3	-6,698	0,391	293,474	0,0001
Katman 4	Intercept4	2,249	0,191	138,252	0,0001
	Gamma_Polarite4	-6,625	0,382	299,658	0,0001
Katman 5	Intercept5	2,308	0,196	138,261	0,0001
	Gamma_Polarite5	-6,417	0,371	297,752	0,0001

**Tablo 5:** Her bir katman için, GLM tahminlerine göre doğru veya yanlış sınıflandırma sayıları ve modellerin doğruluk oranları

Doğrusal Modeller		Tahmin 0	Tahmin 1	Eksik	GLM Doğruluk Oranı
Katman 1	Gözlem 0	105	26	17	0,81
	Gözlem 1	7	137	4	
Katman 2	Gözlem 0	109	12	27	0,82
	Gözlem 1	14	134	0	
Katman 3	Gözlem 0	104	15	29	0,79
	Gözlem 1	15	131	2	
Katman 4	Gözlem 0	104	21	23	0,80
	Gözlem 1	12	134	2	
Katman 5	Gözlem 0	110	15	23	0,84
	Gözlem 1	7	141	0	





Şekil 3: Modellerin doğruluk oranları (sütun grafiği)

## 5. SONUÇ

E-ticaretin gelişmesi ile birlikte hızlı ve maliyetsiz haberleşme ihtiyacı ortaya çıkmıştır. Elektronik posta tamda bu aşamada devreye girmiş. Sadece ticaret de değil hayatın her alanında kullanılan vazgeçilmez unsurlardan biri olmuştur. Ancak ister reklam amaçlı ister kurum ve kuruluşlardan bilinen içerikli gelen e-postaların içeriklerinde yer alan metin içeriklerinde istenmeyen kötü yazılımlar bulunabilmektedir. Zaman zaman ise spam içerik barındırmayan e-postaların spam kutusuna düşmesi. Ya da spam niteliği taşıyan e-postaların bazılarının gelen postalar kutusu içine yerleşebilmektedir. Google istenmeyen içerikli sitelerin, arama sonuçlarında sansürlenmesi için, Outlook/Hotmail/Yahoo gibi e-mail servis sağlayıcılar ise istenmeyen içerikli e-postaların tespiti için bu tarz metin madenciliği algoritmaları kullanmaktadırlar. Buna benzer dünya çapında birçok virüs yazılım firması da bu alanda çalışmalarını sürdürmektedir.

E-postaların içerisinde yer alan istenmeyen postaların belirlenmesi, spam olmayan postaların ise spam kutusuna geldiği anda tespit etmek önemli bir unsurdur. Çalışmada, hazırlanan algoritma ile bu sorunu çözmek amaçlanmıştır. Çalışmada, gama ilişki katsayısı kullanılarak ve tamamıyla otomatik (sezgisel müdahaleden uzak) üretilen değişkenlere dayalı olarak kurgulanan genelleştirilmiş lineer model ile sınıflandırma başarısı ortalama %81,2'ye ulaşılmıştır. Bu algoritma, veri boyutundan bağımsız olarak yapısal olmayan veride metin sınıflandırması yapmaya imkân sağlaması açısından önemli bir katkıdır. Böylece de e-postaların içeriklerinde bulunan bağlantı belgeleri ile spam e-postalarının ayıklanması sağlanmıştır.

## KAYNAKÇA

- Akman,M., (2010).“Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama”, Yüksek Lisans Tezi, Ankara Üniversitesi, ,Sağlık Bil. Enstitüsü, Ankara.
- Cambridge English Dictionary, (2018). <https://dictionary.cambridge.org/dictionary/essential-american-english> adresinden 27.11.2018 tarihinde erişilmiştir.
- Cobb, S., (2003). “The Economics of Spam”.ePrivacy Group, [https://www.cobbsblog.com/spam/economics\\_of\\_spam.pdf](https://www.cobbsblog.com/spam/economics_of_spam.pdf) adresinden 27.11.2018 tarihinde erişilmiştir.
- Çıtlak, O. (2018). “Sosyal Ağlara Yönelik Öğrenmeye Dayalı Bir Spam Hesap Tespit Modeli ve Uygulaması”Gazi Üniversitesi Fen Bilimleri Enstitüsü. Yayınlanmış Yüksek Lisans Tezi.

