



NİCEL DERGİSİ BİLİMLER

JOURNAL of QUANTITATIVE SCIENCES

ISSN: 2667-8993

Tarih/Date: 30.06.2024

Cilt/Volume: 06

Sayı/Number:

01

Sağlık Okuryazarlığı ve Yaşam Kalitesi Arasındaki İlişki: Bir Meta Analiz Çalışması 1-18
The Relationship Between Health Literacy and Quality of Life: A Meta Analysis Study
Dr. Öğr. Üyesi Cuma FİDAN

Psikolojik Esneklik Ölçeği Türkçe Formu: Geçerlilik ve Güvenilirlik Çalışması 19-39
Psychological Flexibility Scale Turkish Form: Validity and Reliability Study
Yasemin ÇAKIR | Prof. Dr. Seher BALCI ÇELİK

Çoklu Bağlantı Olması Durumunda Veri Madenciliği Algoritmaları Performanslarının Karşılaştırılması 40-67
Comparison of Data Mining Algorithms Performances in case of Multicollinearity
Saygın DİLER | Doç. Dr. Yıldırım DEMİR

YSA Sınıflandırma Modellerinde Korelasyon-Hipotez Testi Tabanlı Filtreleme Yoluyla Girdi Seçimi 68-102
Input Selection Through Correlation-Hypothesis Testing Based Filtering in ANN Classification Models
Doç. Dr. Meryem ULUSKAN | Halil Derya ŞENLİ

Quantifying the Impact of Risk Factors on Direct Compensation Property Damage in Canadian Automobile Insurance 103-127
Kanada Otomobil Sigortasında Risk Faktörlerinin Doğrudan Tazmin Edilen Maddi Hasar Üzerindeki Etkisinin Değerlendirilmesi
Pervin BAYLAN | Prof. Dr. Neslihan DEMİREL



Sahibi / Owner

ESİMER adına Prof. Dr. Arzu ALTIN YAVUZ

Editör / Editor

Doç. Dr. Serdar NESLİHANOĞLU

Editör Yardımcıları / Co-Editors

Prof. Dr. Hasan Kıvanç AKSOY

Doç. Dr. Yakup Murat BULUT

Alan Editörler Kurulu / Field Editors Board

Prof. Dr. Zeki YILDIZ

Prof. Dr. Veysel YILMAZ

Prof. Dr. Hasan Kıvanç AKSOY

Prof. Dr. Arzu ALTIN YAVUZ

Doç. Dr. Yakup Murat BULUT

Doç. Dr. Gamze GÜVEN

Dil Editörü/ Language Editor

Prof. Dr. Hasan Kıvanç AKSOY

Prof. Dr. Arzu ALTIN YAVUZ

İstatistik Editörü/ Statistics Editor

Doç. Dr. Yakup Murat BULUT

Dergi Sekreteryası / Journal Secretariat

Abdullah ALTAY

Yayın Türü / Publication Type

Yerel Süreli

Yayın Tarihi / Release Date

Haziran 2024

Nicel Bilimler Dergisi (NBD) hakemli bir dergidir.

Haziran ve Aralık aylarında olmak üzere yılda iki kez yayınlanır.

Dergide yayınlanan yazıların, her türlü bilimsel, imlâ ve hukuki sorumlulukları yazarlarına aittir. Kaynak gösterilmeden alıntı yapılamaz.

Journal of Quantitative Sciences is a peer reviewed journal.

It is published twice a year, in June and December.

All scientific, spelling and legal responsibilities of the articles published in the journal belong to authors. Citation cannot be made without referencing.

Dergi Yazışma Adresi / Journal Correspondence Address

Eskişehir Osmangazi Üniversitesi

Meşelik Kampüsü Fen-Edebiyat Fakültesi F1 Blok 3. Kat ESİMER 26480 ESKİŞEHİR

Esimer: <http://esimer.ogu.edu.tr>

Dergi: <https://dergipark.org.tr/nicel>

Dergi e-posta: ndb@ogu.edu.tr

Yayın Kurulu/ Editorial Board

Unvan / Ad-Soyad Title / Name-Surname	Üniversite / Kurum University / Institution	Ülke Country
Prof. Dr. Abdullah YALAMAN	Eskişehir Osmangazi Üniversitesi	Türkiye
Prof. Dr. Fikret ER	Anadolu Üniversitesi	Türkiye
Prof. Dr. Menderes TARCAN	Eskişehir Osmangazi Üniversitesi	Türkiye
Prof. Dr. Murat Caner TESTİK	Hacettepe Üniversitesi	Türkiye
Prof. Dr. Sinan SARAÇLI	Balıkesir Üniversitesi	Türkiye
Prof. Dr. Cengiz BAL	Eskişehir Osmangazi Üniversitesi	Türkiye
Prof. Dr. Kemal ÖZKAN	Eskişehir Osmangazi Üniversitesi	Türkiye
Prof. Dr. Mustafa Çağatay KORKMAZ	Artvin Çoruh Üniversitesi	Türkiye
Prof. Dr. Şengül CANGÜR	Düzce Üniversitesi	Türkiye
Doç. Dr. Hüseyin GÜRBÜZ	Anadolu Üniversitesi	Türkiye
Doç. Dr. Gül İNAN	İstanbul Teknik Üniversitesi	Türkiye
Dr. Öğr. Üyesi Rifat Aykut ARAPOĞLU	Eskişehir Osmangazi Üniversitesi	Türkiye

*Yayın kurulu liste ünevana göre alfabetik olarak sıralanmıştır.

Bilim/Danışma Kurulu / Science/Advisory Board

Ünvan / Ad-Soyad Title / Name-Surname	Üniversite / Kurum University / Institution	Ülke Country
Prof. Dr. Cihan SEÇİLMİŞ	Eskişehir Osmangazi Üniversitesi	Türkiye
Prof. Dr. Harun SÖNMEZ	Anadolu Üniversitesi	Türkiye
Prof. Dr. Kishore POCHAMPALLY	Southern New Hampshire University	Amerika Birleşik Devletleri
Prof. Dr. Michael GREENACRE	Universitat Pompeu Fabra	İspanya
Prof. Dr. Muhammad NAVEED	The Shaheed Zulfikar Ali Bhutto Institute of Science and Technology	Pakistan
Prof. Dr. Hüseyin ANILAN	Eskişehir Osmangazi Üniversitesi	Türkiye
Doç. Dr. Serhan YAVUZ	Eskişehir Osmangazi Üniversitesi	Türkiye
Doç. Dr. Zeliha KAYGISIZ	Eskişehir Osmangazi Üniversitesi	Türkiye
Dr. Öğr. Üyesi Angelos MARKOS	Democritus University of Thrace	Yunanistan
Dr. Öğr. Üyesi Barış ERGÜL	Eskişehir Osmangazi Üniversitesi	Türkiye
Dr. Öğr. Üyesi Güven Gürkan İNAN	Heriot-Watt Üniversitesi	İngiltere
Dr. Alfonso Lodice D'ENZA	University of Napoli	İtalya

*Bilim/Danışma Kurulu liste ünevana göre alfabetik olarak sıralanmıştır.

İçindekiler / Contents				
Makale Sıra No No	Makale Başlığı Paper Title	Yazar(lar) Author(s)	Makale Türü Paper Type	Sayfa Page
1	Sağlık Okuryazarlığı ve Yaşam Kalitesi Arasındaki İlişki: Bir Meta Analiz Çalışması <i>The Relationship Between Health Literacy and Quality of Life: A Meta-Analysis Study</i>	Dr. Öğr. Üyesi Cuma Fidan	Araştırma Makalesi Research Paper	1-18
2	Psikolojik Esneklik Ölçeği Türkçe Formu: Geçerlik ve Güvenirlik Çalışması <i>Psychological Flexibility Scale Turkish Form: Validity And Reliability Study</i>	Yasemin Çakır Prof. Dr. Seher Balcı Çelik	Araştırma Makalesi Research Paper	19-39
3	Çoklu Doğrusal Bağlantı Olması Durumunda Veri Madenciliği Algoritmaları Performanslarının Karşılaştırılması <i>Comparison of Data Mining Algorithms Performances in Case of Multicollinearity</i>	Saygın Diler Doç. Dr. Yıldırım Demir	Araştırma Makalesi Research Paper	40-67
4	YSA Sınıflandırma Modellerinde Korelasyon-Hipotez Testi Tabanlı Filtreleme Yoluyla Girdi Seçimi <i>Input Selection Through Correlation-Hypothesis Testing Based Filtering in ANN Classification Models</i>	Doç. Dr. Meryem Uluskan Halil Derya Şenli	Araştırma Makalesi Research Paper	68-102
5	Quantifying the Impact of Risk Factors on Direct Compensation Property Damage in Canadian Automobile Insurance <i>Kanada Otomobil Sigortasında Risk Faktörlerinin Doğrudan Tazmin Edilen Maddi Hasar Üzerindeki Etkisinin Değerlendirilmesi</i>	Pervin Baylan Prof. Dr. Neslihan Demirel	Araştırma Makalesi Research Paper	103-127

Değerli okurlarımız,

2024 yılının Haziran sayımızla karşınızda olmaktan mutluluk duyuyoruz. Sayın Rektörümüze, editörümüze, editör yardımcılarımıza, alan editör kurulumuza, yayın ve bilim-danışma kurulumuza, çalışmaların nitelikli hale gelmesinde katkıda bulunan hakemlerimize, ilk sayımızda dergimize çalışmalarıyla katkıda bulunan bütün yazarlara ve teknik süreçte yer alan çalışanlarımıza şahsım ve ekibim adına teşekkür ediyorum.

Dergimiz 2019 yılında kurulmuş olup yayınlanan her sayımızdaki disiplinli çalışma ile daha birçok yıl akademik yazına katkı sunmaya devam edecektir. Dergimize yönelik oluşturduğunuz yoğun taleplere en iyi bir biçimde cevap vermek için çalışmaktayız. Bizden bağımsız olarak işleyen hakemlik süreci, bazı makalelerin yayın sürecinin uzamasına sebep olabilmektedir. Bunun için kıymetli yazarlarımıza özrü bir borç bilip, onların sabırlarını temenni etmekteyiz.

Yayınladığımız bu sayımızda 5 makaleye yer verdik. Titiz bir değerlendirme sürecinin ardından yayımlanan bu makalelerin kendi alanına katkı sağlayan, yenilik getiren ve topluma ışık tutan yayınlar olduğuna inanıyoruz. Özgün ve nitelikli çalışmalarını bilim dünyasına sunma hedefiyle bir sonraki sayıda buluşmak dileğiyle...

Prof. Dr. Arzu ALTIN YAVUZ
ESİMER Müdürü

Yayın Geliş Tarihi (Submitted): 23/07/2023

Yayın Kabul Tarihi (Accepted): 21/03/2024

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

Please Cite As/Atıf için:

Fidan, C. (2024), Sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişki: Bir meta analiz çalışması, *Nicel Bilimler Dergisi*, 6(1), 1-18. doi: 10.51541/nicel.1331638

SAĞLIK OKURYAZARLIĞI VE YAŞAM KALİTESİ ARASINDAKİ İLİŞKİ: BİR META ANALİZ ÇALIŞMASI

Cuma Fidan¹

ÖZ

Çalışmanın amacı, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkinin ve moderatör değişkenlerin bu ilişkiye etkisinin araştırılmasıdır. Araştırma, meta analiz yöntemi kapsamında gerçekleştirilmiştir. Literatürde yapılan çalışmalara 1 Ocak – 4 Nisan 2023 tarihleri arasında YÖK Tez Merkezi, Google Scholar, DergiPark ve TUBİTAK ULAKBİM veri tabanları kullanılarak ulaşılmıştır. Türkçe ve/veya İngilizce yayın dilinde, makale ve tez yayın türünde, tam metne ulaşılabilen, nicel araştırma yöntemi kapsamında Türkiye’de gerçekleştirilen ve sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi araştıran çalışmalar araştırma kapsamına dâhil edilmiştir. Araştırma kapsamına ulaşılan 4.906 çalışmadan 14’ü dâhil edilmiştir. Araştırmanın örneklemini 3.931 birey oluşturmaktadır. Rastgele etki modeline göre ortak etki büyüklüğü hesaplanmıştır. Sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı, pozitif ve orta dereceli bir ilişki bulunmuştur (M:0,275; %95 GA:0,193-0,352). Yayın yanlılığı test edilmemiştir. Meta regresyon analizi sonuçları, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yayın yılı değişkeninin etkilediğini (B:0,051; p<0,05), yaş ortalaması değişkeninin ise etkilemediğini göstermiştir (B:0,001; p>0,05). Sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiye etki edebilecek çeşitli değişkenlerin de göz önünde bulundurulması önerilmektedir.

Anahtar Kelimeler: Meta Analiz, Sağlık Okuryazarlığı, Yaşam Kalitesi.

¹ Sorumlu yazar, Dr. Öğr. Üyesi, Sağlık Yönetimi, Sağlık Bilimleri Fakültesi, Muş Alparslan Üniversitesi, Muş, Türkiye. ORCID ID: <https://orcid.org/0000-0002-8581-5940>

THE RELATIONSHIP BETWEEN HEALTH LITERACY AND QUALITY OF LIFE: A META-ANALYSIS STUDY

ABSTRACT

This study investigated the relationship between health literacy and quality of life and the moderator variables affecting that relationship. The research was conducted within the scope of meta-analysis method. Studies were recruited from YÖK Thesis Center, Google Scholar, DergiPark, and TUBITAK ULAKBIM databases between January 1 and April 4, 2023. The inclusion criteria were (1) being written in Turkish and/or English, (2) being an article or thesis, (3) having full-text access, (4) being a quantitative study, (5) being conducted in Türkiye, and (6) focusing on the relationship between health literacy and quality of life. The literature review yielded 4.906 studies. The sample consisted of 14 studies with a total sample of 3.931 participants. The joint effect size was calculated according to the random effect model. The results indicated a moderate positive relationship between health literacy and quality of life (M:0.275; 95% CI: 0.193-0.352). There was no publication bias. The results of the meta-regression analysis showed that the variable "year of publication" affected the relationship between health literacy and quality of life (B:0.051; $p < 0.05$), but the variable "average age" did not (B:0.001; $p > 0.05$). Researchers should consider possible variables that influence the relationship between health literacy and quality of life.

Keywords: Meta Analysis, Health Literacy, Quality of Life.

1. GİRİŞ

Sağlığın teşvikinde, bakım kalitesinin iyileştirilmesinde ve hastalıkların yönetiminde sağlık okuryazarlığı; bireylerin sağlıkları ile ilgili uygun kararları alabilmeleri için sağlıkla ilgili bilgileri edinme, anlama ve kullanma yeteneklerinde önemli bir rol oynamaktadır (Hossein Mirzaee Beni vd., 2022). Bununla birlikte sağlık okuryazarlığı; sağlık ve esenliğe ulaşmanın bir aracı, sağlık eşitsizliğini azaltabilecek bir faktör, sağlığın ve yaşam kalitesinin önemli bir belirleyicisi olarak değerlendirilmektedir (ALSharit ve Alhalal, 2022). Bu durum düşük sağlık okuryazarlığına sahip bireylerin; ilaç tedavisine düşük bağlılık, koruyucu ve önleyici sağlık hizmetlerinden daha az yararlanma, komorbiditelerin geç teşhisi ve sağlık-hastalık sürecini daha az anlama gibi sağlık riskleri sunan tutum ve davranışlar geliştirmesinden

kaynaklanmaktadır (Cangussú vd., 2022). Bu nedenle sağlık okuryazarlığı, yaşam kalitesi üzerinde önemli bir etkiye sahiptir (Aghajanloo vd., 2022). Dolayısıyla bireylerin sağlık okuryazarlığının iyileştirilmesi, bireylerin sağlık davranışlarındaki olumlu değişikliklerin teşvik edilmesine ve yaşam kalitesinin iyileştirilmesine katkı sağlamaktadır (Guo vd., 2023). Bu nedenle araştırmanın birincil amacı, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi araştırmaktır. Araştırmanın ikincil amacı ise, moderatör değişkenlerin sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiye etkisinin belirlenmesidir.

2. KAVRAMSAL ÇERÇEVE VE ARAŞTIRMA HİPOTEZLERİ

2.1. Sağlık Okuryazarlığı

Sağlık okuryazarlığı “bireyin sağlık bilgilerini ve sağlık hizmetlerini edinme, anlama ve kullanma yeteneği” olarak tanımlanmaktadır. Bununla birlikte sağlık okuryazarlığı, bireyin sağlığını en iyi şekilde yönetme yeteneğinin ve bir sağlık sisteminin hizmetlere adil erişim ve hizmetlerin kullanımını sağlama yeteneğinin anahtar bir belirleyicisidir (Nacanabo vd., 2021). Düşük sağlık okuryazarlığına sahip kişiler; önleyici taramadan daha az yararlanma, doktorların önerilerine daha az uyma, daha az uygun ilaç kullanımı, daha yüksek tıbbi harcamalar, daha yüksek hastaneye yatış oranları, acil servislere daha fazla talep ve daha yüksek ölüm riski sergileme eğilimindedirler (Lee vd., 2022). Küresel bir tehdit olarak değerlendirilen düşük sağlık okuryazarlığı ise; yaşlılarda, göçmenlerde, okuma yazma bilmeyenlerde, düşük gelirli kişilerde, ruh sağlığı durumu düşük kişilerde ve kronik hastalıkları olan kişilerde daha yaygındır (Khandehroo vd., 2022). Bu kişilerde sağlık okuryazarlığının iyileştirilmesi, sağlık hizmeti karar verme sürecine katılımın artmasına ve hastanın güçlendirilmesine katkı sağlar (Rague vd., 2022). Yüksek sağlık okuryazarlık düzeyi; bireylerin sağlık ve tedavi personelinin önerilerine uymasını sağlar, tıbbi konsültasyonların ve sağlığın teşviki ve öz bakım programlarının etkinliğini teşvik eder ve bireylerin tarama programlarına katılma arzusunu artırır (Mehralian vd., 2023). Bu nedenle bireylerin sağlık okuryazarlığı, sağlık bilgisinin etkin kullanımı üzerinde önemli bir etkiye sahiptir (Mikkelsen vd., 2022). Dolayısıyla sağlık okuryazarlığı sağlığın önemli bir belirleyicisi olarak değerlendirilmektedir (ALSharit ve Alhalal, 2022).

2.2. Yaşam Kalitesi

Yaşam kalitesi, bireylerin kendi refahlarını ve fiziksel, psikolojik ve sosyal işlevleri yerine getirme yeteneklerini öznel olarak nasıl değerlendirdiklerini ifade eder. Sağlık ve yaşam standartlarının bir göstergesi olarak da nitelendirilen yaşam kalitesi; bireyin fiziksel, fizyolojik, sosyal, mental ve genel sağlığını içinde barındırır (Zheng vd., 2018). Yaşam kalitesi; bireylerin hedeflerini, beklentilerini, standartlarını ve arzularını, fiziksel ve psikolojik durumlarını, bağımsızlık düzeylerini, sosyal ilişkilerini ve inançlarını önemli ölçüde etkiler (Mehralian vd., 2023). Hasta merkezli bakımın ve klinik sonuçların önemli bir belirleyicisi olarak da değerlendirilen yaşam kalitesi, hasta tarafından algılanılan hastalık ve tedavinin hastanın yaşamı üzerindeki etkisini gösterir (Skoumalova vd., 2022). Bununla birlikte yaşam kalitesi, tedavilerin etkinliği ve güvenliği ile ilgili tıbbi karar vermede sağlık profesyonellerine katkı sağlar (Nacanabo vd., 2021). Sağlık hizmetlerinin önemli bir sonucu olarak görülen yaşam kalitesi, tıbbi müdahalelerde ve nüfus sağlığı araştırmalarında sağlık göstergesi olarak da kullanılmaktadır (Guo vd., 2022).

2.3. Sağlık Okuryazarlığı ve Yaşam Kalitesi

Düşük sağlık okuryazarlığı genellikle düşük öz yeterlik seviyeleri, artan ölüm oranı, kötü sağlık durumu ve düşük yaşam kalitesi gibi kötü sağlık sonuçlarıyla ilişkilidir (Cevik ve Kayabek, 2022). Sağlık okuryazarlığı, okuryazarlığın bilgi ile kullanımı, bireyin endişe verici koşullarda karar vermede, hastalığı önlemede ve yaşam kalitesini iyileştirmede sağlık bilgilerine erişme, anlama, eleştirme ve kullanma yeteneği arasındaki bağlantı faktörüdür (Ataei vd., 2023). Literatürde yapılan bir meta analiz çalışmasının sonucuna göre, sağlık okuryazarlığı yaşam kalitesini etkilemektedir (Zheng vd., 2018). Bu nedenle sağlık okuryazarlığı yaşam kalitesinin önemli bir belirleyicisi olarak kabul edilmektedir (ALSharit ve Alhalal, 2022). Literatürde yapılan randomize kontrol çalışmalarında ise yeterli sağlık okuryazarlığı eğitiminin yaşam kalitesini arttırmada önemli ölçüde etkili olduğuna yönelik sonuç elde edilmiştir (Hosseini Mirzaee Beni vd., 2022; Khandehroo vd., 2022). Bununla birlikte literatürde, yaşam kalitesini iyileştirmek için sağlık okuryazarlığını içeren stratejilere ihtiyaç duyulduğu çıkarımında bulunulmuştur (Rague vd., 2022; Sertkaya vd., 2023). Bu durum, bireylerin sağlık okuryazarlık düzeylerinde yapılacak iyileştirmelerin yaşam kalitesi düzeylerinde de olumlu katkı sağlayacağını göstermektedir (Guo vd., 2023).