- Dagon, D., Qin, X., Gu, G., Lee, W., Grizzard, J., Levine, J. and Owen, H. (2004). Honeystat: Local worm detection using honey pots. In International Workshop on Recent Advances in Intrusion Detection, Springer, Berlin, Heidelberg, 39-58. [https://link.springer.com/chapter/10.1007/978-3-540-30143-1\\_3](https://link.springer.com/chapter/10.1007/978-3-540-30143-1_3) adresinden 27.11.2018 tarihinde erişilmiştir.
- Hambrick, M. E. (2012). Six degrees of information: Using social network analysis to explore the spread of information within sport social networks. *International Journal of Sport Communication*, 5(1), 16-34.
- Kalıkov, A., (2006), “Veri Madenciliği ve Bir E-Ticaret Uygulaması”, Yayınlanmış Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü. Ankara.
- Levy, R. (2012). "Probabilistic Models in the Study of Language" , ch. 6, November 6. [http://idiom.ucsd.edu/~rlevy/pmsl\\_textbook/book\\_draft.pdf](http://idiom.ucsd.edu/~rlevy/pmsl_textbook/book_draft.pdf) adresinden 27.11.2018 tarihinde erişilmiştir.
- Renuka, D. K., Visalakshi P. and Rajamohana, S (2017). “An Ensembled Classifier for Email Spam Classification in Hadoop Environment”. *Appl.Math*, 2017. 11(4): p. 1123-1128. <http://www.naturalspublishing.com/files/published/7ttcu333pd8l38.pdf> adresinden 27.11.2018 tarihinde erişilmiştir.
- Richardson, B.B., (2017). “Aggregating Email”. US Patent.
- Sarıkoz, K. (2010). “Veri Madenciliği Yöntemleri İle Spam Filtreleme”. Gazi Üniversitesi Bilişim Enstitüsü. Yayınlanmış Yüksek Lisans Tezi.
- Shearer, C., (2000), “The Crisp-DM Model: The New Blue print for Data Mining ” *Journal of Data Warehousing*, Cilt 5 No 4, 13-23.
- Şahin, E., (2018). “Makine Öğrenme Yöntemleri ve Kelime Kümesi Tekniği İle İstenmeyen E-Posta / E-Posta Sınıflaması”. Hacettepe Üniversitesi Yayınlanmış Yüksek Lisans Tezi. Ankara.
- Ünver, Ö., Gamgam, H. Ve Altunkaynak, B., (2016). *SPSS Uygulamalı Temel İstatistik Yöntemler Olasılık – Hipotez Testleri – Regresyon Analizi*. Seçkin Yayıncılık. Ankara.

## GENETİK ALGORİTMA VE PİKSELİZASYON YÖNTEMİ İLE MAYIN TARLASI OYUNUNUN ZORLUK SEVİYESİNİ BELİRLEME

\*\*\*

### DETERMINING THE DIFFICULTY LEVEL OF THE MINEFIELD GAME WITH GENETIC ALGORITHM AND PIXELIZATION METHOD

Cemal AKTÜRK\*

#### Öz

Genetik algoritmalar, mühendislik, işletme vb alanlardaki sayısal problemlerin çözümünde arama algoritması olarak kullanılan etkili bir optimizasyon yöntemidir. Darwin'in evrim teorisine, yani içinde bulunulan koşullara daha iyi uyum sağlayan bireylerin hayatta kalması prensibine dayanır. Bireyleri oluşturan popülasyon, genetik algoritma operatörleri uygulanarak değişikliğe uğrar ve bunun sonucunda yeni popülasyonlar oluşur. Genetik algoritma kullanımında amaç, aranan koşullara daha fazla uyum gösterecek bireylere ulaşmaktır. Yapılan çalışmada genetik algoritma ile pikselizasyon yöntemi birlikte kullanılarak mayın tarlası oyununun zorluk seviyesinin istenilen seviyede artırılıp azaltılabileceği gösterilmiştir. Ayrıca genetik algoritma yönteminin gezgin satıcı, rota belirleme, iş akışı dengeleme gibi problemler dışında oyun tasarımında da kullanılabileceği gösterilerek literatüre katkı sağlamak amaçlanmıştır.

**Anahtar Kelimeler:** Genetik Algoritma, Mayın Tarlası, Oyun Tasarımı.

#### Abstract

Genetic algorithms are an effective optimization method which is used as search algorithm for solving numerical problems in engineering, operation etc. fields. Darwin's theory of evolution is based on the principle of survival of individuals who better adhere to the circumstances in which they exist. The population of individuals is modified by applying genetic algorithm operators and as a result new populations are formed. The aim of using genetic algorithms is to reach individuals who will be more adaptable to the desired conditions. In the study, it was shown that the difficulty level of minefield game can be increased or decreased at the desired level by using genetic algorithm and pixelation method together. In addition, it is aimed to contribute to the literature by showing that genetic algorithm method can be used in game design besides traveling seller, route determination and work flow balancing.

**Keywords:** Genetic Algorithm, Minefield, Game Design.

\* Öğr.Gör. Kilis 7 Aralık Üniversitesi, TBMYO, Bilgisayar Teknolojileri Bölümü, [cakturk@kilis.edu.tr](mailto:cakturk@kilis.edu.tr)

ORCID: 0000-0003-3764-3862

## 1. GİRİŞ

Genetik algoritma, karmaşık problemlerin çözümünde oldukça işlevsel olan bir yapay zekâ yöntemidir. Genetik algoritmanın temeli Darwin'in Evrim Teorisi'ne dayanmaktadır. Darwin'in teorisine göre her zaman en iyi uyumu gösterenler hayatta kalmaktadır. Buna doğal seçim de denmektedir (Darwin, 1859). Bu teoriye göre bir canlı topluluğunda daha iyi olanın hayatta kalma şansı daha yüksektir. Bu sebeple iyiler ve genetik özelliklerini kısmen ya da tamamen onlardan alan yavrular daha yüksek ihtimalle yeni nesiller oluşturacak ve genetik bilgilerini yeni nesillere aktarma şansı bulacaklardır (Arslanoğlu, 2006). Canlılarda süregelen bu genetik süreç 1975 yılında Holland tarafından bilgisayar ortamında denenmiştir (Holland, 1975). 1989 yılında ise Holland'ın öğrencisi Goldberg genetik algoritmanın pratikte kullanılabileceğini kanıtlayarak çalışmalarını bir kitap halinde yayınlamıştır (Goldberg, 1989).

Genetik algoritma kullanarak çözüm aranan problemin kendi doğasından kaynaklanan herhangi bir bilgiye ihtiyaç duyulmaz sadece problemin parametre değerlerine ve bu parametrelerin uygunluğunu kıyaslayabilmek için bir kodlama sistemine ihtiyaç duyulur (Daban ve Özdemir, 2004). Çünkü genetik algoritma geleneksel optimizasyon yöntemlerinden farklı olarak problemin parametre kümesini değil bu parametrelerin kodlanmış biçimlerini kullanır. Birçok karmaşık yapıdaki problemlere kolayca adapte edilebilir bir yöntemdir. Çünkü doğrusal yapıdaki algoritmalarındaki kısıtlamaları yoktur. Problem ve parametreleri aranan uyum fonksiyonuna göre doğru bir şekilde kodlanabildiği taktirde, genetik algoritma diğer arama ve optimizasyon yöntemlerinden daha etkin kullanılır.

Genetik algoritmalar, parametrelerin kodlanması, başlangıç popülasyonunun oluşturulması, çoğalma, uygunluk değerlerinin hesaplanması, seçim, çaprazlama ve mutasyon işlemlerinin gerçekleştirilmesinden meydana gelir (Emel ve Taşkın, 2002).

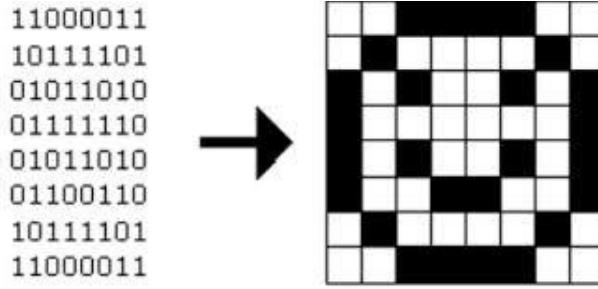
## 2. YÖNTEM

Çalışmada Microsoft Visua Studio ortamında C#.net platformu üzerinde genetik algoritma ve pikselizasyon yöntemleri kullanılarak mayın tarlası oyununun zorluk seviyesinin artırılması için bir uygulama yazılımı geliştirilmiştir.

### 2.1. Kodlama ve Pikselizasyon Yöntemi

Bir problemin genetik algoritma ile çözümünün sağlanabilmesi için, problemin ilgili parametrelerinin algoritma içerisinde kullanılabilir bir bilgiye kodlanması gerekmektedir. Bu bilgi de genelde ikili sayı sistemini oluşturan 1 ve 0 şeklindedir. Kodlama aşamasında arama uzayındaki tüm mümkün çözümler bir dizi olarak kodlanır. Her bir olası çözüm bir bireyi oluşturur ve arama uzayı içerisindeki rastgele oluşturulan bireyler genetik algoritmanın kodlanmış ham verisi olan popülasyonunu oluşturmaktadır. Bir popülasyondaki her birey, iyileştirilmesi planlanan parametreleri içeren kodlanmış bilgiye sahip olmalıdır. Bu bilgilere genetik kavram olarak kromozom denmektedir. Kromozomların uzunluğu arama uzayını tanımlar ve karmaşık yapıların çözümü daha uzun kromozom bilgilerini gerektirir.