Literatürde yapılan çalışmalarda çoğunlukla sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak pozitif ve anlamlı bir ilişki tespit edilmiştir (Akbolat vd., 2022; Ataei vd., 2023; Bessing vd., 2022; Gaffari-Fam vd., 2020; Gao vd., 2023; Guo vd., 2022; Guo

vd., 2023; Hossein Mirzaee Beni vd., 2022; Khandehroo vd., 2022; Liu vd., 2023; Mehralian vd., 2023; Mikkelsen vd., 2022; Nacanabo vd., 2021; Şimşek Küçükkelepçe vd., 2021; Üner ve Dost, 2023; Zheng vd., 2018). Diğer yandan literatürde, sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı ve negatif bir ilişkinin olduğu sonucuna ulaşılan çalışmalar da yer almaktadır (Du vd., 2022; Skoumalova vd., 2022; Topallar, 2019). Bununla birlikte literatürde, sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı bir ilişkinin olmadığını saptayan araştırma da yer almaktadır (Aghajanloo vd., 2022). Bu durum, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi araştıran literatürdeki çalışmalarda birbirinden farklı sonuçların olduğunu göstermektedir. Bununla birlikte, uluslararası literatürde sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi araştıran bir meta analiz çalışmasına (Zheng vd., 2018) ulaşılmış olup ulusal literatürde ise herhangi bir meta analiz çalışmasına rastlanılmamıştır. Literatürde yapılan bu meta analiz çalışmasında ise sağlık okuryazarlığı ile yaşam kalitesi arasında orta derecede bir ilişkinin olduğu sonucuna ulaşılmış, ancak ulaşılan bu sonucun daha fazla kanıtla desteklenmesi gerektiği önerilmiştir (Zheng vd., 2018). Yapılan bu meta analiz çalışmasına 2005-2017 yılları arasında, makale yayın türünde ve tanımlayıcı ve kesitsel ve kohort yöntem türünde gerçekleştirilen çalışmalar dâhil edilmiş; çalışmada yayın yılı değişkeninin sağlık okuryazarlığı ve yaşam kalitesi üzerine etkisi incelenmiş, yaş ortalaması değişkeninin etkisi ise incelenmemiştir (Zheng vd., 2018). Bu meta analiz çalışmasında ise Türkçe ve/veya İngilizce yayın dilinde, makale ve tez yayın türünde, nicel araştırma yöntemi kapsamında Türkiye’de gerçekleştirilen çalışmaların araştırma kapsamına dâhil edilmesi ve yayın yılı ve yaş ortalaması moderatör değişkenlerinin sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkinin araştırılması planlanmıştır. Bu nedenle aşağıdaki hipotezler geliştirilmiştir.

H₁: Sağlık okuryazarlığı ve yaşam kalitesi arasında bir ilişki vardır.

H₂: Sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yayın yılı değişkeni etkilemektedir.

H₃: Sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yaş ortalaması değişkeni etkilemektedir.

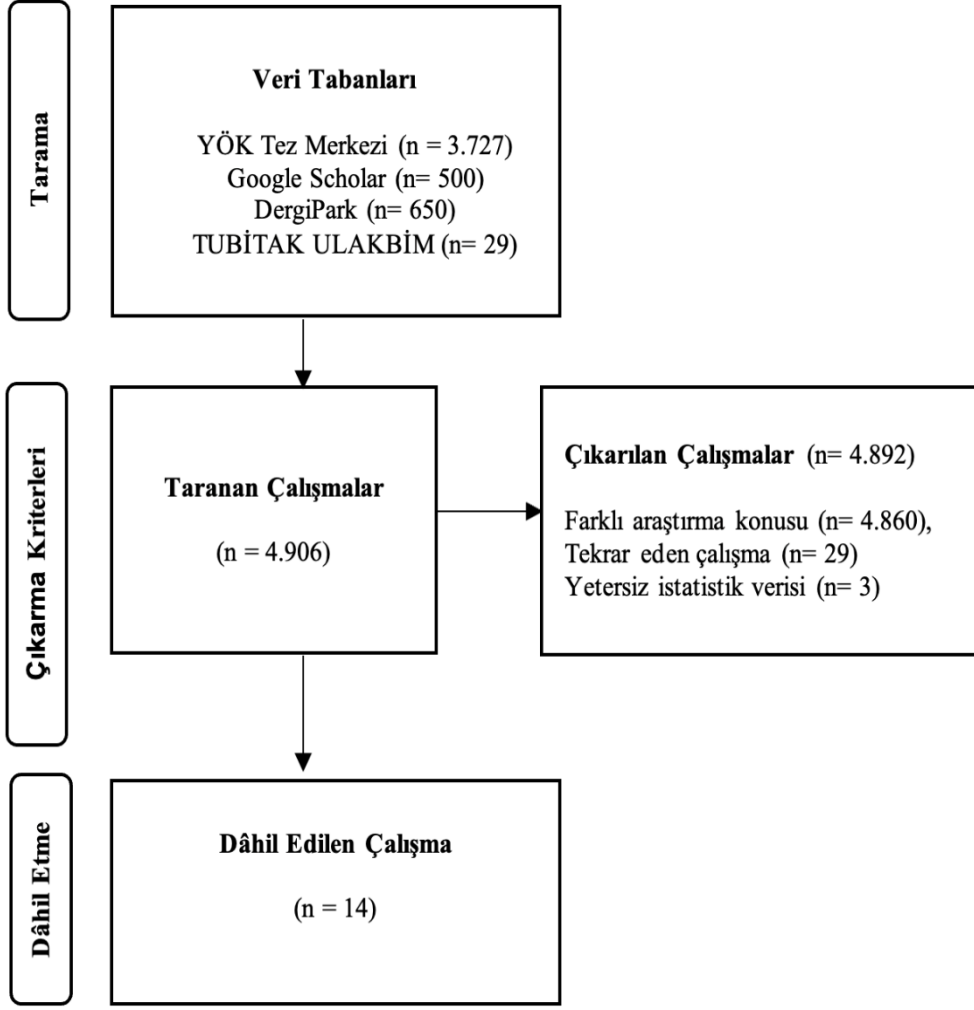
3. MATERYAL VE METOT

Çalışma, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi ve moderatör değişkenlerin bu ilişkiye etkisini araştıran meta analiz yöntemi kapsamında gerçekleştirilmiştir. Meta analiz yöntemi, belirli bir konuyla ilgili literatürde yapılan birçok bireysel çalışmanın

birbirinden farklı bulgularını ortaya koymak ve konuyla ilgili kanıta dayalı sonuçlar elde edilmek istenildiği durumlarda kullanılır (Siddaway vd., 2019).

3.1. Ekleme ve Çıkarma Kriterleri

Literatürde sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi ortaya koyan çalışmalara ulaşmak için “sağlık okuryazarlığı/health literacy, yaşam kalitesi/quality of life” olarak belirlenen Türkçe ve İngilizce anahtar kelimeler kullanılmıştır. Belirlenen anahtar kelimeler kullanılarak YÖK Tez Merkezi, Google Scholar, DergiPark ve TUBİTAK ULAKBİM veri tabanları aracılığıyla 1 Ocak – 4 Nisan 2023 tarihleri arasında tarama yapılmıştır. Çalışmada “Türkçe ve/veya İngilizce yayın dilinde, makale ve tez yayın türünde, tam metne ulaşılabilen, nicel araştırma yöntemi kapsamında Türkiye’de gerçekleştirilen ve sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi araştıran çalışmalar” olarak belirlenen kriterleri sağlayan çalışmalar araştırma kapsamına dâhil edilmiştir. Bu kriterleri sağlamayan çalışmalar araştırma kapsamına dâhil edilmemiştir. Bu süreç PRISMA (Preferred Reporting Items for Systematic Reviews and Meta Analysis) akış diyagramı kullanılarak Şekil 1’de gösterilmiştir (Page vd., 2021).



Şekil 1. PRISMA akış diyagramı

Araştırma kapsamında taranan 4.906 çalışmadan 14'ü çalışma kapsamına dâhil edilmiştir (Şekil 1). Türkiye’de konu ile ilgili ulaşılan ilk çalışmanın 2017 yılında son yapılan çalışmanın ise 2023 yılında yapılması nedeni ile çalışmaların 2017-2023 yılları arasında kapsadığı, çalışmaların makale (n= 3) ve tez yayın türünde (3 uzmanlık tezi, 7 yüksek lisans tezi ve 1 doktora tezi olmak üzere toplam 11 tez çalışması) olduğu, tanımlayıcı ve kesitsel çalışma yönteminde gerçekleştirildiği ve toplam 3.931 birey üzerinde sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkinin araştırıldığı tespit edilmiştir.

3.2. İstatistiksel Analiz

Verilerin analizinde Comprehensive Meta Analysis Versiyon 4 programı kullanılmıştır. Etki büyüklüğünün hesaplanmasında, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi incelemeye kullanılan, korelasyon etki büyüklüğü kullanılmıştır (Bowman, 2012). Yapılan

meta analizde, heterojenitenin istatistiksel olarak anlamlı ($Q: 89,469$; $sd: 13$; $p<0,001$), yüksek heterojenin ($I^2>0,75$) ve çalışmalar arası varyansın olması ($T>0,13$) nedeniyle ortak etki büyüklüğünün hesaplanmasında rastgele etki modeli kullanılmıştır (Tablo 1) (Michael vd., 2019).

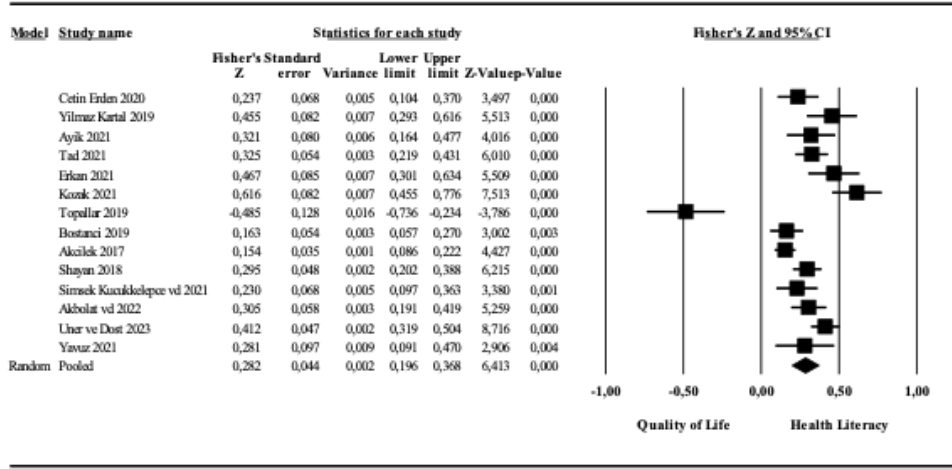
Yayın yanlılığının belirlenmesinde istatistiksel yayın yanlılığı yöntemlerinden Egger'in regresyon testi ve Duval ve Tweedie'nin kırp ve doldur yöntemleri kullanılmıştır. Egger'in regresyon testi, çalışmaların etki büyüklükleri ile standart hataları arasındaki ilişkiyi ve huni grafiğinin asimetrliliğini ölçer. Yöntemden elde edilen p değeri $>0,05$ ise yayın yanlılığının olmadığını gösterir (Egger vd., 1997) Duval ve Tweedie'nin kırp ve doldur yöntemi ise, huni grafiği asimetrik bir yapıda olduğunda eksik olan yayımlanmamış çalışmaların var olduğunu ve bu durumun yayın yanlılığına neden olduğunu varsayar. Yöntemde eksik olan yayımlanmamış çalışma sayısının analize dâhil edilen çalışma sayısından olabildiğince küçük olması istenir ve bu durum yayın yanlılığının olmadığını gösterir (Duval ve Tweedie, 2000). Sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiye etki edebilecek moderatör değişkenler yayın yılı ve yaş ortalaması değişkenleri olarak belirlenmiştir. Meta regresyon analizi ile sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkide yayın yılı ve yaş ortalaması moderatör değişkenlerinin etkisinin olup olmadığı analiz edilmiştir.

3.3. Sınırlılıklar

Araştırma; YÖK Tez Merkezi, Google Scholar, DergiPark ve TUBİTAK ULAKBİM veri tabanları aracılığıyla çalışma kapsamına dâhil edilen 14 çalışmadan elde edilen istatistiki veriler ile sınırlıdır. Sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiye etki edebilecek moderatör değişkenler yayın yılı ve yaş ortalaması değişkenleri ile sınırlıdır.

4. BULGULAR

Araştırma kapsamına dâhil edilen çalışmalara ait etki büyüklükleri Şekil 2'de gösterilmiştir. 13 çalışmada sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak pozitif ve anlamlı bir ilişki ($p<0,005$); 1 çalışmada ise sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak negatif ve anlamlı bir ilişki ($p<0,001$) vardır (Şekil 2).



Şekil 2. Orman grafiği

*p<0,005

Meta analiz sonuçları Tablo 1’de gösterilmiştir. Sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı, pozitif ve orta dereceli bir ilişki (M: 0,275; %95 GA: 0,193-0,352) vardır (H₁ hipotezi kabul) (Tablo 1).

Tablo 1. Meta analiz sonuçları

k	M	% 95 GA		Z	p	Q	sd	p	I ²	T ²
14	0,275	0,193	0,352	6,413	0,000	89,469	13	0,000	85,470	0,022

k: Çalışma sayısı, M: Rastgele etki modeli, GA: Güven aralığı, z: standart normal dağılım, Q: Cochran heterojenlik testi, sd: Serbestlik derecesi, I²: Heterojenlik miktarı, T²: Çalışmalar arası varyans.

*p<0,001

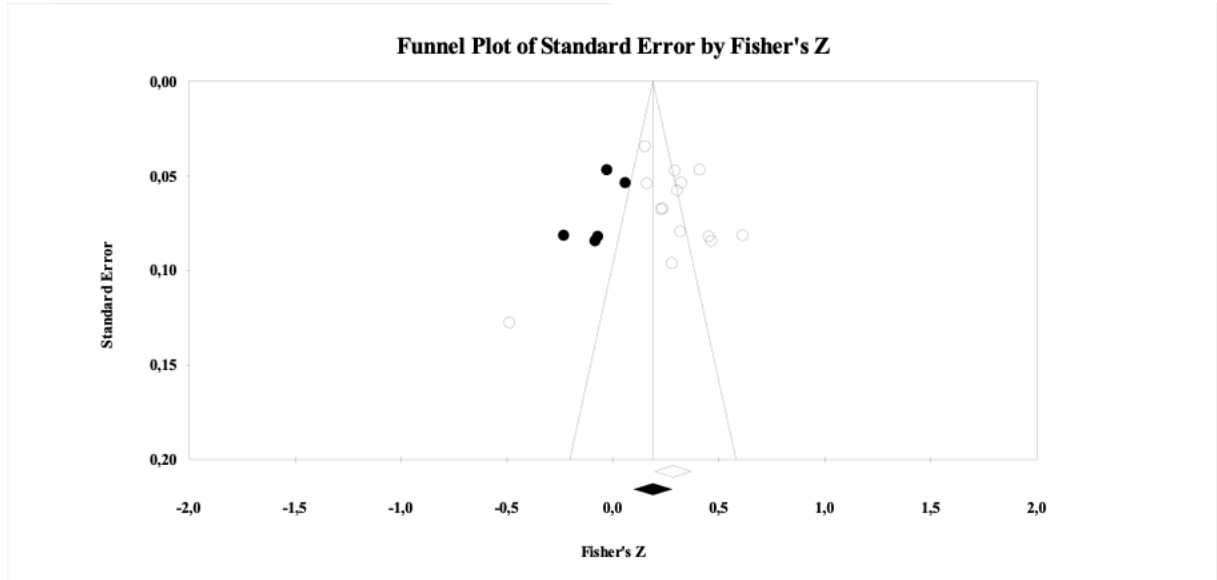
Yayın yanlılığı analiz sonuçları Tablo 2’de ve Duval ve Tweedie’nin kırp ve doldur yöntemine ait huni grafiği Şekil 3’te gösterilmiştir.

Tablo 2. Yayın yanlılığı analiz sonuçları

Yöntem Türü	Test Sonucu
Egger’ın regresyon testi	0,383*
Duval ve Tweedie’nin kırp ve doldur yöntemi	5** (M: 0,186 % 95 GA: 0,096-0,273)

*: $p>0,05$, **: Test değeri $<k$.

Tablo 2’de ve Şekil 3’te görüldüğü gibi, Egger’ın regresyon ($p>0,05$) ve Duval ve Tweedie’nin kırp ve doldur (Test değeri $<k$) yöntemlerine göre yayın yanlılığı yoktur. Duval ve Tweedie’nin kırp ve doldur yöntemine göre, yayınlanmamış çalışmalar da ($n=5$) analize dâhil edildiğinde, araştırma kapsamında ulaşılan sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı ve pozitif bir ilişkinin (M: 0,186; %95 GA: 0,096-0,273) olduğu sonucu farklılık göstermemiştir (Tablo 2 ve Şekil 3).



Şekil 3. Huni grafiği (Duval ve Tweedie’nin kırp ve doldur yöntemi)

Meta regresyon analiz sonuçları Tablo 3'te gösterilmiştir.

Tablo 3. Meta regresyon analiz sonuçları

Model	k	Değişken	B	% 95 GA		Z	p	R ²
Model 1	14	Sabit	-101,898	-199,209	-4,587	-2,05	0,04	0,19
		Yayın yılı	0,051	0,002	0,099	2,06	0,04	
Model 2	14	Sabit	0,244	-0,052	0,540	1,62	0,11	0,00
		Yaş ortalaması	0,001	-0,005	0,006	0,26	0,80	

*p<0,05

Tablo 3'te görüldüğü gibi, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yayın yılı değişkeni etkilemektedir (B: 0,051; p<0,05) (H₂ hipotezi kabul). Diğer yandan, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yaş ortalaması değişkeni etkilememektedir (B: 0,001; p>0,05) (H₃ hipotezi red).

5. TARTIŞMA VE SONUÇ

Sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi ve moderatör değişkenlerin bu ilişkiye etkisini araştıran bu meta analiz çalışmasına dâhil edilen çalışmalar; 2017-2023 yılları arasında, makale (n= 3) ve tez yayın türünde (n=11), tanımlayıcı ve kesitsel çalışma yönteminde gerçekleştirilmiş ve toplam 3.931 birey üzerinde sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişki araştırılmıştır. Literatürde yapılan bir meta analiz çalışmasında ise araştırma kapsamına dâhil edilen çalışmalar (Zheng vd., 2018); 2005-2017 yılları arasında, makale yayın türünde, tanımlayıcı ve kesitsel (n=18) ve kohort (n=5) çalışma yöntemlerinde gerçekleştirilmiş ve toplam 12.303 birey üzerinde sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişki araştırılmıştır. Literatürde yapılan meta analiz çalışmasında (Zheng vd., 2018) ve bu meta analiz çalışmasında araştırma kapsamına dâhil edilen çalışmaların çoğunlukla tanımlayıcı ve kesitsel çalışma yöntem türünde olduğu ve randomize kontrol yöntem türünde çalışmaların araştırma kapsamına dâhil edilmediği ve/veya yapılmadığı tespit edilmiştir. Bu nedenle, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkinin araştırılmasına yönelik randomize kontrol yönteminde çalışmaların yapılması önerilmektedir.