Çalışmada tasarlanacak mayın tarlası için zorlaştırılacak veya kolaylaştırılacak oyunun optimizasyonunun yapılabilmesi için mayınlı ve boş hücrelerin ikili sayı sistemi ile kodlanması gerekmektedir. Bu çalışmada 40 satır ve 25 sütundan oluşan piksel hücreleri tasarlanmıştır. Piksel değerlerine hazırlık aşamasında rastgele 1 ve 0 değerleri atanır. Mayınlı hücreler 1, boş hücreler ise 0 olarak ifade edilir. Oluşan yapıda toplam 1000 adet birim hücre vardır. Bu hücrelerin değerlerini saklamak için 40x25 elemanlı çok boyutlu dizi kullanılmıştır. Piksellerin yapısı ve değerleri ile ilgili örnek bir yapı Şekil 1'de gösterilmektedir. Şekil 1 incelendiğinde; beyaz pikseller (mayınlı hücre) 1, siyah pikseller (boş hücre) 0 değerine karşılık gelmektedir. Bu pikseller artık dijital olarak adreslenmiştir.



Şekil 1. Piksellerin İkilik Sistemde Haritalanması (Diaz ve Sigmund, 2010)

## 2.2. Mayın Tarlası Oyunu

İlk örnekleri 1960'lı yıllarda ortaya çıkmaya başlayan mayın tarlası oyununun günümüzde kullanılan versiyonu 1983 yılında yaygınlaşmıştır. Oyunun amacı mayına denk gelmeyecek şekilde olası tüm boş hücreleri tıklayarak ilgili seviyeyi tamamlamaktır (Tuncer ve ark., 2016). Şekil 2'de oyunun ekran görüntüsü verilmiştir. Oyunun zorluk seviyesini arttırmak için 1 ve 0 olarak tanımlanan piksel hücrelerindeki 1'lerin sayısını arttırmak gerekmektedir ve tıklanan mayın olmayan bir hücre ise bu hücre içerisinde etrafındaki komşu hücrelerdeki mayınların sayısı gösterilir. Yapılan çalışmadaki amaç, bir mayın tarlası oyunu geliştirme çalışması olmayıp, ilgili oyunun matematiksel mantığından yola çıkılarak tasarımına etki edecek zorluk seviyelerini hesaplayan bir genetik algoritması uygulaması geliştirmektir.



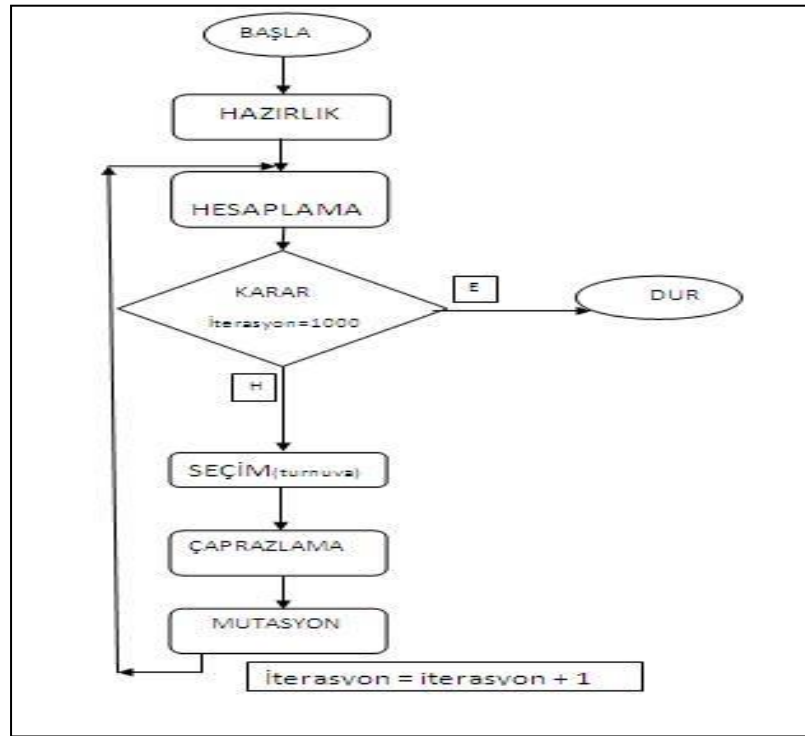
Şekil 2. Mayın Tarlası Oyunu Ekran Görüntüsü (www.rekoroyun.com)