Literatürde nicel araştırma yöntemleri kapsamında yapılan çalışmalarda; sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı ve pozitif bir ilişkinin olduğu (Akbolat vd., 2022; Ataei vd., 2023; Bessing vd., 2022; Gaffari-Fam vd., 2020; Gao vd., 2023; Guo vd., 2022; Guo vd., 2023; Hossein Mirzaee Beni vd., 2022; Khandehroo vd., 2022; Liu vd., 2023; Mehralian vd., 2023; Mikkelsen vd., 2022; Nacanabo vd., 2021; Şimşek Küçükkeleş vd., 2021; Üner ve Dost, 2023), sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı ve negatif bir ilişkinin olduğu (Du vd., 2022; Skoumalova vd., 2022; Topallar, 2019) ve sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı bir ilişkinin olmadığı (Aghajanloo vd., 2022) sonuçlarına ulaşılmıştır. Bu durum sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi araştıran bu çalışmalarda birbirinden farklı sonuçların elde edildiğini göstermektedir. Ayrıca, literatürde yapılan meta analiz çalışmasında (Zheng vd., 2018); sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı, pozitif ve orta dereceli bir ilişkinin olduğu sonucuna ulaşılmıştır. Yapılan bu meta analiz çalışmasında ise; sağlık okuryazarlığı ve yaşam kalitesi arasında istatistiksel olarak anlamlı, pozitif ve orta dereceli bir ilişki saptanmıştır. Yapılan bu meta analiz çalışmasında ulaşılan bu sonuç ile literatürde yapılan meta analiz çalışmasında elde edilen sonuç birbirini desteklemektedir. Dolayısıyla literatürde yapılan meta analiz çalışması ile yapılan bu meta analiz çalışmasında elde edilen sonuçlar, sağlık okuryazarlığının yaşam kalitesi üzerinde pozitif bir etkiye sahip olduğunu göstermektedir. Bununla birlikte çalışmada, Duval ve Tweedie'nin kırp ve doldur istatistiksel yayın yanlılığı yöntemine göre, yayınlanmamış çalışmalar da analize dâhil edildiğinde de araştırma kapsamında ulaşılan sonuç değişiklik göstermemiştir. Bu durum araştırmanın geçerlilik ve güvenilirliğini pozitif yönde arttırmıştır (Duval ve Tweedie, 2000).

Egger'in regresyon ve Duval ve Tweedie'nin kırp ve doldur istatistiksel yayın yanlılığı yöntemlerine göre yayın yanlılığı tespit edilmemiştir. Bu durum araştırmanın geçerlilik ve güvenilirliğini pozitif yönde arttırmıştır (Alinaghi ve Reed, 2018). Zheng vd. (2018) tarafından yapılan meta analiz çalışmasında ise Egger'in regresyon yayın yanlılığı yöntemine göre yayın yanlılığının olmadığı test edilmiştir. Yayın yanlılığına yönelik elde edilen bu sonuç ile çalışmada ulaşılan bu sonuç birbiriyle paralellik göstermektedir.

Literatürde yapılan meta analiz çalışmasında (Zheng vd., 2018); sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yayın yılı değişkeninin etkilediği sonucuna ulaşılmıştır. Yapılan bu meta analiz çalışmasında ise; sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yayın yılı değişkeninin etkilediği saptanmıştır. Dolayısıyla literatürde yapılan meta analiz çalışması ile yapılan bu meta analiz çalışmasında elde edilen sonuçlar, yıllara göre bireylerin sağlık

okuryazarlık düzeylerindeki artışın bireylerin yaşam kalitesi düzeylerini de pozitif yönde arttırdığını göstermektedir. Diğer yandan çalışmada, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi yaş ortalaması değişkeninin bir etkisinin olmadığı sonucuna ulaşılmıştır. Bununla birlikte çalışmadaki sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişki yayın yılı ve yaş ortalaması moderatör değişkenleri ile sınırlıdır. Bu nedenle, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiye etki edebilecek diğer moderatör değişkenlere yönelik çalışmaların yapılması önerilmektedir.

Sonuç olarak, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkinin ve moderatör değişkenlerin bu ilişkiye etkisinin araştırıldığı bu meta analiz çalışmasında, sağlık okuryazarlığının yaşam kalitesi üzerinde pozitif bir etkiye sahip olduğu sonucuna ulaşılmıştır. Meta regresyon analizi sonuçları ise, yayın yılı değişkeninin sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi etkilediğini; yaş ortalaması değişkeninin ise sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiyi etkilemediğini göstermektedir. Dolayısıyla, sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişki yayın yılı ve yaş ortalaması moderatör değişkenleri ile sınırlıdır. Bu nedenle sağlık okuryazarlığı ve yaşam kalitesi arasındaki ilişkiye etki edebilecek diğer moderatör değişkenlerin de göz önünde bulundurulması önerilmektedir.

ETİK BEYAN

“Sağlık Okuryazarlığı ve Yaşam Kalitesi Arasındaki İlişki: Bir Meta Analiz Çalışması” başlıklı çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş, toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

KAYNAKÇA

- Aghajanloo, R., Nadrian, H., Baraei, B., Shirzadi, S., Sarbakhsh, P., Keshavarzian, K., Ghassab-Abdollahi, N. ve Chattu, V.K. (2022), Health literacy, treatment adherence, bone mass density and health-related quality of life among Iranian older adults with osteoporosis, *International Journal of Health Promotion and Education*.
- Akbolat, M., Amarat, M. ve Doğanyığıt, P.B. (2022), Sağlık okuryazarlığının yaşam kalitesine etkisinde özbakım gücünün aracı rolü, *Balıkesir Sağlık Bilimleri Dergisi*, 11(1), 69-75.

- Akcilek, E. (2017), *Üniversite öğrencilerinde sağlık okuryazarlığı ve yaşam kalitesinin incelenmesi*, Yüksek Lisans Tezi, İstanbul Medipol Üniversitesi Sağlık Bilimleri Enstitüsü, İstanbul.
- Alinaghi, N. ve Reed, W.R. (2018), Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work?, *Research Synthesis Methods*, 9(2), 285-311.
- ALSharit, B.A. ve Alhalal, E.A. (2022), Effects of health literacy on type 2 diabetic patients' glycemic control, self-management, and quality of life, *Saudi Med J.*, 43(5), 465-472.
- Ataei, M., Esmaelzadeh Saeieh, S., Yazdkhasti, M. ve Jashni Motlagh, A. (2023), Coping with identity threat and health literacy on the quality of life and mental health in students: Structural equation modeling, *Neuropsychopharmacology Reports*.
- Ayık, M.E. (2021), *Üniversite hastanesinin onkoloji kliniğinde tedavi gören kanser hastalarında sağlık okuryazarlığı ile yaşam kalitesi arasındaki ilişkinin incelenmesi*, Yüksek Lisans Tezi, Hasan Kalyoncu Üniversitesi Lisansüstü Eğitim Enstitüsü, Gaziantep.
- Bessing, B., van der Mei, I., Taylor, B.V., Honan, C.A., Blizzard, L. ve Claflin, S.B. (2022), Evaluating the impact of the Understanding Multiple Sclerosis online course on participant MS knowledge, health literacy, resilience, self-efficacy, quality of life, and MS symptom severity, *Multiple Sclerosis and Related Disorders*, 60, 103717.
- Bostancı, Ş. (2019), *Hipertansiyon hastalarında sağlık okuryazarlığı düzeyinin yaşam kalitesi üzerine etkisi*, Yüksek Lisans Tezi, Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü, Bursa.
- Bowman, N.A. (2012), Effect sizes and statistical methods for meta-analysis in higher education, *Research in Higher Education*, 53, 375-382.
- Cangussú, L.R., Alho, E.A.S., Silva, A.L., Fonsêca, D.V., Lopes, J.M., de Almeida Barbosa, R.H. ve Lopes, M.R. (2022), Low health literacy and quality of life in patients with systemic arterial hypertension, *Dialogues in Health*, 1, 100036.
- Cevik, C. ve Kayabek, I. (2022), Health literacy and quality of life among people in semi-urban and urban areas, *Rev Esc Enferm USP*, 56, e20210495.
- Çetin Erden, M. (2020), *Hipertansif hastalarda tedavi uyumu ve sağlık okuryazarlığı düzeylerinin yaşam kalitesi üzerine etkisi*, Uzmanlık Tezi, İzmir Katip Çelebi Üniversitesi İzmir Atatürk Eğitim ve Araştırma Hastanesi, İzmir.

- Du, S., Feng, Z., Wang, W., Tian, L. ve Wang, Y. (2022), A structural equation model linking health literacy, self efficacy and quality of life in adults with coronary heart disease, *BMC Cardiovascular Disorders*, 22.
- Duval, S. ve Tweedie, R. (2000), Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis, *Biometrics*, 56(2), 455-463.
- Egger, M., Smith, G.D., Schneider, M. ve Minder, C. (1997), Bias in meta-analysis detected by a simple, graphical test, *Bmj*, 315, 629-634.
- Erkan, R. (2021), *Üriner inkontinanslı kadınlarda sağlık okuryazarlığının ve yaşam kalitesinin değerlendirilmesi*, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Gaffari-Fam, S., Lotfi, Y., Daemi, A., Babazadeh, T., Sarbazi, E., Dargahi-Abbasabad, G. ve Abri, H. (2020), Impact of health literacy and self-care behaviors on health-related quality of life in Iranians with type 2 diabetes: a cross-sectional study, *Health Qual Life Outcomes*, 18, 357.
- Gao, Y., Yan, K., Yan, X., Xi, N., Gao, J. ve Ren, H. (2023), Correlation between health literacy and health-related quality of life in patients with diabetic peripheral neuropathy: The mediating role of self-management, *Nursing Open*, 10(5), 3164-3177.
- Guo, S., Yu, X., Naccarella, L., Armstrong, R. ve Davis, E. (2022), Health literacy and health-related quality of life in beijing adolescents: A path analysis, *Health Literacy Research and Practice*, 6(4), e300-e309.
- Guo, Y., Liu, Y., Ding, R., Yan, X., Tan, H., Wang, Y., Wang, X. ve Wang, L. (2023), A structural equation model linking health literacy, self-efficacy, and quality of life in patients with polycystic ovary syndrome, *BMC Women's Health*, 23, 98.
- Hossein Mirzaee Beni, Z., Maasoumi, R., Pashaeypoor, S. ve Haghani, S. (2022), The effects of self-care education based on the health literacy index on self-care and quality of life among menopausal women: a randomized clinical trial, *BMC Women's Health*, 22, 1-10.
- Khandehroo, M., Sany, S.B.T., Oakley, D. ve Peyman, N. (2022), Health literacy intervention and quality of life in menopausal women: A randomized controlled trial, *International Journal of Health Promotion and Education*, 60(2), 114-126.

- Kozak, F. (2021), *Aydın Adnan Menderes Üniversitesi Uygulama ve Araştırma Hastanesi Diyet Polikliniğine başvuran yaşlılarda sağlık okuryazarlığı ile yaşam kalitesi ilişkisi*, Yüksek Lisans Tezi, Aydın Adnan Menderes Üniversitesi Sağlık Bilimleri Enstitüsü, Aydın.
- Lee, C.J., Ho, M.H., Joo, J.Y., Montayre, J., Lin, Y.K., Chang, C.C. ve Liu, M.F. (2022), Gender differences in the association between oral health literacy and oral health-related quality of life in older adults, *BMC Oral Health*, 22.
- Liu, A.K., Liu, Y.Y., Su, J., Gao, J., Dong, L.J., Lyu, Q.Y. ve Yang, Q.H. (2023), Health literacy and quality of life of patients with coronary heart disease in Tibet, China: The mediating role of self-efficacy and self-management, *Heart & Lung*, 57, 271-276.
- Mehralian, G., Yusefi, A.R., Davarani, E.R., Ahmadidarrehsima, S. ve Nikmanesh, P. (2023), Examining the relationship between health literacy and quality of life: evidence from older people admitted to the hospital, *BMC Geriatrics*, 23.
- Michael, H., Thornton, S., Xie, M. ve Tian, L. (2019), Exact inference on the random-effects model for meta-analyses with few Studies, *Biometrics*, 75(2), 485-493.
- Mikkelsen, H.T., Skarstein, S., Helseth, S., Småstuen, M.C., Haraldstad, K. ve Rohde, G. (2022), Health-related quality of life, health literacy and COVID-19-related worries of 16-to 17-year-old adolescents and parents one year into the pandemic: a cross-sectional study, *BMC Public Health*, 22.
- Nacanabo, R., Debussche, X., Rouamba, M., Kamouni, P., Mancini, J. ve Kouanda, S. (2021), Health literacy and health-related quality of life in type 2 diabetes: A cross-sectional study in Burkina Faso, *Diabetes Epidemiology and Management*, 3.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akkl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. ve Moher, D. (2021), The PRISMA statement: An updated guideline for reporting systematic reviews, *BMJ*, 372, n71.
- Rague, J.T., Kim, S., Hirsch, J., Meyer, T., Rosoklija, I., Larson, J.E., Swaroop, V.T., Bowman, R., Bowen, D.K., Cheng, E.Y., Gordon, E.J., Holmbeck, G., Chu, D.I., Isakova, T., Yerkes, E.B. ve Chu, D.I. (2022), The association of health literacy with health-related

- quality of life in youth and young adults with Spina Bifida: A cross-sectional study, *The Journal of Pediatrics*, 251, 156-163.e2.
- Sertkaya, Z., Koyuncu, E., Nakipoğlu Yüzer, G. F. ve Özgirgin, N. (2023), Investigation of health literacy level and its effect on quality of life in patients with spinal cord injury, *The Journal of Spinal Cord Medicine*, 46(1), 62-67.
- Shayan, N.A. (2018), *Diyabetli olan ve olmayan hemodiyaliz hastalarının sağlık okuryazarlığı ve yaşam kalitelerinin değerlendirilmesi*, Doktora Tezi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Siddaway, A.P., Wood, A.M. ve Hedges, L.V. (2019), How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses, *Annual Review of Psychology*, 70, 747-770.
- Skoumalova, I., Madarasova Geckova, A., Rosenberger, J., Majernikova, M., Kolarcik, P., Klein, D., de Winter, A.F., van Dijk, J.P. ve Reijneveld, S.A. (2022), Low health literacy is associated with poorer physical and mental health-related quality of life in dialysed patients, *Int. J. Environ. Res. Public Health*, 19, 13265.
- Şimşek Küçükkeleşçe, D.Ş., Gölbaşı, Z., Bayer, N. ve Ağırbaş, H. (2021), An investigation of the relationship between health literacy and quality of life in pregnant women, *Adıyaman Üniversitesi Sağlık Bilimleri Dergisi*, 7(3), 213-222.
- Tad, S.İ. (2021), *Hipertansiyon hastalarında sağlık okuryazarlığının ilaç tedavisine uyum ve yaşam kalitesine etkisi*, Yüksek Lisans Tezi, Nevşehir Hacı Bektaş Veli Üniversitesi Fen Bilimleri Enstitüsü, Nevşehir.
- Topallar, B. (2019), *Koroner arter hastalığı olan bireylerin yaşam kalitesi ve sağlık okuryazarlık düzeyinin incelenmesi*, Yüksek Lisans Tezi, Ufuk Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Üner, E. ve Dost, A. (2023), Determination of the relationship between the level of health literacy and quality of life of vocational school of health services students during the covid-19 pandemic, *YOBU Sağlık Bilimleri Fakültesi Dergisi*, 4(1), 1-10.
- Yavuz, N. (2021), *Hemodiyaliz hastaları için geliştirilen sağlık okuryazarlığı ölçeği kullanılarak sağlık okuryazarlığının yaşam kalitesi ve tedavi etkinliği ile ilişkisinin değerlendirilmesi*, Uzmanlık Tezi, Gazi Üniversitesi Tıp Fakültesi, Ankara.

Yılmaz Kartal, İ.B. (2019), *Kronik hastalığı olan bireylerde sağlık okuryazarlığının yaşam kalitesi ile ilaç uyumu ile ilişkisi*, Uzmanlık Tezi, Atatürk Üniversitesi Tıp Fakültesi, Erzurum.

Zheng, M., Jin, H., Shi, N., Duan, C., Wang, D., Yu, X. ve Li, X. (2018), The relationship between health literacy and quality of life: a systematic review and meta-analysis, *Health and Quality of Life Outcomes*, 16(1).

Yayın Geliş Tarihi (Submitted): 26/08/2023

Yayın Kabul Tarihi (Accepted): 29/05/2024

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

Please Cite As/Atıf için:

Çakır, Y. ve Çelik, S. B. (2024), Psikolojik esneklik ölçeği Türkçe formu: geçerlik ve güvenilirlik çalışması, *Nicel Bilimler Dergisi*, 6(1), 19-40. doi: 10.51541/nicel.1350589

PSİKOLOJİK ESNEKLİK ÖLÇEĞİ TÜRKÇE FORMU: GEÇERLİK VE GÜVENİRLİK ÇALIŞMASI

Yasemin Çakır¹ ve Seher Balcı Çelik²

ÖZ

Bu çalışmanın amacı Ben-Itzhak, Bluvstein ve Maor (2014) tarafından geliştirilen Psikolojik Esneklik Ölçeği'nin Türkçeye uyarlanmasıdır. Araştırmaya 2021-2022 eğitim-öğretim yılında üniversite 1., 2., 3. ve 4. Sınıf düzeyinde öğrenim görmekte olan toplam 337 üniversite öğrencisi katılmıştır. Katılımcıların 204'ü kadın; 133'ü erkektir. Psikolojik Esneklik Ölçeği'nin orijinal ölçek formunda olduğu gibi beş faktörlü modeli "Olumlu Değişim Algısı", "Kendini Esnek Biri Olarak Görme", "Kendini Açık ve Yenilikçi Biri Olarak Görme", "Dinamik ve Değişen Bir Gerçeklik Algısı" ve Çok Yönlü Bir Gerçeklik Algısı" sergileyip sergilemediğini tespit edebilmek amacı ile doğrulayıcı faktör analizi yapılmıştır. DFA sonucuna göre program tarafından üretilen değişiklik önerileri incelenmiştir. Aynı alt ölçekte yer alan maddeler incelenerek benzer içeriğe sahip olan maddelere dair hata kovaryansları önerilen biçimde modele eklenerek analizler tekrar yapılmıştır. Ayrıca aynı anda birden fazla faktöre yüklenerek ölçeğin faktör yapısı üzerinde olumsuz etkide bulunan maddeler ölçek formundan çıkarılmıştır. Psikolojik Esneklik Ölçeği'nin ölçüt geçerliğini tespit edebilmek amacıyla Kabul ve Eylem Formu II Ölçeği ile arasındaki korelasyona bakılmıştır. Korelasyon analizleri sonucunda ulaşılan değerler de ölçek formunun uyum geçerliğinin sağlandığını ortaya koymaktadır. Ölçeğin güvenilirliği Cronbach alfa iç tutarlılık katsayısı ve test tekrar

¹ Sorumlu Yazar, Psikolojik Danışman/Rehber Öğretmen, Ondokuz Mayıs Üniversitesi, Lisansüstü Eğitim Enstitüsü, Doktora Öğrencisi, Rehberlik ve Psikolojik Danışmanlık Ana Bilim Dalı, Samsun, Türkiye. ORCID ID: <https://orcid.org/0000-0002-1940-4406>

² Prof. Dr., Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Rehberlik ve Psikolojik Danışmanlık Ana Bilim Dalı, Samsun, Türkiye. ORCID ID: <https://orcid.org/0000-0001-9506-6528>

yöntemleri ile analiz edilmiştir. Araştırma bulguları psikolojik esneklik ölçeğinin geçerli ve güvenilir bir ölçme aracı olduğunu göstermektedir.

Anahtar Kelimeler: Psikolojik esneklik, kabul ve kararlılık, ACT terapisi

PSYCHOLOGICAL FLEXIBILITY SCALE TURKISH FORM: VALIDITY AND RELIABILITY STUDY

ABSTRACT

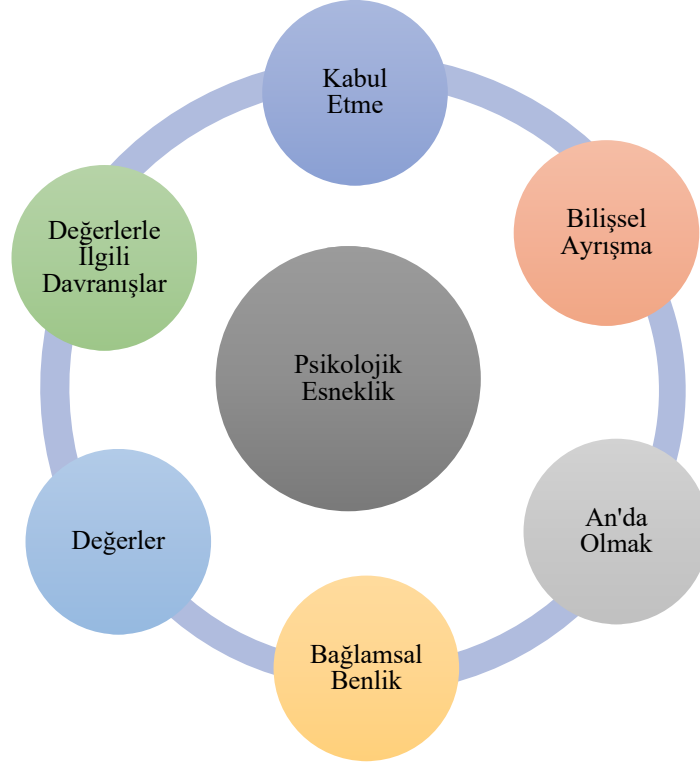
The aim of this study is to adapt the Psychological Resilience Scale developed by Ben-Itzhak, Bluvstein, and Maor (2014) into Turkish. A total of 337 university students studying at the 1st, 2nd, 3rd and 4th grades of the university in the 2021-2022 academic year participated in the research. 204 of the participants were women (60.53%); 133 of them are male (39.47%). In order to determine the sample size, the number of items in the scale was taken as a basis. Participants with a size of at least 10 times the number of items were included in the study. As in the original scale form of the Psychological Resilience Scale, its five-factor model is "Perception of Positive Change", "Seeing Oneself as Flexible", "Seeing Oneself as Open and Innovative", "A Dynamic and Changing Perception of Reality" and a Versatile Reality. Confirmatory factor analysis was carried out in order to determine whether it exhibits "perception". According to the DFA results, the suggestions for changes produced by the program were examined. By examining the items in the same subscale, the error covariances of the items with similar content were added to the model as suggested, and the analyzes were repeated. In addition, the items that were loaded on more than one factor at the same time and had a negative effect on the factor structure of the scale were removed from the scale form. In the last case, it is determined that $\chi^2/df=2.15$. Other fit index values are CFI=0.97, RMSEA=0.07, GFI=0.93. These values indicate a good level of goodness of fit of the scale. In order to determined the criterion validilty of the Psychological Resilience Scale, the correlation between it and the Acceptance and Action Form II Scale was examined. The values obtained as results of the corelation analyze also reveal that the congruent validity of the scale form is provide. The reliability of the scale was analyzed by Cronbach's alpha internal consistency coefficient and test-repeat methods. Research findings show that the psychological flexibility scale is a valid and reliabled measurement tool.

Keywords: Psychological flexibility, acceptance and commitment, ACT therapy

1. GİRİŞ

Steven Hayes, otuz yaşındayken bilimsel bir toplantıda panik atak geçirmeye başlamıştır ve yaklaşık iki yıl boyunca panik atak nöbetleri ile mücadele etmiştir. Steven Hayes bu yaşadığı ve deneyimlediği süreçte, acının insan hayatı için kaçınılmaz olduğunu ve bu acının kabul edilmesi gerektiğini fark etmiştir. Buradan hareketle olumsuz düşüncelerle savaşmak yerine o olumsuz düşünceleri yaşamın bir parçası olarak kabul etmeyi vurgulayan Kabul ve Kararlılık Terapisi (ACT) yaklaşımını ortaya atmıştır (Cloud, 2006). Acıyı bastırmaya veya acıdan kaçmaya çalışmak, hayatta genellikle daha fazla acı çekilmesine neden olmaktadır (Follette, Pistorello, 2007). Benzer biçimde kötü bir anıyı düşünmemek için sarf edilen çaba genellikle o hatırayı daha fazla ortaya çıkarmaktadır (Bilgen, 2022). Hoş olan kadar hoş olmayan duygu ve düşüncelere sahip olmak psikolojik olarak sağlıklıdır (Hayes, S. ve Hayes, W. 2011).

Kabul ve Kararlılık Terapisi ‘üçüncü kuşak’ olarak nitelendirilen Bilişsel Davranışçı Terapiler arasında yer almaktadır (Bilgen, 2022). ACT’ta altı ana terapötik basamak bulunmaktadır. Bu basamaklar: kabul etme, bilişsel ayrışma, an’da olma, bağlamsal benlik, değerler ve değerlerle ilgili davranışlar (eylem) aşamalarını içermektedir. Bu basamakların sonunda bireyin “psikolojik esnekliğe” ulaşması hedeflenmektedir (Harris, 2016). Kabul ve Kararlılık Terapisi psikolojik esnekliği artırmak için tasarlanmış bir yaklaşımdır (Hayes ve Levin, 2012). Bu nedenle psikolojik esneklik, ACT’ın temel kavramıdır (Akt. ve Ulubay, 2020). Şekil 1’de Psikolojik Esneklik modeline yer verilmektedir.



Şekil 1. Psikolojik Esneklik Modeli (Blackledge, Ciarrochi ve Deane, 2009)

Kabul ve Kararlılık Terapisi yaklaşımında istenmeyen duygu ve düşüncelerin “durdurulması” ve “kontrol edilmesi” yerine duygu ve düşüncelerin anlaşılması, keşfedilmesini, gözlemlenmesini ve kabul edilmesi vurgulanmaktadır (Wright vd., 2014)

Kabul etme, bireyin psikolojik esnekliğe ulaşması aşamasında diğer ACT kavramlarının temelinde yer almaktadır. (Yalnız vd., 2019). Kabul etme ile bir sonraki basamağa geçilmektedir. Kabul etme olmadığı sürece birey sürekli sıkıntı yaşamakta ve acı çekmektedir (Bilgen, 2022). Kabul etme, bireylerin yaşamış olduğu problemlerden kaçınmamasını, problemleri ortadan kaldırmaya/yok etmeye çalışmamasını ve olumsuz duygularını bir süreliğine kendisine uğrayan bir ziyaretçi olarak görmesi manasına gelmektedir. Fakat bu durum, bireyin yaşamış olduğu problemlerden ya da yaşadığı acılardan hoşlandığı anlamına gelmemektedir (Yalnız vd. 2019). Kabul etme aracılığı ile birey, yaşamış olduğu duygular ile savaşmaz ve duygularını sürekli bir biçimde kontrol etme ihtiyacı duymaz (Harris, 2016).

Bilişsel ayrışma, dili gözlemeleme, duyguların ve düşüncelerin doğasını tanıma ve tüm bunlardan ayrışma becerisidir (Hayes, Strosahl ve Wilson, 2012). Bilişsel ayrışma bir adım geriye gidip düşüncelere uzaktan bakmak ve onların varlığı gözlemelemek demektir (Tekin Larsson, Ergin, 2022). Ayrışmanın amacı düşünceyi değiştirmektense, düşünce ile düşüncenin sahibi 'ben' arasındaki ilişkiyi değiştirmektir. Kelime, cümle ve düşünce ile kaynaşmak demek, bireyin dünyayı bir gözlük ile görmesi demektir. Gözünüzde sarı bir gözlük varsa dünyayı sarı olarak algıyorsunuz (Bilgen, 2022).

An'da olmak, bireyin yaşadığı dünya ile doğrudan bağlantı kurması ve dolayısı ile bireyin daha esnek davranışlarda bulunmasını amaçlamaktadır (Hayes vd., 2006). ACT, bireyleri şimdiki ana yönlendirerek, bireylerin şimdi ve buradaki deneyimlerine dair farkındalıklarını geliştirmeye teşvik eder (Gordon vd., 2017). Birey, geçmişte yaşamış olduğu olumsuz deneyimler ve gelecekte karşılaşılabilecek olduğu olası problemlerin endişesi ile an'da olmakta zorluk çekebilmektedir. Ayrıca birey, hoş olmayan içsel deneyimlerinin ortaya çıkmasından korktuğu için de an'da olmaktan kaçınabilmektedir. Bu durum sürekli bir şekilde değerlendirme yapmaya, inceleme yapmaya ve analiz yapmaya neden olarak bireyin yaşam kalitesini düşürmektedir (Ögel, 2015; Gordon, Borushok ve Polk, 2017).

Bağlamsal benlik "kendini gözlemeleme" olarak da tanımlanmaktadır (Dubler, 2018). Bu durum bireyin benliğini gözlemelemesinin benzersiz bir yoludur (Gordon, Borushok, Polk, 2017). Birey, yaşantılarını duyu organları aracılığıyla algıladığı fiziksel bir benliğe, bilişsel özellikler aracılığıyla algıladığı düşünsel bir benliğe ve bu iki benliği izleyen -gözlemeleyen- bir benliğe sahiptir (Harris, 2009). Bağlamsal benlik, bireyin tüm benliğini tek bir yaşantısına göre değil yaşantısının tamamına bakarak incelemesi biçiminde ifade edilmektedir (Dubler, 2018).

Değerler; eylem, kabul etme ve ayrışmanın arkasındaki temel itici güçlerdir (Wright, Turkington, Kelly, Davies, Jacobs, Hopton, 2014, s.10). Değerleri bir birey için kişisel olarak kimin ve neyin önemli olduğuna dair benzersiz şekilde seçilmiş ilke ve standartlardır (Gordon, Borushok, Polk, 2017, s.14). Değerlerin yokluğunda birey çevresindeki diğer bireyler açısından "iyi" ve "doğru" olana göre yaşamakta ve çevresindeki diğer bireylerin gözünde iyi ve doğru olmak için çaba göstermektedir (Bond, Hayes ve Barnes-Holmes, 2006). Değerler ile hedefleri birbirine karıştırmamak gerekir. Çünkü hedeflerin bir bitiş noktası bir sonu/sonucu var iken değerler yaşam boyu sürmektedir (Yalnız vd., 2019).

Değerlerle ilgili davranışlar, bireylerin seçmiş olduğu değerlere uygun olarak küçük ya da büyük hedefler belirlemesi ve belirlediği hedefleri gerçekleştirmeye başlamasıdır. Bir başka deyişle değerlerle ilgili davranışlar bireylerin değerlerine ulaşabilmek amacı ile eyleme geçmesi olarak tanımlanmaktadır (Bramwell ve Richardson, 2018).

Psikolojik esnekliğin, bireylerin ruhsal sağlığı ve iyi oluş düzeyleri ile arasındaki ilişki klinik olarak saptandığından itibaren bireylerin psikolojik esnekliğini geliştirmek önemli bir amaç olmuştur (Hayes vd., 2006; Kashdan, 2010). Literatür incelendiğinde, bireylerin psikolojik esneklik düzeylerini ölçmek için geliştirilen çeşitli ölçekler bulunmaktadır (Bond vd., 2011, BenItzhak, Bluvstein ve Maor, 2014; Rolffs, Rogge ve Wilson, 2018). Bu alanda geliştirilen veya uyarlanan psikolojik ölçme araçlarının literatüre katkı sağlayacağı düşünülmektedir. BenItzhak, Bluvstein ve Maor (2014) tarafından geliştirilen Psikolojik Esneklik Ölçeği, diğer Psikolojik Esneklik Ölçeklerinin ölçmediği ve psikolojik esnekliğin önemli özelliklerini dışarıda bırakan önemli özellikleri ölçmektedir. Ayrıca psikolojik esnekliği ölçmeye yardımcı diğer ölçekler, psikolojik esnekliğe benzer kavramları ancak doğrudan psikolojik esneklikle ilgili olmayan yönleri de içermektedir (Ben-Itzhak, Bluvstein ve Maor, 2014). Dolayısıyla bu çalışmada Ben-Itzhak, Bluvstein ve Maor (2014) tarafından geliştirilen Psikolojik Esneklik Ölçeği'nin Türkçeye uyarlama çalışması yapılması amaçlanmıştır.

2. YÖNTEM

2.1. Çalışma Grubu

Araştırmaya 2021-2022 eğitim-öğretim yılında üniversite 1., 2., 3. ve 4. sınıfta öğrenim görmekte olan 337 üniversite öğrencisi gönüllü olarak katılmıştır. Katılımcıların 204'ü kadın (%60.53); 133'ü erkektir (%39.47). Örneklem büyüklüğünü belirleyebilmek amacı ile ölçekte yer alan madde sayısı esas alınmıştır. Madde sayısının en az 10 katı büyüklüğünde katılımcı araştırmaya dahil edilmiştir (Şencan, 2005).

2.2. Veri Toplama Araçları

Bu çalışmada veri toplama aracı olarak Psikolojik Esneklik Ölçeği, Kabul ve Eylem Formu II Ölçeği ve araştırmacılar tarafından oluşturulan kişisel bilgi formu kullanılmıştır.

2.2.1. Kişisel Bilgi Formu

Formda üniversite öğrencilerinin cinsiyeti ve öğrenim gördüğü sınıf düzeyine ilişkin bilgiler yer almaktadır.

2.2.2. Kabul ve Eylem Formu II

Kabul ve Eylem Formu-II (KEF-II), Bond ve arkadaşları (2011) tarafından bireylerin psikolojik esneklik düzeylerini ölçebilmek amacı ile geliştirilmiştir. Ölçek formunun Türkçe uyarlama çalışması Yavuz ve arkadaşları (2016) tarafından yapılarak literatüre kazandırılmıştır. Araştırmanın çalışma grubunu kaygı bozukluğu tanısı, anti-sosyal kişilik bozukluğu tanısı, depresyon tanısı veya iki uçlu bipolar bozukluk tanısından en az birini almış olan 207 birey ve hiç tanısı olmayan 267 birey olmak üzere toplamda 474 katılımcı oluşturmaktadır. Ölçeğin Türkçe formunun güvenirlik düzeyini ölçmek amacıyla iç tutarlılık ile zamana göre tutarlılık analizi yapılarak elde edilen değerler incelenmiştir. Ölçeğin Cronbach Alfa iç tutarlılık katsayı ortalaması 0.84'tür. 60 günlük ara ile gerçekleştirilen test-tekrar test analizinin sonucunda ise korelasyon katsayısının 0.85 olduğu tespit edilmiştir.

Ölçeğin yapı geçerliğini ölçebilmek amacıyla açımlayıcı ve doğrulayıcı faktör analizleri gerçekleştirilmiştir. Kaiser-Meyer-Olkin değerine göre ($r=0.83$) yapı geçerliği açısından ölçeğin uygun bir faktör yapısına sahip olduğu tespit edilmiştir. Toplam varyans açıklama oranının %51.76 olması, ölçeğin tek faktör olarak açıklanması bağlamında uygun olmaktadır. Bu sayede, Türkçe Kabul ve Eylem Formunun 7 maddeden meydana gelen tek faktörlü bir yapısının olduğu tespit edilmiştir. Ölçekte tüm maddeler ters kodlanmaktadır. Tüm maddelerden elde edilen puanlar toplanmaktadır. Ölçekten alınan yüksek puanlar (en çok 49 puan) bireyin psikolojik katılık düzeyini göstermekte iken ölçekten alınan düşük puanlar (en az 7 puan) ise bireyin psikolojik esneklik düzeyini göstermektedir. Bulgular birlikte incelendiğinde, KEF-II klinik örneklemeler ile klinik olmayan örneklemeler üzerinde, bireylerin psikolojik esnekliğini ölçebilen tek boyuta sahip geçenli ve güvenilir bir ölme aracıdır.

2.2.3. Psikolojik Esneklik Ölçeği

Ölçek Maor, Ben-Itzhak ve Bluvstein (2014) tarafından geliştirilerek literatüre kazandırılmıştır. Araştırmanın çalışma grubunu 197 üniversite mezunu öğrenciler oluşturmaktadır. Ölçeğin ön formu 60 maddeden oluşmaktadır. Ölçeğin ön uygulaması 90 katılımcıya uygulanmıştır. Uzman görüşleri ve ön uygulama kapsamında elde edilen veriler üzerinde yapılan analizler sonucunda 40 madde ölçekten çıkarılmıştır. Katılımcıların yaş

ortalaması 25'tir. Ölçek 5 alt boyut ve 20 maddeden oluşmaktadır. Yapı geçerliğini ölçmek amacı ile açımlayıcı ve doğrulayıcı faktör analizleri gerçekleştirilmiştir. Cronbach Alfa iç tutarlılık katsayısı ortalamasının 0.90 olduğu tespit edilmiştir. Yapı geçerliği incelendiğinde temel bileşen analizi ile Kaiser-Meyer-Olkin değerine göre ($r=0.67$) ölçeğin uygunluk gösteren bir faktör yapısının olduğu tespit edilmiştir. Toplam varyans açıklama oranı %66.8'dir. Ölçek beş faktörlü bir yapıdan oluşmaktadır. Ölçekte ters kodlanan madde bulunmamaktadır. Ölçek 6'lı Likert tipli bir ölçektir. 1: Hiçbir zaman/asla ve 6: her zaman anlamına gelecek şekilde maddeler 1'den 6'ya kadar puanlanmaktadır. Ölçekten en az 6 puan; en yüksek 120 puan alınmaktadır. Psikolojik Esneklik Ölçeği'nin psikolojik esnekliği ölçlebilen beş faktörlü bir yapıdan oluşan güvenilir ve geçerli bir ölçek olduğu görülmektedir (Maor, Ben-Itzhak ve Bluvstein, 2014).

2.3. Verilerin Toplanması

Ölçme araçları, 2021-2022 eğitim-öğretim yılında üniversite öğrencilerine uygulanmıştır. Ölçekler Giresun Üniversite'sine gidilerek uygulanmıştır. Uygulamanın yapıldığı günlerde okulda bulunmayan öğrencilere, sınıf iletişim grupları aracılığı ile uygulama formu online platform üzerinden ulaştırılmıştır. Ayrıca ölçek formu sosyal medya hesaplarında da paylaşılarak üniversite öğrencilerinin erişimine açılmıştır. Uygulama sürecinde katılımcılara yönelik olarak bu çalışmanın amacına, çalışmaya katılım için gönüllülüğün temel alındığına, uygulama formunda verilen yanıtların gizli tutulacağına ve uygulama formunun nasıl doldurulacağına ilişkin bilgilendirme yapılmıştır. Bu bağlamda uygulama formu araştırmacı tarafından katılımcılara sınıfta dağıtılmıştır veya katılımcılara online platform üzerinden iletilmiştir. Uygulama aşaması yaklaşık olarak 12-16 dakika aralığında tamamlanmıştır.

2.4. İşlem

Ölçek geliştirme makalesinde, ölçeği kullanmak için araştırma makalesini referans göstermenin yeterli olacağı bilgisi yer almaktadır. Makale sebeple ölçek geliştirme makalesi referans gösterilerek uyarılama çalışmasına başlanmıştır. Literatür incelendiğinde bahsi geçen ölçeğin Türkçeye uyarılama çalışmasının yapılmadığı görülmüştür. Orijinal dili İngilizce olan ölçeğin çeviri çalışması, iki İngilizce öğretmeni ve bir uzman psikolojik danışmandan oluşmak üzere üç kişilik bir ekip tarafından birbirinden bağımsız biçimde yapılmıştır. Çeviriler 4 uzman psikolojik danışman ve Ondokuz Mayıs Üniversitesi Psikolojik Danışmanlık ve Rehberlik Anabilim Dalı'nda görev yapan bir profesör tarafından incelenerek

düzenlenmiş ve geri çeviri işlemleri için tek bir form oluşturulmuştur. Ölçek maddelerinin değerlendirilmesi amacıyla ölçek formu Ondokuz Mayıs Üniversitesi Psikolojik Danışmanlık ve Rehberlik Anabilim Dalı'nda görev yapan bir profesör, bir doçent ve bir doktor öğretim üyesi akademisyene iletilmiştir. Uzman görüşlerinin ardından ölçek formuna son şekli verilmiştir.

Ölçeğin Türkçe formunun geri çevirisini yapmak üzere Orta Doğu Teknik Üniversitesi'nde İngilizce dilinde eğitim alan bir mezuna ve iki İngilizce öğretmeni tarafından İngilizceye çevrilmiştir. Çeviri formu ile orijinal formun karşılaştırılması yapılmıştır. Geri çeviriler Erasmus projesinde görev yapan iki İngilizce öğretmeni tarafından incelenerek değerlendirilmiştir.

Uzman görüşleri alındıktan ve değerlendirmeler yapıldıktan sonra gerekli düzenlemeler yapılarak ölçeğe son şekli verilmiştir. Devamında çeviri ölçek formu ile orijinal ölçek formu arasındaki eş değerliğini tespit etme aşamasına geçilmiştir. Çalışmada İngilizce ve Türkçe formlar art arda verilmiştir. Çalışmaya Giresun Üniversitesi İngiliz Dili ve Edebiyatı programında öğrenim görmekte olan 27 üniversite 4. Sınıf öğrencisi katılmıştır.