## 2.3. Genetik Algoritma

Mayın tarlası oyununun zorluk seviyesi genetik algoritma kullanılarak arttırılmaktadır. Bu yöntemin seçilmesindeki amaç doğrusal veya doğrusal olmayan bir çok problemin çözümünde oldukça etkili bir yöntem olmasıdır (Altıparmak ve ark., 2006). Tasarlanacak mayın tarlası oyununun matematiksel ifadesi karmaşık olmadığından dolayı piksel yapısı ve karşılık gelen değerler genetik operasyonlara uğrayacak verileri sağlamaktadır. Piksel yapısındaki satır sayısı olarak kullanılan birey sayısı 40'tır. Her satır 25 hücreden oluştuğu için bireylerin kromozomları 25 gen uzunluğundadır. Çaprazlama işlemine tabi tutulacak bireyler turnuva yöntemi ile seçilmiştir (Ouedraogo, 2011). 40 satırlı yapıdan rastgele dört tanesi seçilir. Seçilen elemanlar kendi içerisinde 1 sayılarına göre karşılaştırılır. İki karşılaştırmanın kazananları ise çaprazlama işlemine tabi tutulur. Çaprazlama işlemi, 1 ile 24 arasındaki bir bitin rastgele seçilerek tek noktadan yapılmasıyla

uygulanır. Tablo 1’de çaprazlama işlemine örnek olarak seçilen nokta ve Tablo 2’de çaprazlama işlemi sonrası oluşturulan yavru kromozomların durumu gösterilmiştir. Bu çaprazlama işlemi için yalnız bir tane ayırım noktası kullanıldığından tek noktalı çaprazlama olarak isimlendirilmektedir (Arslanoğlu, 2006).

Ardından mutasyon işlemi için 40 bireyden 0.05 oranla rastgele seçim yapılır. Seçilen bireyin rassal bir biti mutasyona tabi tutularak 1 ise 0, 0 ise 1 yapılır. Mutasyon işlemi yapıldıktan sonra yavru bireydeki değişim Tablo 3’te gösterilmiştir. Bu işlemlerden sonra Şekil 3’te gösterilen algoritma yapısındaki gibi yeniden uygunluk değerleri kontrol edilir ve döngüye devam edilir. Döngünün iterasyon sayısı 1000 olduğu zaman durdurma kriteri koşulu sağlandığından işlemler tamamlanır. İyileştirmenin amaç fonksiyonu yani uygunluk kriteri olarak mayın tarlası piksel tasarımının her satırındaki 1’lerin sayısı kullanılmıştır. Oyunu zorlaştırmak için 1 sayısı fazla olanlar en uygun değere sahip kabul edilerek her satırdaki 1 değerlerinin sayısını arttırmak amaçlanır. Ters istenirse oyunu kolaylaştırmak için 0’ların sayısını arttırmak amaçlanabilir.



Şekil 3. Genetik Algoritma Akış Diyagramı

Tablo 1. Çaprazlama işleminin uygulandığı kromozomlar

Kromozom1	1	1	0	1	1	0	1	0
Kromozom2	0	0	1	1	0	1	0	0

Tablo 2. Çaprazlama işleminden sonra oluşan yeni yavru kromozomlar

Yavru1	0	0	1	1	1	0	1	0
Yavru2	1	1	0	1	0	1	0	0

**Tablo 3.** Mutasyon işlemi sonrasında oluşan yeni kromozom

Yavru1	0	0	1	1	1	0	1	0
Mutasyondan Sonra	0	0	1	0	1	0	1	0

Başlangıç popülasyonunu oluştururken ve genetik operasyonlardan sonra yapılan uygunluk değeri hesaplamasının C# kodları Şekil 4’te gösterilmektedir. Şekil 4’teki kodlar incelenecek olursa 40 satır ve 25 sütundan oluşan tabloda her bir satırdaki 1 (mayınlı hücre) sayısı hesaplanır ve her satırın son bitine ilgili satırdaki toplam 1 sayısı, genetik algoritmanın yapacağı iyileştirmeyi takip etmek için kaydedilir.

```
for (int i = 0; i < 40; i++)
{
    satir = "";
    sayac = 0;
    for (int j = 0; j < 25; j++)
    {
        if (stDizi[i, j] == 1)
        {
            sayac++;
        }
    }
    stDizi[i, 25] = sayac;
}
```

**Şekil 4.** Uygunluk Değerinin Hesaplanarak 1’lerin Sayısının Bulunması

Şekil 5’te turnuva seçim operatörünün uygulanması için geliştirilen C# kodları gösterilmektedir. Öncelikle 40 satır içerisinde rastgele dört eleman seçilir ve seçilen bireyler ikiye bölünür. Uygunluk kriterlerine (1 sayısına) göre karşılaştırılır. Her karşılaştırmadan galip gelen iki birey sonraki işlem olan çaprazlama operatörüne tabi tutulacaktır.

```
int[,] secilenler = new int[4, 2];
int[,] kazananlar = new int[2, 2];
for (int i = 0; i < 4; i++)
{
    rturnuva = r.Next(0, 40);
    secilenler[i, 0] = rturnuva;
    secilenler[i, 1] = stDizi[rturnuva, 25];
}
if (secilenler[0, 1] > secilenler[1, 1]) //1. karşılaştırma
{
    kazananlar[0, 0] = secilenler[0, 0]; // kazananın indis değeri
    kazananlar[0, 1] = secilenler[0, 1]; // kazananın 1 sayısı
}
else
{
    kazananlar[0, 0] = secilenler[1, 0]; // kazananın indis değeri
    kazananlar[0, 1] = secilenler[1, 1]; // kazananın 1 sayısı
}
if (secilenler[2, 1] > secilenler[3, 1]) //1. karşılaştırma
{
    kazananlar[1, 0] = secilenler[2, 0]; // kazananın indis değeri
    kazananlar[1, 1] = secilenler[2, 1]; // kazananın 1 sayısı
}
else
```

```
{  
kazanalar[1, 0] = secilenler[3, 0]; // kazananın indis değeri  
kazanalar[1, 1] = secilenler[3, 1]; // kazananın 1 sayısı  
}
```