2.5. Verilerin Analizi

Verilerin analizini yapmak amacı ile SPSS 27.00 ve AMOS 21 paket programları kullanılmıştır. Dilsel eşdeğerlik çalışmasından sonra madde analizi, geçerlik ve güvenirlik analizleri yapılmıştır. Güvenirlik analizinde Cronbach alfa iç tutarlılık katsayısı, katılımcıların Türkçe ve İngilizce formlara verdiği yanıtlardan elde edilen veriler üzerinde Dilsel Geçerliği ve uyum geçerliğini saptamak amacıyla Pearson Momentler Çarpımı Korelasyon Analizinden ve geçerlik analizinde ise doğrulayıcı faktör analizinden faydalanılmıştır. Veri analizinde Psikolojik Esneklik Ölçeği'nin kaç faktörlü bir yapıya sahip olduğunu tespit edebilmek amacı ile doğrulayıcı faktör analizi (DFA) yapılmıştır. Ölçek formunun yapısına dair bilgi veren deneysel çalışmalar olduğu için DFA yapılmıştır (Byrne, 2010). Bu kapsamda DFA ile olası faktör yapısı bilinmek üzere var olan verinin olası faktör yapısı ile ne ölçüde uyumlu olduğu test edilmektedir. DFA'da çoklu benzerlik modeli kullanılarak sunulan modelin veriye uyum derecesinin incelenmesinde çoğunlukla önerilen (Hoper, Couhlan ve Mullen, 2008; Kline, 2005) farklı uyum iyiliği istatistik değerlerinden faydalanılmıştır. Bu istatistiksel değerler; kare-kare (χ^2), kare-kare'nin serbestlik derecesine oranı (χ^2 / sd), karşılaştırmalı uyum indeksi (CFI) ve yaklaşık hataların ortalama karekökü (RMSEA)'dür. Uyum iyiliği istatistiklerinin değerlendirilmesi Tablo 1'de yer alan bilgiler eşliğinde yapılmıştır.

Tablo 1. Uyum iyiliği indeksleri

Model Uyum İyiliği İstatistikleri	İyi Uyum1	Kabul Edilebilir Uyum1
χ^2	$0 \leq \chi^2 \leq 2df$	$2df \leq \chi^2 \leq 3df$
χ^2/df	0-2	2-3
GFI	$\geq .90$	$\geq .85$
AGFI	$\geq .90$	$\geq .85$
CFI	$\geq .97$	$\geq .95$
IFI	$\geq .95$.94- .90
RMR	$\leq .05$.06- .08
RMSEA	$\leq .05$.06- .08

(Seçer, 2018)

Ölçek formunun güvenirlik katsayısının tespit edilmesi aşamasında Cronbach alpha iç tutarlılık katsayı değeri hesaplanmıştır. Psikolojik Esneklik Ölçeği'nin Türkçe formunun ölçüt geçerliliğini incelemek amacıyla Yavuz ve arkadaşları (2016) tarafından Türkçe uyarlama çalışması yapılarak literatüre kazandırılan Kabul ve Eylem Formu II Ölçeği kullanılmıştır. Ölçüt geçerliliğinin incelenmesi amacıyla yukarıda belirtilen ölçekler arasındaki ilişkinin belirlenebilmesi amacı ile Pearson momentler çarpımı korelasyon katsayı değeri hesaplanmıştır.

3. BULGULAR

3.1. Dilsel Eşdeğerliğe İlişkin Bulgular

Çevirisi yapılan ölçek maddelerinin İngilizce formu ve Türkçe formu arasındaki dilsel eşdeğerliği belirleyebilmek amacı ile Giresun Üniversitesi İngiliz Dili ve Edebiyatı Programı 4. sınıf düzeyinde eğitim alan 27 katılımcıya uygulama yapılmıştır. Katılımcıların Türkçe ve İngilizce formlardan aldıkları puanlar arasındaki ilişki Pearson Momentler Çarpımı Korelasyon Katsayısı aracılığı ile tespit edilmiştir.

Tablo 2. Psikolojik Esneklik Ölçeği'nin Dilsel Eşdeğerliğini Belirleyebilmek Amacı ile Gerçekleştirilen Pearson Momentler Çarpımı Korelasyon Katsayısı Testi Sonuçları

Alt Boyutlar	T.1	İ.1	T.2	İ.2	T.3	İ.3	T.4	İ.4	T.5	İ.5	T. Toplam	İ. Toplam
T.1	1											
İ.1	,887*	1										
T.2	,975*	,870*	1									
İ.2	,849*	,967*	,871*	1								
T.3	,947*	,842*	,966*	,840*	1							
İ.3	,868*	,981*	,853*	,953*	,851*	1						
T.4	,970*	,842*	,963*	,835*	,969*	,847*	1					
İ.4	,850*	,979*	,838*	,951*	,821*	,983*	,819*	1				
T.5	,952*	,832*	,943*	,805*	,931*	,848*	,954*	,829*	1			
İ.5	,834*	,939*	,840*	,927*	,824*	,951*	,811*	,959*	,878*	1		
T. Toplam	,989*	,873*	,989*	,857*	,977*	,869*	,987*	,847*	,970*	,850*	1	
İ. Toplam	,874*	,991*	,871*	,980*	,850*	,988*	,846*	,989*	,847*	,966*	,874*	1

Kısaltmalar: T. 1: Türkçe 1. Alt Boyut: Olumlu Değişim Algısı; İ. 1: İngilizce 1. Alt Boyut: Olumlu Değişim Algısı; T. 2: Türkçe 2. Alt Boyut: Kendini esnek biri olarak görme; İ. 2: İngilizce 2. Alt Boyut: Kendini esnek biri olarak görme; T. 3: Türkçe 3. Alt Boyut: Kendini açık ve yenilikçi biri olarak görme; İ. 3: İngilizce 3. Alt Boyut: Kendini açık ve yenilikçi biri olarak görme; T. 4: Türkçe 4. Alt Boyut: Dinamik ve değişen bir gerçeklik algısı; İ. 4: İngilizce 4. Alt Boyut: Dinamik ve değişen bir gerçeklik algısı; T. 5: Türkçe 5. Alt Boyut: Çok yönlü bir gerçeklik algısı; İ. 6: İngilizce 6. Alt Boyut: Çok yönlü bir gerçeklik algısı; T. Toplam: Türkçe Toplam Puan: Psikolojik Esneklik; İ. Toplam: İngilizce Toplam Puan: Psikolojik Esneklik.

Tablo 1’de görüldüğü üzere Psikolojik Esneklik Ölçeği’nin dilsel eşdeğerlik çalışmasının sonucuna göre Türkçe form ile orijinal form arasındaki korelasyon katsayısı değerleri, Olumlu Değişim Algısı alt boyutu için $r=0.88$; Kendini Esnek Biri Olarak Görme alt boyutu için $r=0.87$; Kendini Açık ve Yenilikçi Biri Olarak Görme alt boyutu için $r=0.86$; Dinamik ve Değişen Bir Gerçeklik Algısı alt boyutu için $r=0.81$ ve Çok Yönlü Bir Gerçeklik Algısı alt boyutu için $r=0.87$ olarak hesaplanmıştır. Ulaşılan bu değerler Türkçe ölçek formu ile orijinal ölçek formunun eşdeğer olduğunu ortaya koymaktadır.

3.2. Psikolojik Esneklik Ölçeği Geçerlik Çalışmaları

3.2.1. Yapı geçerliliği

Psikolojik Esneklik Ölçeği’nin orijinal ölçek formunda (Maor, Ben-Itzhak, Bluvstein, 2014) olduğu gibi beş faktörlü modeli “Olumlu Değişim Algısı”, “Kendini Esnek Biri Olarak Görme”, “Kendini Açık ve Yenilikçi Biri Olarak Görme”, “Dinamik ve Değişen Bir Gerçeklik Algısı” ve Çok Yönlü Bir Gerçeklik Algısı” sergileyip sergilemediğini tespit edebilmek amacı ile DFA gerçekleştirilmiştir.

3.2.2. Doğrulayıcı faktör analizi (DFA)

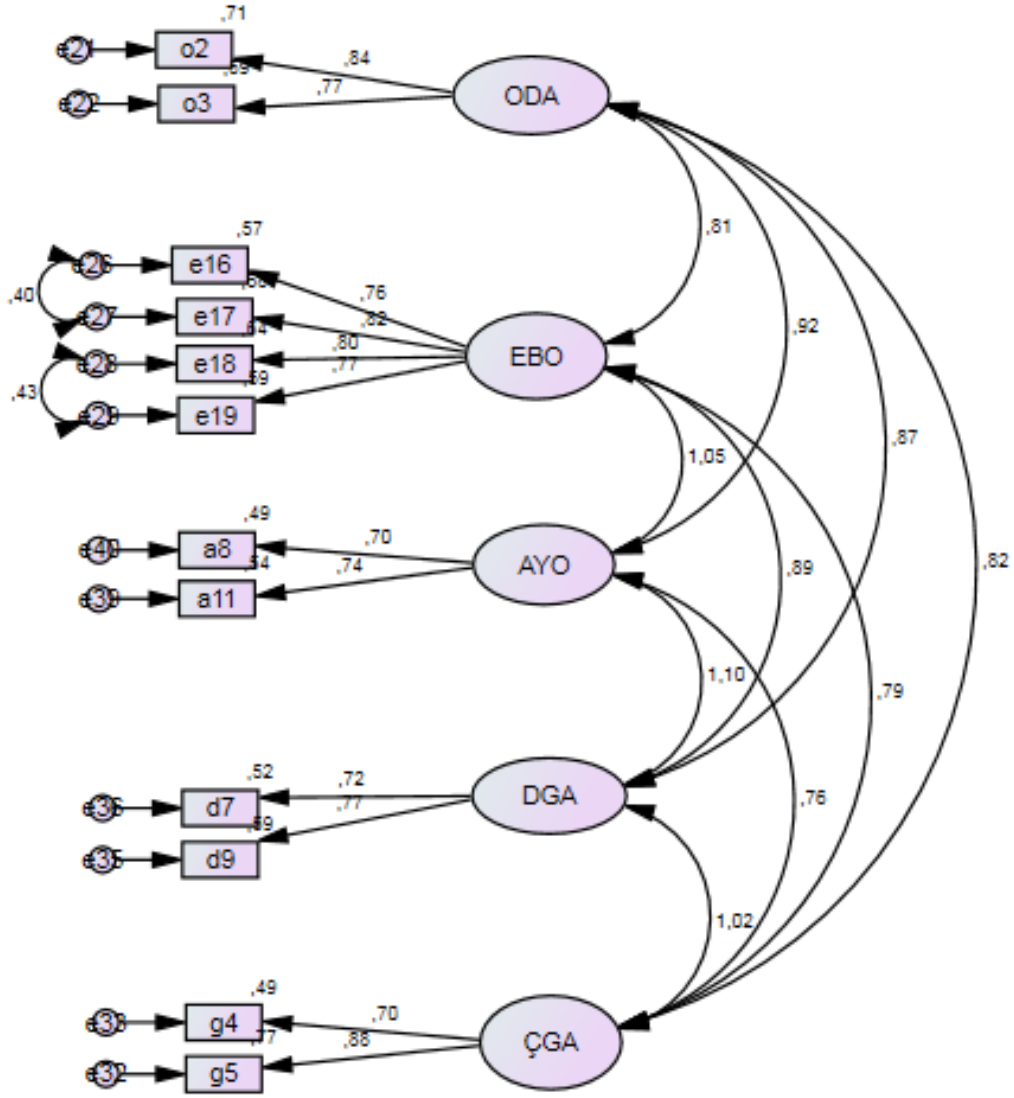
DFA sonucuna göre Kikare değeri (755.042, $p<0.001$) 0.05’ten küçük ve anlamlı düzeydedir. Kikare değeri, örneklemden etkilemesi sebebi ile anlamlılık gösterebilmektedir (Akt. Büyükgöze Kavas, Şanlı, İslam, 2021). Dolayısı ile ek olarak χ^2/sd oranının da hesaplanması önerilmektedir (Kline, 2005). Bu bağlamda, χ^2/sd oranının (755.042/160=4.719) 3’ten büyük ve ulaşılan diğer uyum değerleri (CFI=0.83, RMSEA=0.134, GFI=0.73) modelde iyi uyum olmadığını ortaya koymaktadır. Bu kapsamda, DFA sonucuna göre program tarafından üretilen değişiklik önerileri ele alınmıştır. 18. madde (*Çok alışılmamış davranış biçimlerini düşünmek benim için kolaydır*) ile 19. maddenin (*Değişikliklere/değişimlere açık hissediyorum*) hataları arasında ilişki olduğu tespit edilerek analizler tekrar yapılmıştır (Tablo 1). Tekrar yapılan analizler sonucunda χ^2/sd oranı 4.51 olarak hesaplanmıştır. Diğer uyum iyiliği değerleri ise CFI=0.84, RMSEA=0.130, GFI=0.75’tir. Daha sonra aynı anda birden fazla faktöre yüklenerek ölçek formunun faktör yapısı üzerinde olumsuz etkide bulunan 13. madde (*Her bir bireyden öğrenmek benim için önemlidir*), 14. madde (*Kendimi duruma göre değişebilen ve bunlara uyum sağlayabilen biri olarak tanımlarım*), 15. madde (*Gelecekteki değişiklikleri kabul etmeye hazır hissediyorum*), 20. madde (*Bir görev verildiği zaman, önceden kullandığım yollara kıyasla, görevi*

tamamlamanın özgün bir yolunu bulabilirim) maddeler modelden çıkarılarak analiz tekrar yapılmıştır (Tablo 3). Yapılan analizsonucunda $\chi^2/sd=3.19$, CFI=0.92, RMSEA=0.10, GFI=0.86 olarak hesaplanmıştır. Değişiklik önerileri tekrar incelendiğinde 16. madde (*Değişime olan ihtiyacıma dayalı olarak bazı zamanlar önemli kararlar verebilirim*) ve 17. madde (*Bir hedefi başarmak için zorlukla karşılaştığım zaman, sayısız farklı çözümler deneyebilirim*) hataları arasında ilişkinin mevcut olduğu tespit edilmiştir. Aynı alt boyutta bulunan bu maddeler incelenmiştir. Bu maddelerde birbirine benzer içeriklerin var olduğu saptanmıştır. Bu kapsamda belirtilen maddelere dair hata kovaryanslarının tavsiye edilen biçimde modele eklenmesi ile analizler yeniden yapılmıştır (Byrme, 2010). Bu işlemin ardından $\chi^2/sd=3$, CFI=0.92, RMSEA=0.09, GFI=0.86 olarak hesaplanmıştır. Ardından tekrar aynı anda birden fazla faktöre yüklenerek ölçek formunun faktör yapısı üzerinde olumsuz biçimde etkide bulunan 1. madde (*Gerçek asla kesin değildir*), 10. madde (*Bir değişikliği sık sık bir meydan okuma olarak görürüm*), 12. madde (*Gerçekliğin çok çeşitli belirtilerini, çoğu zaman önemli ölçüde farklı olabilen ve hatta birbirleriyle çelişebilen belirtilerini, kabul etmeyi kolay buluyorum*), 6. madde (*Farklı ve sıra dışı olanı deneyimlemeye açık biriyim*) maddeler modelden çıkarılarak analiz tekrarlanmıştır (Tablo 1). Son durumda, $\chi^2/sd=2.15$ olarak hesaplanmıştır. Diğer uyum indeks değerleri ise CFI=0.97, RMSEA=0.07, GFI=0.93'tür. Bu değerler uyum iyiliğinin iyi düzeyde olduğuna işaret etmektedir.

Gözlenen farklılıkların anlamlılığının değerlendirilmesinde Kikare değişim testinin, örneklem büyüklüğüne karşı oldukça hassas olması sebebi ile (Akt. Büyükgöze Kavas, Şanlı, İslam, 2021), Cheung ve Rensvold'un (2002) tavsiyesi bağlamında modellerin karşılaştırılması aşamasında CFI ve RMSEA fark kriterleri ($\Delta CFI \leq 0.01$, $\Delta RMSEA \leq 0.01$) dikkate alınmıştır (Tablo 1). Değişiklik önerileri kapsamında tekrar yapılan Doğrulayıcı Faktör Analizi sonucuna göre ölçeğin beş faktörlü bir yapıya sahip olduğunun doğrulandığı ifade edilebilir (Şekil 2). Standartlaştırılmış regresyon katsayısı değerlendirildiğinde bütün maddelere ait istatistiki verilerin anlamlı seviyede oldukları tespit edilmiştir. Ölçeğin faktör yükleri 0.70 ile 0.88 arasında değişmektedir.

Tablo 3. Uyum İyiliği İstatistik Değerleri

Model	χ^2	sd	CFI	RMSEA [90% CI]		Δ CFI	Δ RMSEA
A. Orijinal Model	755,042	160	,831	,134			
B. 18. ve 19. maddelere kovaryans eklenmiş	717,747	159	,841	,130	A – B	.010	.004
C. 13. madde çıkarılmış	563,106	141	,871	,120	B – C	.030	.010
D. 14. madde çıkarılmış	462,278	124	,886	,115	C – D	.015	.005
E. 15. madde çıkarılmış	356,299	108	,909	,105	D – E	.023	.010
F. 20. madde çıkarılmış	296,935	93	,917	,103	E – F	.008	.002
G. 16. ve 17. maddelere kovaryans eklenmiş	276,552	92	,925	,098	F- G	.008	.005
H. 1. madde çıkarılmış	211,873	78	,942	,091	G – H	.017	.007
İ. 10. madde çıkarılmış	166,935	65	,952	,087	H – I	.010	.004
J. 12. madde çıkarılmış	125,142	53	,962	,081	I – J	.010	.006
K. 6. madde çıkarılmış	90,443	42	,972	,075	J – K	.010	.006



Kısaltmalar: ODA: Olumlu Değişim Algısı Alt Boyutu, EBO: Kendini Esnek Biri Olarak Görme Alt Boyutu, AYO: Kendini Açık ve Yenilikçi Biri Olarak Görme Alt Boyutu, DGA: Dinamik ve Değişen Bir Gerçeklik Algısı Alt Boyutu, ÇGA: Çok Yönlü Bir Gerçeklik Algısı Alt Boyutu

Şekil 2. Psikolojik Esneklik Ölçeğine ilişkin beş faktörlü model: Standardize edilmiş katsayılar

3.2.3. Ölçüt Geçerliği

Psikolojik Esneklik Ölçeği'nin Türkçe formunun ölçüt geçerliliğini incelemek amacıyla KEF-II kullanılmıştır. Ölçüt geçerliliğinin incelenmesi amacıyla yukarıda belirtilen

ölçekler arasındaki ilişkilerin belirlenebilmesi için Pearson momentler çarpımı korelasyon katsayı değerleri tespit edilmiştir (Tablo 2).

Tablo 4. Alt Boyutlara dair ortalama, standart sapma ve korelasyon değerleri

	\bar{x}	s	1.	2.	3.	4.	5.	6.
1. Olumlu Değişim Algısı	21,2523	4,47233	1					
2. Kendini Esnek Biri Olarak Görme	20,1963	3,19031	,717*	1				
3. Kendini Açık ve Yenilikçi Biri Olarak Görme	15,9720	3,84477	,907**	,556**	1			
4. Dinamik ve Değişen Bir Gerçeklik Algısı	17,0467	2,90212	,800**	,914**	,633**	1		
5. Çok Yönlü Bir Gerçeklik Algısı	11,9720	2,85650	,882**	,508**	,767**	,545**	1	
6. Kabul ve Eylem Formu II	17,1308	8,93697	-,627**	-,521**	-,423**	-,650**	-,640**	1

** $p < .001$

Pearson korelasyon analizi sonucuna göre 0 – 0.29 arası düşük düzeyde ilişkiye; 0.30 –0.69 arası orta düzeyde ilişkiye; 0.70 ve üzeri yüksek düzeyde ilişkiye işaret etmektedir (Büyüköztürk, Çokluk ve Köklü, 2010). Buna göre Psikolojik Esneklik Ölçeği'nin birinci faktörü olan olumlu değişim algısı ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişki ($r=-0.62$); kendini esnek biri olarak görme ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişki ($r=-0.52$); kendini açık ve yenilikçi biri olarak görme ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişki ($r=-0.42$); dinamik ve değişen bir gerçeklik algısı ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişki ($r=-0.65$) ve çok yönlü bir gerçeklik algısı ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişki ($r=-0.64$) vardır. Kabul ve Eylem Formu II'de tüm maddeler ters puanlanmaktadır. Ölçekten alınan yüksek puan düşük psikolojik esneklik düzeyine, düşük puan ise yüksek psikolojik esneklik düzeyine işaret etmektedir (Yavuz ve ark., 2016). Bu sebeple Psikolojik Esneklik Ölçeği ile Kabul ve Eylem Formu II arasındaki ilişkinin negatif yönde olması beklenen bir sonuçtur.