**Şekil 5.** Turnuva Seçim İşlemi ve Kazananların Belirlenmesi

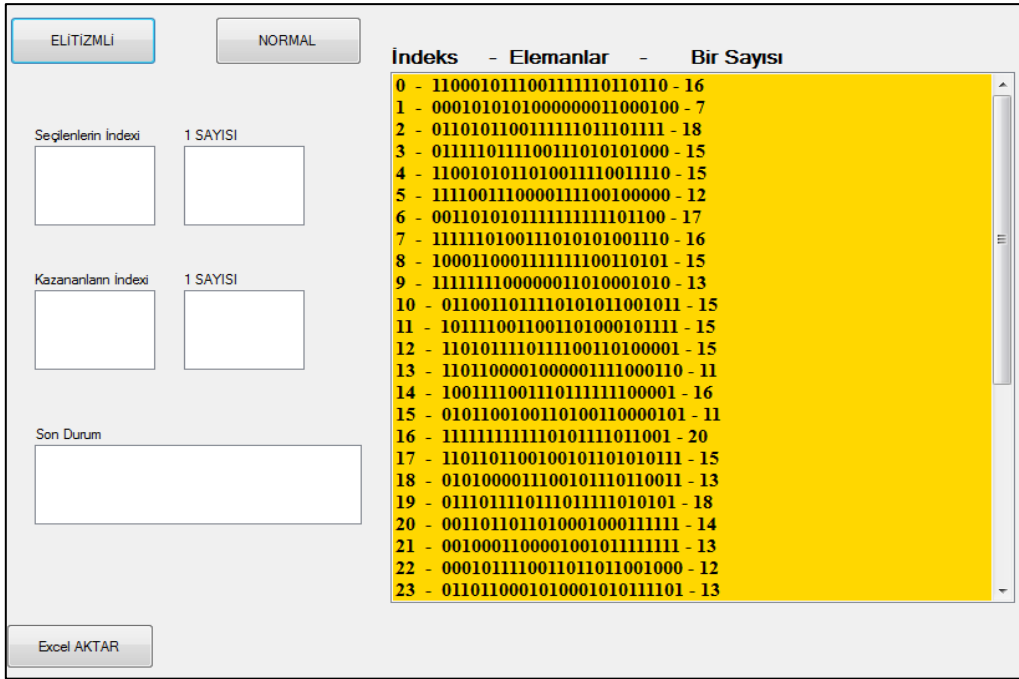
Şekil 6’da turnuvadan galip gelen bireyler için yapılan çaprazlama operatörünün ve sonrasında rastgele seçilen bir bireye ait yapılan mutasyon işlemlerinin C# kodları gösterilmektedir. Çaprazlama işlemi Şekil 6’dan incelendiğinde; öncelikle 1 ile 25 arasında rastgele bir çaprazlama noktası belirlenir. Daha sonra turnuvadan galip gelen ve çaprazlama işlemine tabi tutulan iki bireyin belirlenen çaprazlama noktasına kadar olan bit değerleri karşılıklı yer değiştirilerek çaprazlama işlemi tamamlanmış olur. Mutasyon işlemi için de 0 ile 40 arasında rastgele bir sayı oluşturularak öncelikle mutasyona tabi tutulacak birey belirlenir. Daha sonra kaçınıcı bitin mutasyon işlemine tabi tutulacağını belirlemek için 0 ile 25 arasında tekrar rastgele sayı üretilerek bu bitin değeri 1 ise 0, 0 ise 1 yapılarak mutasyon işlemi tamamlanmış ve genetik çeşitlilik sağlanmış olur. Genetik algoritma operasyonları bitince tekrar tüm bireyler için Şekil 3’te gösterildiği gibi uygunluk kriteri hesaplanır ve durdurma kriteri olarak belirtilen iterasyon sayısı veya koşulun sağlanıp sağlanmadığı kontrol edilir. Eğer koşul sağlanmamışsa işlemlere tekrar döngü içerisinde devam edilir.

```
//ÇAPRAZLAMA  
rcapraz = r.Next(1, 25);  
for (int i = 0; i <= rcapraz; i++)  
{  
temp1 = stDizi[kazanalar[0, 0], i];  
temp2 = stDizi[kazanalar[1, 0], i];  
stDizi[kazanalar[0, 0], i] = temp2;  
stDizi[kazanalar[1, 0], i] = temp1;  
}  
  
//MUTASYON  
int rmutsecim = 0;  
rmutsecim = r.Next(0, 40);  
rmut = r.Next(0, 25);  
if (stDizi[rmutsecim, rmut] == 1)  
stDizi[rmutsecim, rmut] = 0;  
else if (stDizi[rmutsecim, rmut] == 0)  
stDizi[rmutsecim, rmut] = 1;
```