2.3. Ölçeğin Güvenirliğine İlişkin Bulgular

Ölçeğin güvenilirlik katsayısını belirleyebilmek amacı ile Cronbach alpha iç tutarlılık katsayısı (α) analiz edilmiştir. Psikolojik Esneklik Ölçeği veri grubundan ulaşılan sonuca göre iç tutarlılık katsayısı Olumlu Değişim Algısı alt boyutu için 0.80; Kendini Esnek Biri Olarak

Görme alt boyutu için 0.88; Kendini Açık ve Yenilikçi Biri Olarak Görme alt boyutu için 0.69; Dinamik ve Değişen Bir Gerçeklik Algısı alt boyutu için 0.79 ve Çok Yönlü Bir Gerçeklik Algısı alt boyutu için 0.84 olarak hesaplanmıştır. Ölçek geliştirme ve uyarlama çalışmalarında ölçeğin güvenilir düzeyde olduğunun kabul edilebilmesi için 0.70 ve üzerinde bir güvenilirlik değerinin olması beklenmektedir. Ulaşılan bu değerler ölçek formunun yeterli düzeyde iç tutarlığa sahip olduğunu göstermektedir (Landis ve Koch, 1977; Robinson, Shaver ve Wrightsman, 1991).

4. SONUÇ, TARTIŞMA VE ÖNERİLER

Bu çalışmanın amacı Ben-Itzhak, Bluvstein ve Maor (2014) tarafından geliştirilen Psikolojik Esneklik Ölçeği'nin Türkçeye uyarlanmasıdır. Ölçeğin Türk kültürüne uyarlanması aşamalarında öncelikle dil geçerliği sağlanılmaya çalışılmıştır. Psikolojik Esneklik Ölçeği'nin dilsel eşdeğerlik çalışması sonucuna göre Türkçe form ile orijinal form arasındaki korelasyon katsayısı değerleri, Olumlu Değişim Algısı alt boyutu için $r=0.88$; Kendini Esnek Biri Olarak Görme alt boyutu için $r=0.87$; Kendini Açık ve Yenilikçi Biri Olarak Görme alt boyutu için $r=0.86$; Dinamik ve Değişen Bir Gerçeklik Algısı alt boyutu için $r=0.81$ ve Çok Yönlü Bir Gerçeklik Algısı alt boyutu için $r=0.87$ olarak hesaplanmıştır. Bu değerler Türkçe ölçek formu ile orijinal ölçek formunun eşdeğer olduğunu ortaya koymaktadır.

Psikolojik Esneklik Ölçeği'nin orijinal ölçek formunda (Maor, Ben-Itzhak, Bluvstein, 2014) olduğu gibi beş faktörlü modeli "Olumlu Değişim Algısı", "Kendini Esnek Biri Olarak Görme", "Kendini Açık ve Yenilikçi Biri Olarak Görme", "Dinamik ve Değişen Bir Gerçeklik Algısı" ve "Çok Yönlü Bir Gerçeklik Algısı" olup olmadığını tespit edebilmek amacı ile doğrulayıcı faktör analizi yapılmıştır. DFA sonucuna göre program tarafından üretilen değişiklik önerileri incelenmiştir. Aynı alt ölçekte var olan maddeler incelenerek benzer içeriğe sahip olan maddelere dair hata kovaryansları tavsiye edilen biçimde modele dahil edilerek analiz yeniden yapılmıştır (Byrme, 2010). Ayrıca aynı anda birden fazla faktöre etki ederek ölçeğin faktör yapısı üzerinde negatif bir etkiye neden olan maddeler ölçekten çıkarılmıştır. Son durumda, $\chi^2/sd=2.15$ olduğu tespit edilmiştir. Diğer uyum indeks değerleri ise CFI=0.97, RMSEA=0.07, GFI=0.93'tür. Bu değerler, son durumda iyi düzeyde bir uyum iyiliğine işaret etmektedir (Cole, 1987, Hu ve Bentler, 1999; Kline, 2011; Marcoulides ve Schumacher, 2001).

Ölçeğin güvenirlik katsayısını belirleyebilmek amacı ile Cronbach alpha iç tutarlılık katsayı değeri hesaplanmıştır. Psikolojik Esneklik Ölçeği'nin veri grubundan ulaşılan iç tutarlılık katsayı değeri Olumlu Değişim Algısı alt boyutu için 0.80; Kendini Esnek Biri Olarak Görme alt boyutu için 0.88; Kendini Açık ve Yenilikçi Biri Olarak Görme alt boyutu için 0.69; Dinamik ve Değişen Bir Gerçeklik Algısı alt boyutu için 0.79 ve Çok Yönlü Bir Gerçeklik Algısı alt boyutu için 0,84 olarak hesaplanmıştır. Ölçek geliştirme ve uyarlama çalışmalarında ölçeğin güvenilir düzeyde olduğunun kabul edilebilmesi için 0.70 ve üzerinde bir güvenirlik değerinin olması beklenmektedir. Ulaşılan bu değerler ölçek formunun yeterli düzeyde iç tutarlığa sahip olduğunu göstermektedir (Landis ve Koch, 1977; Robinson, Shaver ve Wrightsman, 1991).

Psikolojik Esneklik Ölçeği'nin ölçüt geçerliğini tespit edebilmek amacıyla KEF-II ile arasındaki korelasyon değeri incelenmiştir. Pearson'ın momentler çarpımı korelasyon analizi sonucunda Psikolojik Esneklik Ölçeği'nin birinci faktörü olan olumlu değişim algısı ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişkinin; kendini esnek biri olarak görme ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişkinin; kendini açık ve yenilikçi biri olarak görme ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişkinin; dinamik ve değişen bir gerçeklik algısı ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişkinin ve çok yönlü bir gerçeklik algısı ile KEF-II arasında negatif yönde orta düzeyde anlamlı bir ilişkinin olduğu tespit edilmiştir.

Kabul ve Eylem Formu II'de tüm maddeler ters puanlanmaktadır. Ölçekten alınan yüksek puan düşük psikolojik esneklik düzeyine, düşük puan ise yüksek psikolojik esneklik düzeyine işaret etmektedir (Yavuz ve ark., 2016). Bu sebeple Psikolojik Esneklik Ölçeği ile Kabul ve Eylem Formu II arasındaki ilişkinin negatif yönde olması beklenen bir sonuçtur.

Psikolojik Esneklik Ölçeği "Olumlu Değişim Algısı", "Kendini Esnek Biri Olarak Görme", "Kendini Açık ve Yenilikçi Biri Olarak Görme", "Dinamik ve Değişen Bir Gerçeklik Algısı" ve "Çok Yönlü Bir Gerçeklik Algısı" olmak üzere beş alt boyuttan meydana gelmektedir. Ölçekte bulunan alt boyutların her biri ayrı ayrı puanlanarak değerlendirilmeye alınmaktadır. Sonuç olarak Türkçeye uyarlama çalışması yapılan Psikolojik Esneklik Ölçeği üniversite öğrencilerinin psikolojik esneklik düzeylerini ölçmek için kullanılabilecek geçerli ve güvenilir bir ölçme aracıdır.

ETİK BEYAN

“Psikolojik Esneklik Ölçeği Türkçe Formu: Geçerlik ve Güvenirlik Çalışması” başlıklı çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş, toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

KAYNAKÇA

- Ben-Itzhak, S., Bluvstein, I. and Maor, M. (2014), The Psychological Flexibility Questionnaire (PFQ): Development, reliability and validity, *Webmed Central Psychology*, 5(4), 1-10.
- Bilgen, İ. (2022). *Terapide Psikolojik Esneklik Kabul ve Adanmışlık Terapisi*. İstanbul: Epsilon Yayınevi.
- Blackledge, J. T., Ciarrochi, J., and Deane, F. (Eds.). (2009), *Acceptance and Commitment Therapy: Contemporary Theory Research and Practice*. Bowen Hills, Australia: Australian Academic Press.
- Bond, F. W., Hayes, S. C. and Barnes-Holmes, D. (2006), *Psychological flexibility, ACT and organizational behavior*, *Journal of Organizational Behavior Management*, 26(1-2), 25-54.
- Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T. and Zettle, R. D. (2011), Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance, *Behavior Therapy*, 42(4), 676-688.
- Bramwell, K. and Richardson, T. (2018). Improvements in depression and mental health after acceptance and commitment therapy are related to changes in defusion and values-based action, *Journal of Contemporary Psychotherapy*, 48(1), 9-14.
- Büyükgöze Kavas, A., Şanlı, E. ve İslam, Ü. Y. (2021), Kariyer psikolojik danışmanlığı öz-yeterlik ölçeği Türkçe Formu: Geçerlik ve güvenilirlik çalışması, *Mehmet Akif Ersoy Eğitim Fakültesi Dergisi*, 58, 105-130.
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2010), *Sosyal bilimler için istatistik* (6. baskı), Ankara: Pegem Akademi.

- Cloud, J. (2006), Happiness isn't normal. *Time*, 167(7), 58-67.
https://facultystaff.richmond.edu/~bmayer/pdf/HappinessNotNormal_TIME.pdf
Erişim tarihi: 22.07.2022
- Dubler, B. A. (2018), Psychological flexibility as a key mechanism in Acceptance and Commitment Therapy: Shifts in flexibility as mediators of treatment effects (*Doctoral dissertation*). University of Rochester, USA.
- Follette, V. and Pistorello, J. (2007), *Finding life beyond trauma: Using acceptance and commitment therapy to heal from post-traumatic stress and trauma-related proble*. New Harbinger Publications.
- Gordon, T., Borushok, J. and Polk, K. L. (2017), *The ACT approach: A comprehensive guide for acceptance and commitment therapy*. PESI Publishing & Media.
- Harris, R. (2007), *Acceptance and Commitment Therapy (ACT) Introductory*. Workshop Handout. PESI, USA.
- Harris, R. (2009), *ACT with Love: Stop Struggling, Reconcile Differences, and Strengthen Your Relationship with Acceptance and Commitment Therapy*. Oakland, CA: New Harbinger.
- Harris, R. (2016), *Act'i kolay öğrenmek* (Çev. Karatepe, H. K. ve Yavuz, K. F.) İstanbul: Litera Yayınları.
- Hayes, S. C. and Levin, M. E. (Eds.). (2012), *Mindfulness and acceptance for addictive behaviors: Applying contextual CBT to substance abuse and behavioral addictions*. New Harbinger Publications.
- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A. and Lillis, J. (2006), Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1-25.
- Hayes, S. C., Strosahl K. D. and Wilson, K. G. (2012), *Acceptance and commitment therapy: The process and practice of mindful change*, New York: Guilford Press.
- Hayes, S. C., Strosahl, K. D. and Wilson, K. G. (2011), *Acceptance and commitment therapy: The process and practice of mindful change*. Guilford Press.

- Hooper, D., Coughlan, J., Mullen, M. R., ve AL., E. (2008), Evaluating model fit: A synthesis of the structural equation modelling literatüre, *Electronic Journal of Business Research Methods*, 6(1), 53–60.
- Kashdan, T. B. (2010), Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, 30(7), 865-878.
- Kline, T. (2005), *Psychological Testing: A practical approach to design and evaluation*. *Psychological Testing: A Practical Approach to Design and Evaluation*. Sage Publications.
- Landis, J. R. ve Koch, G. G. (1977), The measurement of observer agreement for categorical data, *Biometrics*, 33, 159-174.
- Ögel, K. (2015), *Farkındalık ve kabullenme temelli terapiler* (2. baskı), Ankara: HYB Yayıncılık.
- Robinson, J. P., Shaver, P. R. ve Wrightsman, L. S. (1991), *Criteria for scale selection and evaluation in measure of personality and social psychological attitudes*, San Diego: California Academic Press
- Rolffs, J. L., Rogge, R. D. and Wilson, K. G. (2018), Disentangling components of flexibility via the hexaflex model: Development and validation of the Multidimensional Psychological Flexibility Inventory (MPFI), *Assessment*, 25(4), 458-482.
- Seçer, İ. (2018), *Psikolojik test geliştirme ve uyarlama süreci: SPSS ve LISREL uygulamalar*, Ankara: Anı yayıncılık.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik*, Ankara, Seçkin Yayınevi.
- Tekin Larsson, E. G. Ve Ergin, B. (2022), *Act Metaforları Kabul ve Adanmışlık Terapisinde Deneyime Dayalı Alıştırmalar, Uygulamalı Metaforlar Rehberi*, İstanbul, Epsilon Yayınevi.
- Ulubay, G. (2020), *Çok Boyutlu Psikolojik Esneklik Envanterinin Türkçeye Uyarlanması*, Yüksek Lisans Tezi, Balıkesir Üniversitesi, Sosyal Bilimler Enstitüsü, Eğitim Bilimleri Anabilim Dalı, Balıkesir.
- Wright, N. P., Turkington, D., Kelly, O. P., Davies, D., Jacobs, A. M., and Hopton, J. (2014), *Treating psychosis: A clinician's guide to integrating acceptance*

and commitment therapy, compassion-focused therapy, and mindfulness approaches within the cognitive behavioral therapy tradition. New Harbinger Publications.

Yalnız, A. (2019), *Kabul ve Kararlılık Terapisi Yönelimli Psikoeğitim Programının Akran Zorbalığı Üzerindeki Etkisi*, Doktora Tezi, Sakarya Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitim Bilimleri Anabilim Dalı, Sakarya.

Yavuz, K. F., Ulusoy, S., Işkın, M., Esen, F. B., Burhan, H. Ş., Karadere, M. E. and Yavuz, N. (2016), Turkish version of Acceptance and Action Questionnaire-II (AAQ-II): A Reliability and Validity Analysis in Clinical and Non-clinical Samples, *Bulletin of Clinical Psychopharmacology*, 26(4), 397-408.

Yayın Geliş Tarihi (Submitted): 05/10/2023

Yayın Kabul Tarihi (Accepted): 27/05/2024

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

Please Cite As/Atıf için:

Diler, S. ve Demir, Y. (2024), Çoklu doğrusal bağlantı olması durumunda veri madenciliği algoritmaları performanslarının karşılaştırılması, *Nicel Bilimler Dergisi*, 6(1), 40-67. doi: 10.51541/nicel.1371834

ÇOKLU DOĞRUSAL BAĞLANTI OLMASI DURUMUNDA VERİ MADENCİLİĞİ ALGORİTMALARI PERFORMANSLARININ KARŞILAŞTIRILMASI

Saygın Diler¹ ve Yıldırım Demir²

ÖZ

Bilgisayar teknolojilerindeki gelişmelere paralel olarak veri madenciliği algoritmaları ile yapılan çalışmalarda artış yaşanmaktadır. Sınıflandırma algoritmaları ile yapılan çalışmalarda veri kalitesinin bozulması algoritmaların performansında önemli rol oynamaktadır. Bu çalışmada veri kalitesini bozan etmenlerden birisi olan çoklu doğrusal bağlantının veri setinde bulunması durumunda sınıflandırma algoritmalarının performansının nasıl etkilendiği incelenmiştir. Çoklu doğrusal bağlantının varlığını tespit etmek için veri setlerine ait korelasyon grafikleri incelenmiş daha sonrasında ise koşul endeksi ile çoklu doğrusal bağlantının derecesi belirlenmiştir. Sınıflandırma algoritmalarından olan Naive Bayes (NB), Lojistik Regresyon (LR) ve K-En Yakın Komşu Algoritması (kNN), Destek Vektör Makineleri (SVM) ve Aşırı Gradyan Arttırma Algoritması (XGBoost) ile uygulamalar gerçekleştirilmiştir. Yöntemlerin performanslarının incelenmesi için simülasyon çalışması ve gerçek veri setleri ile uygulamalar yapılmış, sonuçlar tablolar halinde sunulmuştur. Analiz sonuçlarına göre, çoklu doğrusal bağlantı varlığında büyük örneklem hacimli veri setlerinde doğruluk ve F-ölçütü metriklerine göre XGBoost algoritmasının diğer algoritmalarından dikkate değer performans farklılığı gösterdiği belirlenmiştir. Çoklu doğrusal bağlantıdan performansı en olumsuz etkilenen algoritmanın ise Naive Bayes olduğu gözlemlenmiştir.

¹ Sorumlu yazar, Türkiye İstatistik Kurumu, Ankara, Türkiye. ORCID ID: <https://orcid.org/0000-0002-9056-412X>

² Doç. Dr., İktisadi ve İdari Bilimler Fakültesi, Van Yüzüncü Yıl Üniversitesi, Van, Türkiye. ORCID ID: <https://orcid.org/0000-0002-6350-8122>

Anahtar Sözcükler: Çoklu doğrusal bağlantı, Sınıflandırma, Veri madenciliği

COMPARISON OF DATA MINING ALGORITHMS PERFORMANCES IN CASE OF MULTICOLLINEARITY

ABSTRACT

As advancements in computer technologies progress, there has been an increase in research utilizing data mining algorithms. In studies involving classification algorithms, the degradation of data quality plays a significant role in algorithm performance. This study investigates the impact of multicollinearity, one of the factors that compromise data quality, on the performance of classification algorithms. To identify the presence of multicollinearity, correlation graphs of the datasets were examined, followed by the determination of the degree of multicollinearity using the condition index. The classification algorithms, namely Naive Bayes (NB), Logistic Regression (LR), k-Nearest Neighbors (kNN), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost), were implemented for the analysis. Simulation studies and real dataset analyses were conducted to assess the performance of these methods, and the results were presented in tabular form. According to the analysis results, it has been determined that XGBoost algorithm shows a notable performance difference compared to other algorithms in terms of accuracy and F-measure metrics in the presence of multicollinearity in large sample-sized datasets. On the other hand, Naive Bayes was observed to be the algorithm most adversely affected by multicollinearity, showing diminished performance.

Keywords: Classification, data mining, multicollinearty.

1. GİRİŞ

Teknolojik gelişmelerin günümüzde hızla ilerlemesine bağlı olarak büyük miktarda ve çeşitlilikte veri üretilmektedir. Her geçen gün daha da artarak veri tabanlarında biriken karmaşık ve büyük veri setlerinden değerli bilgilerin ortaya çıkarılması için geliştirilmiş farklı yöntemler bulunmaktadır. Veri madenciliği yöntemi büyük veri setlerinden bilgilerin keşfedilmesi için geliştirilmiş yaklaşımlar arasında yer almaktadır (Han vd., 2012). Veri setinin kalitesi istatistiksel analiz süreçlerinde ve veri madenciliği sınıflandırma algoritmaları

ile yapılan çalışmalarda büyük bir öneme sahiptir. Çünkü bu algoritmalar ile yapılan analizler veri setinin kalitesinden etkilenmekte ve dolayısıyla algoritmaların etkin performans göstermesi veri kalitesiyle ilişkili olmaktadır (Batista ve Monard, 2002). Veri kalitesini bozan çeşitli etmenler bulunmakta ve çoklu doğrusal bağlantı bu etmenler arasında yer almaktadır. Veri madenciliği yöntemleri ile yapılan analizlerin başarısı, büyük ölçüde algoritma ve veri kalitesi ile ilişkilidir. Verilerin analize uygunluğu veri madenciliği çalışmalarının varsayımları arasında bulunmaktadır. Veri yapısı ve kalitesini etkileyen sorunlarla karşılaşmak istatistik biliminin doğal bir parçası olarak görülebilir. Bu tür sorunlar, veri madenciliği çalışmalarında sonuçları etkileyebilir. Veri kalitesi düşük olduğunda klasik veri madenciliği yöntemleri etkisiz hale gelebilmekte ve bu, algoritma performanslarını etkilemektedir (Zhu vd., 2018).

Burada yöntem performanslarının kötü etkilenmesi iki şekilde ele alınabilir: birincisi veriyi temsil edememek ve düşük doğruluk, keskinlik gibi değerlere sebep olmak, ikincisi makine öğrenmesi yöntemlerinde sıkça karşılaşılan aşırı uyum problemi ki bu da performans kriterlerinin yüksek çıkmasına sebep olmaktadır. Bu çalışmada söz konusu sorun, birinciden ziyade ikinci sorun olan aşırı uyum problemidir. Bu durum klasik regresyon modellerinde daha çok çoklu doğrusal bağlantıdan kaynaklanmaktadır. Aşırı uyum sorununun çözümü, farklı makine öğrenmesi yöntemleriyle birçok farklı alanda araştırılmıştır (Roelofs vd., 2019; Ying, 2019). Bu araştırmanın temel motivasyonlarından birisi, çözümden bağımsız şekilde bu sorunun farklı sınıflandırma algoritmalarını nasıl etkilediğinin ortaya çıkarılması ve karşılaştırma yapılarak yöntemlerin dayanıklılığı veya avantaj ve dezavantajları hakkında çıkarımlar yapmaktır.

Çoklu doğrusal bağlantı sorunu için istatistik alanında oldukça fazla literatür bulunmasına rağmen veri madenciliği, makine öğrenmesi gibi alanlarda çok az literatür bulunmaktadır (Garg ve Tai, 2013). Bu bağlamda veri kalitesini bozan etmenler arasında yer alan çoklu doğrusal bağlantının veri setinde olması durumunda sınıflandırma algoritmalarına ait performansların incelenmesi çalışmanın temel motivasyonunu oluşturmaktadır.