**Şekil 6.** Çaprazlama ve Mutasyon İşlemleri

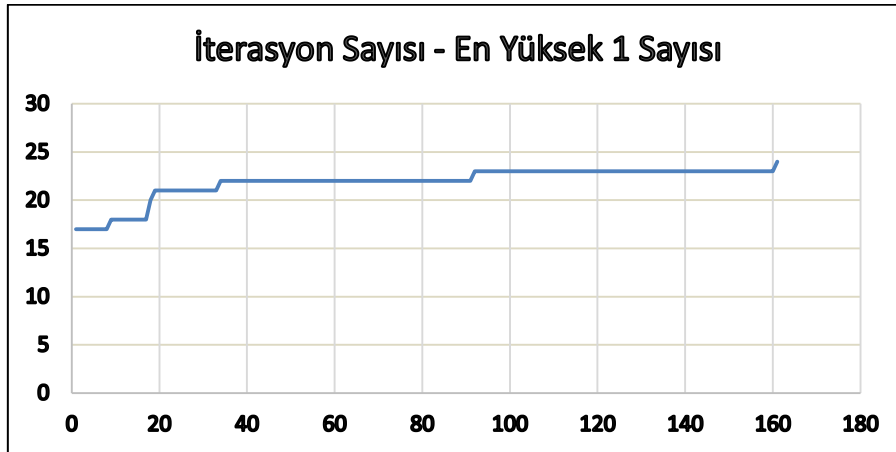
Uygulama yazılımının ekran görüntüsü Şekil 7’de gösterilmektedir. Çalıştırılan örnek bir uygulama için genetik algoritma operasyonları sonrasında bireylerin son durumu sarı renkli listbox nesnesinde indis değeri- eleman değeri – bir sayısı şeklinde gösterilmektedir. Ayrıca iyileştirmeyi takip etmek için her iterasyon sonucunda en iyi uygunluk değerine sahip olan bireyin 1 sayısı da kayıt altına alınarak excel ortamına aktarılabilir.





Şekil 7. Uygulama Yazılımı Ekran Görüntüsü

Şekil 8’de örnek olarak çalıştırılan uygulamadan elde edilen, her iterasyon sonrasındaki en iyi uygunluk değerine sahip bireyin 1 sayılarının genetik algoritma işleminde iterasyona göre değişimi grafik olarak gösterilmiştir. Grafikten anlaşılacağı üzere başlangıç durumunda en iyi uygunluk değerine sahip bireyin 17 adet 1 biti bulunmaktadır. 160 iterasyonun sonucunda birlerin sayısında iyileştirmeler yapılmış olup en iyi uygunluk değerine sahip olan bireyin 24 adet 1 biti bulunmaktadır.



Şekil 8. Örnek Bir Uygulamadaki Genetik Algoritma İyileştirmesi

### 3. SONUÇ

Genetik algoritma; içerisindeki seçim, çaprazlama ve mutasyon gibi operatörler sayesinde belirli bir veri grubunda (bireylerde) çeşitlilikler sağlamaktadır. Bu çeşitlilikler, daha iyi olanın hayatta kalma prensibine göre uygulandığında belirli bir koşulu sağlayanların hayatta kalacağı, diğerlerinin yok olacağı düşünülecek olursa etkili bir optimizasyon aracı olarak genetik algoritmanın tercih edilmesine imkan vermektedir.

Yapılan çalışmada genetik algoritma operasyonlarından kısaca bahsedilmiş olup, sonraki çalışmalara yol göstermesi amacıyla geliştirilen uygulamadaki genetik algoritma operatörlerinin kodları paylaşılmıştır. Ayrıca literatürde daha çok gezgin satıcı problemi (İlkuçar ve Çetinkaya, 2018), (Çakır ve Yılmaz, 2015), (Ertuğrul ve Özçil, 2016); çizelgeleme problemleri (Calp ve Akcayol, 2018), (Doğan ve Takçı, 2015); siber güvenlik (Özgür ve Erdem, 2018), elektromanyetik malzeme tasarımları (Çor ve Saka, 2018), (Alaybeyoğlu ve Ugranlı, 2018), (Aktürk ve ark., 2015), gibi çeşitli sayısal problemlerin çözümünde kullanılan genetik algoritmanın, matematiksel olarak ifade edilebilen oyunların senaryosunun tasarlanmasında da kullanılabileceği mayın tarlası örneği ile gösterilmiştir.

### KAYNAKÇA

- Aktürk, C., Karaaslan, M., Ozdemir, E., Ozkaner, V., Dincer, F., Bakir, M., & Ozer, Z. (2015). Chiral Metamaterial Design Using Optimized Pixelated Inclusions With Genetic Algorithm. *Optical Engineering*, 54(3), 035106.
- Alaybeyoğlu, E., & Ugranlı, F. (2018, May). A New Approach For Electronic Design Automation Of Analog Building Blocks. In 2018 26th Signal Processing And Communications Applications Conference (SIU) (Pp. 1-4). IEEE.
- Altıparmak, F., Gen, M., Lin, L., & Paksoy, T. (2006). A Genetic Algorithm Approach For Multi-Objective Optimization Of Supply Chain Networks. *Computers & Industrial Engineering*, 51(1), 196-215.
- Arslanoğlu, Y., 2006. Genetic Algorithm For Personnel Assignment Problem With Multiple Objectives. Doctoral Dissertation, Middle East Technical University, Ankara.
- Calp, M. H., & Akcayol, M. A. (2018). Optimization Of Project Scheduling Activities İn Dynamic CPM And PERT Networks Using Genetic Algorithms. *Süleyman Demirel University Journal Of Natural And Applied Sciences (SDU J Nat Appl Sci)*, 22(2), 615-627.
- Çakır, M., & Yılmaz, G. (2015, May). Traveling Salesman Problem Optimization With Parallel Genetic Algorithm. In Signal Processing And Communications Applications Conference (SIU), 2015 23th (Pp. 2557-2560). IEEE.
- Çor, İ., & Saka, B. (2018, May). Analysis And Optimization Of Wideband Radomes. In 2018 26th Signal Processing And Communications Applications Conference (SIU) (Pp. 1-4). IEEE.
- Daban, F., & Ozdemir, E.. 2004. Eğitimde Verimliliği Artıran Ders Programlarının Hazırlanması İçin Genetik Algoritma Kullanımı. *Journal Of Educational Sciences & Practices*, 3(6).
- Darwin, C. (1859). *On The Origins Of Species By Means Of Natural Selection*. London: Murray, 247, 1859.
- Diaz, A. R., & Sigmund, O. (2010). A Topology Optimization Method For Design Of Negative Permeability Metamaterials. *Structural And Multidisciplinary Optimization*, 41(2), 163-177.
- Doğan, N. Ö., & Takçı, E. (2015). Bir Tekstil İşletmesinde Simülasyon Yardımıyla Süreç İyileştirme. *Ege Academic Review*, 15(2).

- Emel, G. G., & Taşkın, Ç.. 2002. Genetik Algoritmalar Ve Uygulama Alanları. Uludağ Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi, Cilt XXI, Sayı 1, 2002, S. 129-152.
- Ertuğrul, İ., & Özçil, A. (2016). Siyasi Parti Mitinglerinin Gezgin Satıcı Problemi Yaklaşımı İle Analizi. Siyaset, Ekonomi Ve Yönetim Araştırmaları Dergisi, 4(4).
- Goldberg, D. E.. 1989. Genetic Algorithms In Search, Optimization, And Machine Learning Reading Menlo Park: Addison-Wesley, Vol. 412.
- Holland, J. H.. 1975. Adaptation In Natural And Artificial Systems: An Introductory Analysis With Applications To Biology, Control, And Artificial Intelligence. U Michigan Press.
- Ilkucar, M., & Cetinkaya, A. (2018). Optimization Of Local Travelling Route Supported With Mobile Phone And Google Maps: Case Study Of Burdur. Journal Of Mehmet Akif Ersoy University Economics And Administrative Sciences Faculty, 5(1), 64-74.
- Ouedraogo, R. O. (2011). Topology Optimization Of Metamaterials And Applications To RF Component Design. Doctoral Thesis. Michigan State University. Electrical Engineering.
- Özgür, A., & Erdem, H. (2018). Saldırı Tespit Sistemlerinde Genetik Algoritma Kullanarak Nitelik Seçimi Ve Çoklu Sınıflandırıcı Füzyonu. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 33(1).
- Tuncer, T., Avcı, D., & Avcı, E. (2016). İkili İmgeler İçin Mayın Tarlası Oyunu Tabanlı Yeni Bir Veri Gizleme Algoritması. Journal Of The Faculty Of Engineering & Architecture Of Gazi University, 31(4).