Literatürde, çoklu doğrusal bağlantı olması durumunda veri madenciliği yöntemlerini kullanan çalışmalardan bazılarında kısaca değinecek olunursa; Garg ve Tai (2013) vücuttaki yağ içeriğinin tahmin edilmesi için çoklu doğrusal bağlantıya sahip veri setinde uygulama gerçekleştirmiştir. Çalışmada makine öğrenmesi algoritmaları kullanılmış ve Yapay Sinir Ağır (YSA) ile genetik programlamanın başarılı sonuçlar verdiği gözlenmiştir. Blommaert vd. (2014) tarafından yapılan çalışmada, cezalandırılmış genelleştirilmiş tahmin denklemleri kullanılarak çoklu doğrusallık ve zamana bağımlılık altında uzunlamasına veriler için veri

madenciliği algoritmaları ile uygulama yapılmıştır. Dumancas ve Bello (2015) büyük veri analitiği ve yüksek performanslı veri madenciliğinde çoklu doğrusallığı ele almada makine öğrenmesi tekniklerinin karşılaştırılması üzerine araştırma gerçekleştirmiştir. Senawi vd. (2017), çoklu doğrusallığı azaltmak için özellik seçimi yöntemi olarak maksimum ilişki-minimum çoklu doğrusallık (MRmMC) yöntemi kullanmış ve sınıflandırma algoritmalarında iyi performans elde etmiştir. Obite vd. (2020) çoklu doğrusal bağlantıya sahip gerçek ve simülasyon verileri ile uygulama gerçekleştirmiştir. Uygulamada, Yapay Sinir Ağır (YSA) modelinin, Sıradan En Küçük Kareler Regresyonundan daha iyi bir uyum ve tahmin elde ettiği belirtilmiştir. Zhang vd. (2021), çoklu doğrusallığa sahip hisse senedi fiyat hareketlerini tahmin etmek için derin çarpanlara ayırma makinesi ve dikkat mekanizmasına dayalı (FA-CNN) sinir ağı modeli üzerine odaklanmıştır. Çalışma, girdi özellikleri arasındaki gün içi etkileşimleri ve yan sanayi endeksi gibi ek bilgilerin tahmin doğruluğunu arttırdığını göstermişlerdir. Rahman vd. (2021), meme kanserinin erken teşhisindeki performansı artırmak amacıyla çoklu doğrusallık analizi ve makine öğrenimi modelleri kullanarak meme kanseri belirteçlerini ve teşhis sistemini geliştirmeyi hedeflemektedir. Çalışmada destek vektör makineleri algoritması en başarılı algoritma olarak belirlenmiştir. Urooj vd. (2022), Android işletim sistemi uygulamalarındaki kötü amaçlı yazılımları tespit etmek için makine öğrenimi sınıflandırma algoritmalarını kullanarak çoklu doğrusallığa sahip verileri ele almış ve önerilen modelin %96.24 doğrulukla kötü amaçlı yazılımları tespit etmekte etkili olduğunu göstermişlerdir. Chan vd. (2022), çoklu doğrusallığı azaltmak için değişken seçimi ve değiştirilmiş tahmin yöntemlerinin kullanıldığına değinmiş, son araştırmalarda makine öğrenimi ile optimizasyon yaklaşımının çoklu doğrusallığı daha iyi ele aldığını belirtmişlerdir. Derraz vd. (2023), pirinç biyokütlesinin tahmininde temel ve topluluk makine öğrenmesi algoritmaları kullanmışlardır. Çalışmada çoklu doğrusal bağlantı olan ve çoklu doğrusal bağlantı olmayan veriler ile pirinç biyokütle tahmini gerçekleştirilmiştir. Çalışmada, topluluk makine öğrenmesi algoritmalarının temel makine öğrenmesi algoritmalarına göre daha başarılı performans gösterdiği belirlenmiştir.

Bu çalışmada, çoklu doğrusal bağlantı olması durumunda veri madenciliği sınıflandırma algoritmalarına ait sınıflandırma performansları incelenmektedir. Analiz öncesinde çoklu doğrusal bağlantının varlığını incelemek için çoklu korelasyon ısı grafikleri ve daha sonrasında ise çoklu doğrusal bağlantının derecesini belirlemek için koşul indeksi incelenmiştir. Çoklu doğrusal bağlantıya sahip iki gerçek veri seti ve simülasyon verileri ile uygulamalar gerçekleştirilmiş ve algoritmaların performansları karşılaştırılmıştır.

2. MATERYAL VE YÖNTEM

Veri setinde çoklu doğrusal bağlantı olduğunda sınıflandırma algoritmalarının performanslarını karşılaştırılmak amacıyla ilk önce iki gerçek veri setiyle ve daha sonra ise simülasyon çalışmasıyla uygulamalar gerçekleştirilmiştir. Veri setleri eğitim (%75) ve test (%25) verisi olacak şekilde ikiye ayrılmıştır. Eğitim veri setiyle modeller oluşturulmuş ve daha sonra test veri setiyle bu modellerin performansları ölçülmüştür. Model performanslarını karşılaştırmada ise kesinlik, seçicilik, duyarlılık, doğruluk ve F-ölçüt kriterleri kullanılmıştır. Çalışmada, Destek Vektör Makineleri, Lojistik Regresyon, Naïve Bayes, k-En Yakın Komşu (kNN) ve Aşırı Gradyan Arttırma algoritmaları ile uygulamalar gerçekleştirilmiş ve R programlama dili kullanılmıştır.

2.1. Sınıflandırma Algoritmaları

Sınıflandırma yöntemlerinde amaç bağımsız değişkenler yardımıyla genellikle kategorik verilerden oluşan bağımlı değişkeni tahmin etmektir. Sınıflandırma için çalışma yapısı ve varsayımları birbirinden farklılık gösteren birçok algoritma geliştirilmiştir (Davidson ve Tayi, 2009).

2.1.1. K-En Yakın Komşu Algoritması

kNN algoritması, veri madenciliğinde en çok kullanılan algoritmalarından birisidir. Algoritma, sınıfları bilinen veri setinde sınıfları bilinmeyen yeni verileri en yakın komşularına atama mantığı ile çalışmaktadır (Mucherino vd., 2009). Bu çalışmada, kNN algoritması için “k” değeri doğruluk ölçütünü maksimum yapabilecek şekilde çapraz geçerlilik yöntemiyle 1 ile 20 arasında bir değer seçilerek optimize edilmiştir. Zira kNN algoritmasının doğru sınıflandırma yapabilmesi büyük ölçüde komşu sayısının (k değeri) uygun seçimiyle ilgilidir.

2.1.2. Naive Bayes Algoritması

Bayesci sınıflama yöntemi istatistik tabanlı algoritmalar arasında yer almakta ve algoritma gözlemlere dayalı olasılıklar ile olasılık dağılımındaki parametreleri hesaplamaktadır (McNamara vd., 2006). Yöntem, veri setinde sınıfları belli olan verileri kullanarak yeni bir verinin mevcut sınıflardan birine girme olasılığını belirlemektedir (Silahtaroglu, 2013).

Sınıfları belirlemede algoritma koşullu olasılıkları kullanmakta ve olasılıklar,

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

olarak hesaplanmaktadır. Burada C_1, C_1, \dots, C_m sınıf değerlerini ve $X = \{x_1, x_2, \dots, x_n\}$ nitelik değerlerinden oluşan ve sınıfı bilinmeyen veri örneğini göstermektedir.

Bilinmeyen bir örneği sınıflandırmak amacıyla maksimum değer Eşitlik (1) ile hesaplanmakta ve bilinmeyen X örneğinin bu sınıfta olabileceğine karar verilmektedir. Eşitlik (2) sonlu olasılıkları kullanmakta ve en büyük sonlu sınıflandırma yöntemi olarak bilinmektedir. Bayes sınıflayıcısı Eşitlik (3)'ü kullanmaktadır (Özkan, 2008).

$$\arg \max_{C_i} \{P(X|C_i)P(C_i)\} \quad (2)$$

$$C_{MAP} = \arg \max_{C_i} \prod_{k=1}^n P(X_k|C_i) \quad (3)$$

2.1.3. Lojistik Regresyon Analizi

Lojistik regresyonda bağımlı değişkenler kategorik verilerden oluştuğu için doğrusal regresyon analizinin özel bir hali olarak ifade edilmektedir (Lewis, 2017). Lojistik regresyon analizi, veri madenciliğinde sınıflandırma amacı ile kullanılmaktadır. Bağımsız değişken sayısı p olduğunda lojistik fonksiyon Eşitlik (4) ile gösterilmekte ve değişken sayısı 1 olduğunda bu fonksiyon sigmoid fonksiyon olarak adlandırılmaktadır (Harrington, 2012).

$$P(Y = 1 | x) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}} \quad (4)$$

Y genellikle 0 ve 1 değerlerini alan kategorik bağımlı değişkeni, x açıklayıcı değişkenleri ve β model parametrelerini ifade etmektedir. Bağımsız değişkenlerin değerleri bilindiğinde bağımlı değişkenin olasılığını $P(Y | x)$ göstermektedir. Logaritmik dönüşüm uygulanarak bu ilişki Eşitlik (5)'deki gibi doğrusal bir şekilde incelenebilir.

$$\ln \left(\frac{P(Y = 1 | x)}{1 - P(Y = 1 | x)} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (5)$$

$P(Y = 1 | x)/(1 - P(Y = 1 | x))$, odds oranı olup bir olayın gerçekleşme olasılığının gerçekleşmeme olasılığına oranını ifade etmektedir (Hosmer vd., 2013).

2.1.4. Destek Vektör Makineleri

Destek vektör makinaları (Support vector machine-SVM) el yazılarını, yüz, nesne, örüntü tanıma, zaman serileri, konuşmacı tanıma metin kategorizasyonu, gibi çok çeşitli alanlarda hem sınıflandırma hem de regresyon amacı ile kullanılmaktadır (Burges, 1998; Campbell ve Yiming, 2011). Destek vektör makinaları doğrusal ve doğrusal olmayan sınıflandırma olmak üzere iki kısımda incelenmektedir (Han vd., 2012).

Doğrusal destek vektör makinelerinde sınıflandırma yapılırken girdi uzayındaki örnekler dikkate alınır ve sınıflar arasında çizilebilecek sonsuz sayıdaki hiperdüzlemler arasından sınıfları birbirinden olabildiğince uzak sınıflandıran hiperdüzlem optimum seçim kabul edilir (Cervantes vd., 2020).

Hard-Marjin Destek Vektör Makineleri: D boyutlu (nitelik) örnek sayısı L ve x_i ($i=1, \dots, L$) girdileri yalnızca iki sınıfa ait olsun. Verilerin birbirlerinden doğrusal olarak ayrıldığı varsayımına dayanarak $D=2$ olduğu durumda iki sınıf birbirinden bir doğru ile ayrılırken, $D>2$ olduğu durumda ise iki sınıf bir hiperdüzlem ile birbirlerinden ayrılabilir. w hiperdüzleminin ağırlık vektörü, $\|w\|$ ise w 'nin öklit normunu gösterirse $b/\|w\|$ hiperdüzlemden orijine dik uzaklık olmak üzere hiperdüzleme ait denklem Eşitlik (6)'daki gibi yazılabilir (Burges, 1998).

$$w * x + b = 0 \quad (6)$$

Destek vektörleri hiperdüzleme en yakın noktalar olup amacı her iki sınıfın birbirlerine yakın üyelerine hiperdüzlemi en uzak biçimde yönlendirmektir. Destek vektör makinaları w ve b 'nin seçimi olarak da ifade edilebilir (Kartal ve Balaban, 2019). Bu durumda eğitim veri seti Eşitlik (7) ile yazılabilir (Cristianini ve Taylor, 2000).

$$\forall_i \text{ için } y_i(x_i * w + b) - 1 \geq 0 \quad (7)$$

Ayrılcı hiperdüzlemin konumu da destek vektörleri ile belirlenmekte ve iki destek düzlemi arasında herhangi bir veri bulunmamaktadır. Destek düzlemleri arasındaki uzaklık $\frac{1}{\|w\|}$ 'ye eşit olup marjin olarak ifade edilmektedir. Ayrılcı optimal hiperdüzlemin destek vektörden mümkün oldukça uzak olması için marjinin maksimize edilmesi gerekmektedir (Kartal ve Balaban, 2019). $\frac{1}{\|w\|}$ ifadesinin maksimizasyonu $\|w\|$ normunun minimizasyonu ile mümkün olmaktadır (Uğuz, 2019). Bu minimizasyon yönteminin çözümünde Lagrange çarpanları kullanılmakta ve \forall_i için Lagrange çarpanları $\alpha_i \geq 0$ olur (Burges, 1998).

$$\forall_i \text{ için } y_i(x_i * w + b) - 1 \geq 0 \text{ olmak üzere; } \min \frac{1}{2} \|w\|^2 \quad (8)$$

olur. Eşitlik (8)'i minimize eden w ve b değerleri ile aynı eşitliği maksimize eden α değerinin bulunması istenmektedir ($\alpha_i \geq 0, \forall_i$). Bunun için; Karush-Kush-Tucker koşullarına bağlı kalarak sırasıyla w 'ye ve b 'ye göre Eşitlik (9)'un kısmı türevi alınıp sıfıra eşitlenmektedir.

$$L_P(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^L \alpha_i y_i(x_i * w + b) + \sum_{i=1}^L \alpha_i \quad (9)$$

Soft-Marjin Destek Vektör Makineleri: Hard-Marjin destek vektör makineleri gibi sıfır hatalı sınıflandırma yapmak yerine bazı verilerin sınırın karşı sınıfında yer almasına izin vermektedir. Bu sayede daha esnek, aşırıyı öğrenmeye (overfitting) duyarlı ve iyi bir genelleme yeteneğine sahip bir model elde edilmektedir (Uğuz, 2019).

Soft-Marjin destek vektör makinelerinde ξ_i , negatif olmayan aylak (slack) değişkenini tanımlamaktadır (Cortes ve Vapnik, 1995);

$$\forall_i \text{ için } y_i(x_i * w + b) - 1 + \xi_i \geq 0 \quad (10)$$

Bu yöntemde, ξ_i değerlerinin toplamının az olması istenmektedir (Uğuz, 2019). $\xi_i = 0$ ise örneklem doğru sınıflandırılmış, $0 < \xi_i < 1$ ise örneklem sınırın içerisinde ve doğru sınıflandırılmış, $\xi_i \geq 1$ ise örneklem yanlış sınıflandırılmıştır. Böylece ξ_i 'nin 1'den büyük olduğu örneklerin sayısı yanlış sınıflandırılan örnekleri vermektedir (Öz, 2019). Aylak değişken (ξ_i) ile marjin arasında bir denge sağlamak için bir C parametresi kullanılmaktadır. C 'nin büyük olması; daha az sayıda hatalı sınıflandırılan örnekler sağlamlasına ve marjinin küçük olmasına yol açmaktadır. C 'nin küçük olması ise tersini vermektedir. En uygun C parametresi, çapraz doğrulama ve ızgara arama yöntemleriyle belirlenmektedir (Uğuz, 2019). Hard-Marjinde tanımlanan ilk problem, Soft-Marjinde Eşitlik (11)'deki gibi ifade edilebilir (Kartal ve Balaban, 2019). $\forall_i \text{ için } y_i(x_i * w + b) - 1 + \xi_i \geq 0$ ve $\xi_i \geq 0$ olmak üzere;

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L \xi_i \quad (11)$$

olur.

Bazı durumlarda veri setleri bir düzlem ile birbirlerinden doğrusal olarak ayrılamamaktadır. Bu tarz veriler doğrusal olarak birbirinden ayrılamadığından, iki boyutlu bir uzaydan üç boyutlu bir uzaya taşınarak bir düzlem yardımıyla birbirlerinden ayrılmaktadır

(Uğuz, 2019). Verilerin daha yüksek boyutlu bir uzaya dönüştürülmesi çekirdek fonksiyonlarıyla gerçekleşmekte ve $\Phi: R^n \rightarrow H$ biçiminde haritalanmaktadır. Böylece veriler Hilbert uzayı denilen daha yüksek boyutlu uzaya yerleştirilir (Öz, 2019). $\Phi(x)$ 'in uygulanmasından sonra veriler, daha yüksek boyutlu uzayda doğrusal olarak ayrılabilen ve doğrusal bir karar sınırı bulunabilmektedir (Kartal ve Balaban, 2019). Farklı bir uzaydaki vektörlerin iç çarpımı sonuçlarını döndüren ve öznitelik uzayına taşınan vektörlerin iç çarpımını ifade eden $K(x, y)$ fonksiyonuna çekirdek fonksiyon (kernel) denilmektedir (Stoian ve Stoian, 2014).

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (12)$$

Bir doğrusal sınıflandırma probleminin amaç fonksiyonunda $\langle x_i, x_j \rangle$ şeklinde vektörlerin bir iç çarpımı yer alıyorsa, bu iç çarpım yerine uygun bir $K(x_i, x_j)$ çekirdek fonksiyonu yazılabilir. Bu durumda dual optimizasyon problemi Eşitlik (13)'deki gibi güncellenebilir (Uğuz, 2019).

$$\max_a L(a) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (13)$$

$$\text{Kısıtlar; } \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq c, \quad i = 1, \dots, n$$

Sınıflandırma performansını doğrudan etkilediği için doğrusal olmayan SVM'de çekirdek fonksiyonun seçimi çok önemlidir (Öz, 2019). Bu çalışmada gerçek veri ve simülasyon uygulamalarında, SVM algoritması için Radyal çekirdek kullanılmış ve cost hiperparametresi (0.1, 0.5, 1, 2) aralığında, gamma hiperparametresi ise (0.1, 0.5, 1, 1.5, 2) aralığında doğruluk kriterini maksimize edecek şekilde çapraz geçerlilikle optimize edilmiştir.

2.1.5. XGBoost Algoritması

Chen ve Guestrin (2016) tarafından literatüre kazandırılan Aşırı Gradyan Arttırma algoritması (Extreme Gradient Boosting) XGBoost olarak adlandırılmaktadır. Temelde Karar Ağaçları ve Gradyan Arttırma yöntemine dayanan algoritma, büyük ölçekli veri setlerinde daha yüksek sınıflandırma başarısı elde etmek için yaygın olarak kullanılmaktadır. Boosting yöntemi ile paralel işlem yaparak büyük veri setlerini hızlı bir şekilde işleyebildiği için

XGBoost, sınıflandırma problemlerinde yüksek sınıflandırma performansı sağlayabilir (Singh vd., 2022).

XGBoost, birden çok ağacın tahmin sonuçlarını toplayarak nihai bir model oluşturur. Tahmin edilen değer ile mevcut değer arasındaki hataları (artık) sürekli olarak iyileştirilmesi ve güçlü bir model elde edilmesi algoritmanın çalışma mantığını oluşturur (Dong vd., 2023). Algoritma, model performansını iyileştirmek için kayıp fonksiyonun çözümünde Newton yöntemini kullanır, kayıp fonksiyonu Taylor serisini ikinci mertebeden genişletir ve genişletilmiş seriyi minimuma götüren bir dizi iterasyon izler. Ek olarak, kayıp fonksiyonuna düzenleme terimleri eklenir. Eğitim sırasında amaç fonksiyonu, gradyan kaldırma algoritmasının kaybı ve düzenleme terimi olmak üzere iki kısımdan oluşmaktadır (Yan vd., 2022).

Veri seti: $D = (x_i, y_i)$, $i = 1, \dots, n$, $x_i \in \mathbb{R}^m, y_i \in \mathbb{R}$ varsayımına sahip ve n gözlem ile m değişken olursa, tahmin edilen \hat{y}_i değeri Eşitlik (14) ile tanımlanır (Asselman vd., 2021).

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in K \quad (14)$$

Chen ve Guestrin (2016), XGBoost algoritmasını Eşitlik (15) ile tanımlamaktadır. Gradyan artırma algoritmasının her bir iterasyonunda artık değerler önceki tahmin edicinin düzeltilmesi için manipüle edilerek kayıp fonksiyonu optimize edilmektedir. $f_k(x_i)$, k ağacın i . örneği için tahmin değerini belirtmektedir. f_k fonksiyon seti amaç fonksiyonu minimize edilerek öğrenilebilir (Chen ve Guestrin, 2016; Asselman vd., 2021).

$$Obj = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (15)$$

Eşitlik (15)'de l ile belirtilen ilk terim kayıp fonksiyonu temsil etmekte ve tahmin edilen \hat{y}_i değeri ile gerçek y_i değeri arasındaki farkı ölçmektedir. Ω ise regülasyon terimini temsil etmekte ve f_k ağacının karmaşıklığını ölçmektedir.

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \omega^2 \quad (16)$$

Burada γ ve λ parametreleri regülasyon derecelerini, T yaprak düğümleri sayısını ve ω ise her yapraktaki puanı belirtir.

XGBoost algoritmasına özgü birçok hiperparametre bulunmakta (Asselman vd., 2021) ve bunların değerleri veri setinin yapısına göre değişkenlik göstermektedir. Bu çalışmanın gerçek veri ve simülasyon uygulamalarında, XGBoost algoritması için nround hiperparametresi 100 ila 1000 aralığında 100 artacak şekilde, eta hiperparametresi (0.01, 0.05, 0.1, 0.3) aralığında çapraz geçerlilik yöntemiyle doğruluk kriterini maksimize edecek şekilde optimize edilmiştir. Diğer hiperparametreler için ise varsayılan değerler kullanılmıştır.

2.2. Çoklu Doğrusal Bağlantı

Çoklu doğrusal bağlantı (multicollinearty) çoklu doğrusal regresyon problemlerinde karşılaşılan önemli sorunlardan birisidir. Çoklu doğrusal bağlantı, bağımsız değişkenlerin kendi aralarında tam ya da güçlü ilişki yapısını ifade etmektedir (Demir, 2020). Ancak, regresyon uygulamaları bağımsız değişkenler arasında ilişki olmaması varsayımına dayalıdır. Uygulamalarda genellikle bu varsayım ihmal edilmekte ve bağımsız değişkenler kendi aralarında ilişkili olabilmektedir (Alpar, 2013). Veri setinde çoklu doğrusal bağlantı hatalı katsayılar ve bağımsız değişkenler arasında yüksek korelasyon gibi çeşitli problemlere yol açabilir. Hatalı regresyon katsayıları büyük standart hatalara neden olmakta, dolayısıyla model güvenilirliğini ve tahminleri olumsuz etkilemektedir. Bağımsız değişkenler arasındaki yüksek korelasyon ise modelde R^2 değerinin büyümesine yol açmaktadır (Chan vd., 2022).

Yapay zeka, veri madenciliği ya da makine öğrenmesi alanlarındaki çalışmalarda da çoklu doğrusal bağlantı sorunuyla karşılaşılabilen ve bu bağlantının belirlenmesi için çeşitli yöntemler bulunmaktadır. İlk yöntem, korelasyon matrisinde değişkenler arasındaki ikili korelasyon değerlerine bakmaktır. İkili değişkenler arasındaki korelasyonun 0,8'den büyük olması çoklu doğrusal bağlantı göstergesi olabilir (Mason ve Perreault, 1991).

Korelasyon matrisinin özdeğer ($\lambda_1, \lambda_2, \dots, \lambda_p$) ve özvektörleri (v_1, v_2, \dots, v_p) çoklu doğrusal bağlantıyı belirlemede kullanılan bir diğer yöntemdir. Özdeğerlerin sıfıra yakın olması çoklu doğrusal bağlantı hakkında fikir vermekte, ancak çoklu doğrusal bağlantının derecesi hakkında fikir edinmek için Eşitlik (17)'de verilen ve koşul indeksinde yer alan maksimum değeri ifade eden koşul numarasına bakılmaktadır (Alin, 2010).

$$K = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \quad (17)$$

Koşul endeksinin 10 ila 30 arasında olması orta düzeyde çoklu doğrusal bağlantıya, 30'dan büyük olması ise ciddi derecede çoklu doğrusal bağlantıya işaret etmektedir (Chan vd., 2022).

2.3. Sınıflandırma Performanslarının Değerlendirilmesi

Tablo 1'de verilen gerçek ve tahmin değerlerinden elde edilen Eşitlik (18) ile (22) arasındaki 5 metrik kullanılarak model başarıları değerlendirilebilir (Mulla vd., 2021).

Tablo 1. Karmaşıklık matrisi

	Tahmin Edilen Sınıf	
	Pozitif	Negatif
Gerçek Sınıf	DP (Doğru Pozitif)	YN (Yanlış Negatif)
	YP (Yanlış Pozitif)	DN (Doğru Negatif)

$$Doğruluk = \frac{DP + YN}{DP + DN + YP + YN} \quad (18)$$

$$Duyarlılık = \frac{DP}{DP + YN} \quad (19)$$

$$Seçicilik = \frac{DN}{DN + YP} \quad (20)$$

$$Kesinlik = \frac{DP}{DP + YP} \quad (21)$$

$$F \text{ Ölçütü} = 2 \times \frac{Duyarlılık \times Kesinlik}{Duyarlılık + Kesinlik} \quad (22)$$

3. UYGULAMA

İki gerçek veri seti ve bir simülasyon uygulaması ile çalışma gerçekleştirilmiştir. Lojistik regresyon algoritması parametrik bir model olup sıkı varsayımları bulunmakta ve bunlar arasında çoklu doğrusal bağlantı olmaması da yer almaktadır. Gerçek veri ve simülasyon çalışmalarının sonuçları yorumlanırken bu varsayım üzerinde ayrıca durulmuştur.

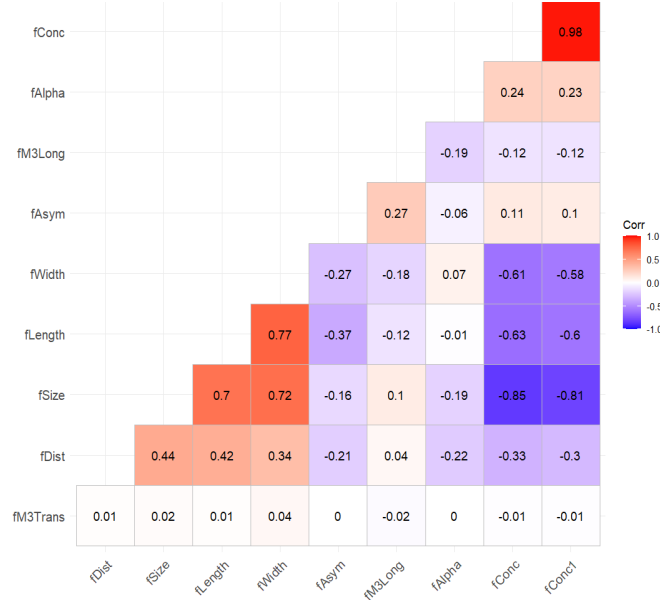
3.1. MAGIC Gamma Teleskop Veri Seti

19020 gözlem, 10 bağımsız ve 1 bağımlı değişkenden oluşmakta olan MAGIC gamma teleskop (Telescope) veri seti, <https://archive.ics.uci.edu> (UCI Machine Learning Repository) adresinden temin edilmiştir. Veri setine ait değişkenler ve özet istatistikler Tablo 2’de verilmiştir.

Tablo 2. MAGIC gamma teleskop veri setine ait değişkenler

Öznitelik Adı	Veri Türü	Nümerik=Min, Max Kategorik= Değerler	Nümerik=Ort.±s.s Kategorik= n(%)
fOdak uzunluğu	Nümerik	4.2-334.1	53.3±42.4
fOdak genişliği	Nümerik	0.0-256.3	22.2±18.3
fAlan boyutu	Nümerik	1.9-5.32	2.83±0.47
fÖzellik yoğunluğu1	Nümerik	0.01-0.89	0.38±0.18
fÖzellik yoğunluğu2	Nümerik	0.00-0.67	0.21±0.11
fAsimetri	Nümerik	-457.9-575.2	-4.33±59.2
fM3Uzunluk	Nümerik	-331.7-238.3	10.5±51.0
fM3Yatay	Nümerik	-205.8-179.8	0.25±20.8
fAlfa	Nümerik	0.-90	27.6±26.1
fMesafe	Nümerik	1.2-495.5	193.8±74.7
Bağımlı Değişken (Sınıf)	Kategorik	0, 1	0: n=12332 (%64.8) 1: n= 6688 (%35.2)

Veri setinde yer alan bağımsız değişkenlere ait korelasyon grafiği Şekil 1’de verilmiştir.



Şekil 1. MAGIC gamma teleskop veri seti korelasyon ısı grafiği

Şekil 1 incelendiğinde bağımsız değişkenler arasında pozitif ve negatif yönlerde yüksek ilişkilerin varlığı gözlenmektedir. Ayrıca çoklu doğrusal bağlantının tespiti için koşul endeksine bakılmış, en büyük özdeğer 173.588 ve en küçük özdeğer 1.683 olmak üzere koşul endeksi yaklaşık 103 olarak belirlenmiştir. Bu değer 30'dan çok büyük olduğu için veri setinde yüksek düzeyde çoklu doğrusal bağlantının olduğu söylenebilir.

MAGIC Gamma Teleskop veri seti için sınıflandırma algoritmalarına ait performans sonuçları Tablo 3'de verilmiş ve değerlendirmede beş sınıflandırma performans kriteri kullanılmıştır.

Tablo 3. MAGIC gamma teleskop veri setine ait sınıflandırma performansları

Ölçüt / Algoritma	kNN	Naive Bayes	Lojistik Reg.	SVM	XGBoost
Doğruluk	0.837	0.725	0.790	0.871	0.886
Duyarlılık	0.956	0.919	0.799	0.958	0.942
Seçicilik	0.617	0.367	0.765	0.710	0.772
Kesinlik	0.821	0.728	0.903	0.859	0.884
F-Ölçütü	0.883	0.812	0.848	0.906	0.909

Çoklu doğrusal bağlantıya sahip MAGIC gamma teleskop veri seti ile yapılan analizde en yüksek sınıflandırma performansı %88.6 doğrulukla XGBoost algoritmasına, en düşük performans ise %72.5 doğrulukla Naive Bayes algoritmasına ait olduğu görülmektedir. Ayrıca seçicilik ve F ölçütüne göre de en yüksek performansın XGBoost ve en düşük performansın ise Naive Bayes tarafından sergilendiği gözlenmiştir. Duyarlılıkta en yüksek değer SVM ve

yakın bir değer kNN ile en düşük değer ise Lojistik Regresyon ile elde edilmiştir. Kesinlik ölçütüne göre en yüksek performans Lojistik Regresyon ve en düşük performans ise Naive Bayes ile elde edilmiştir.

MAGIC gamma teleskop veri seti için bir değerlendirme yapıldığında XGBoost algoritması ile SVM algoritmasının başarılı performans sergilediği ve Naive Bayes algoritmasının düşük performansa sahip olduğu ve iyi performans sergilemediği söylenebilir.

3.2. Kredi Kartı Müşterilerinin Temerrüdü Veri Seti

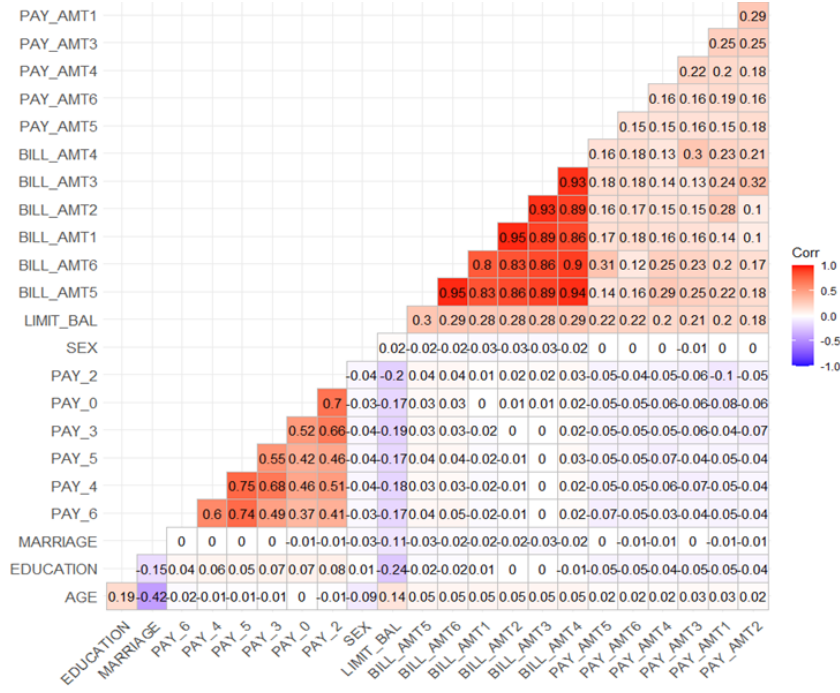
29602 gözlem, 23 bağımsız ve 1 bağımlı değişkenden oluşan Kredi Kartı Müşterilerinin Temerrüdü (Default of Credit Card Clients) veri seti, <https://archive.ics.uci.edu> (UCI Machine Learning Repository) adresinden elde edilmiştir. Veri setinde yer alan değişkenler ve bu değişkenlere ait özellikler Tablo 4'de verilmiştir.

Tablo 4. Kredi kartı veri setinde yer alan değişkenler

Öznitelik Adı	Veri Türü	Nümerik=Min, Max Kategorik= Değerler	Nümerik=Ort.±s.s Kategorik= n(%)
Limit Bakiyesi*	Nümerik	10.000-100.000	167.551±129.944
Fatura Tutarı (1, 2, 3, 4, 5, 6)	Nümerik	-69.777-1.664.089	44.820±66.655
Ödeme Tutarı (1, 2, 3, 4, 5, 6)	Nümerik	0-896.040	5.258±17.835
Ödeme (0,1,2,3,4,5,6) ¹	Kategorik	0, 1, 2, 3, 4, 5, 6, 7, 8	0: n=152.720 (%86.0) 1: n=3.696 (%2.1) 2: n=18.861 (%10.6) 3: n=1.423 (%0.8) 4: n=451 (%0.3) 5: n=135 (%0.1) 6: n=74 (%0.04) 7: n=218 (%0.1) 8: n=28 (%0.02)
Yaş	Nümerik	21-79	35.5± 9.21
Cinsiyet	Kategorik	1, 2	1: n=11.746 (%39.7) 2: n=17.855 (%60.3)
Eğitim ¹	Kategorik	1, 2, 3, 4	1: n=10.581 (%35.7) 2: n=14.024 (%4.4) 3: n=4.873 (%16.5) 4: n=123 (%0.4)
Evlilik ¹	Kategorik	1, 2, 3	1: n=13.477 (%45.5) 2: n=15.806 (%53.4) 3: n=318 (%1.1)
Bağımlı Değişken (Ödeme durumu)	Kategorik	0, 1	0: n=22.996 (%77.7) 1: n=6.605 (%22.3)

Kredi kartı veri setinde yer alan bağımsız değişkenlere ait korelasyon grafiği Şekil 2’de verilmiştir.

Şekil 2 incelendiğinde bazı bağımsız değişkenler arasında pozitif yönde yüksek derecede ilişkilerin olduğu gözlenmektedir. Ayrıca bu veri seti için koşul endeksi yaklaşık 893 olarak hesaplanmış ve böylece veri setinde çok yüksek düzeyde çoklu doğrusal bağlantı olduğu söylenebilmektedir.



Şekil 2. Kredi kartı veri setinde bağımsız değişkenlere ait korelasyon ısı grafiği

Kredi kartı veri seti için sınıflandırma algoritmalarına ait performans sonuçları Tablo 5’de verilmiş ve değerlendirmede beş sınıflandırma performans değerlendirme kriteri kullanılmıştır.

Tablo 5. Kredi kartı veri setine ait sınıflandırma performansları

Ölçüt / Algoritma	kNN	Naive Bayes	Lojistik Reg.	SVM	XGBoost
Doğruluk	0.816	0.771	0.823	0.822	0.826
Duyarlılık	0.945	0.828	0.837	0.955	0.952
Seçicilik	0.368	0.572	0.707	0.360	0.389
Kesinlik	0.839	0.871	0.958	0.839	0.844
F-Ölçütü	0.889	0.849	0.894	0.893	0.895

Tablo 5 incelendiğinde, en yüksek sınıflandırma performansının %82.6 doğrulukla XGBoost algoritmasına ve en düşük performansın ise %77.1 doğrulukla Naive Bayes algoritmasına ait olduğu görülmektedir. Duyarlılığa göre en yüksek değer SVM ve en düşük değer ise Naive Bayes ile sağlanmıştır. Seçicilik ve kesinlik ölçütlerine göre en yüksek değer Lojistik Regresyon ile elde edilirken, seçicilik için en düşük değer SVM ile kesinlik için ise SVM ve kNN ile elde edilmiştir. F Ölçütüne göre algoritmaların birbirine çok yakın değerler verdiği ve en yüksek performans %89.5 değeriyle XGBoost algoritmasından, en düşük performans ise %84.9 değeriyle Naive Bayes algoritmasından elde edilmiştir.

Kredi kartı veri seti için bir değerlendirme yapıldığında XGBoost, SVM ve Lojistik Regresyon algoritmalarının başarılı performans sergilediği söylenebilir. Naive Bayes algoritması bu veri seti için de düşük performans gösterdiği söylenebilir.

3.3. Simülasyon Çalışması

Simülasyon çalışması için veri üretim mekanizması Tablo 6’da sunulmuştur. Her bir simülasyon senaryosu için tekrar sayısı 500 olarak kararlaştırılmıştır. Simülasyon çalışmalarında bağımlı değişkenin iki sınıf içerdiği varsayılmıştır. Simülasyon kurgusu göz önüne alındığında, sınıflandırma algoritmalarının performansları için üç farklı çoklu doğrusal bağlantı oranı (ÇDB), iki farklı örneklem büyüklüğü ve iki farklı bağımsız değişken sayısı dikkate alınmıştır. Eğitim ve test veri setleri gerçek veri setlerinde olduğu gibi sırasıyla %75 ve %25 olarak alınmıştır. Sonuçlar, gerçek veri sonuçlarıyla karşılaştırılarak uyum ve uyumsuzluklar tartışılmıştır.

Tablo 6. Simülasyon kurgusu ve veri üretimi

Simülasyon kurgusu		
Örneklem büyüklüğü	Çoklu doğrusal bağlantı oranı	Açıklayıcı değişken sayısı
$n = 50, 150$	$\rho = (0.70, 0.85, 0.99)$	$p = (5, 10)$
Verilerin üretilmesi		
Bağımsız değişkenler	Olasılıklar üretilmesi (sınıflar için)	Kategorik bağımlı değişken
$X \sim MV[\mu_x, \Sigma_x] \in \mathbb{R}^{n \times p}$	$z = 1 + \theta_p X,$ $P_y = \frac{1}{(1 + e^{-z})}$	$y = binom(n, P_y)$

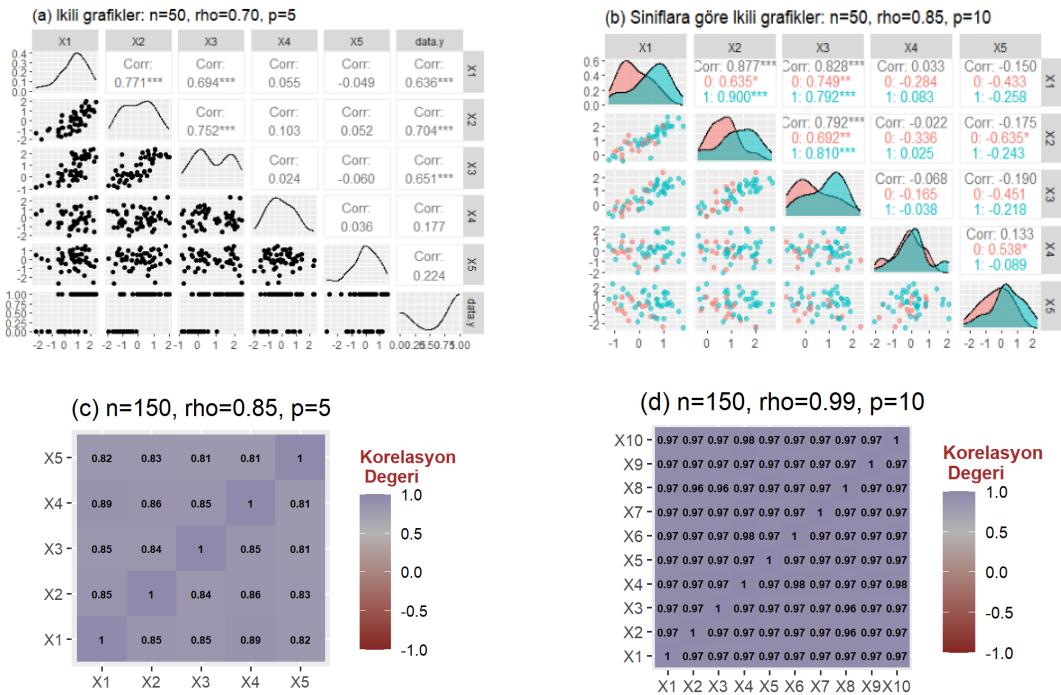
Üretilen verilerin, çoklu doğrusal bağlantı durumlarını gözlemleyebilmek için Şekil 3 verilmiştir. Şekil 3, her biri farklı simülasyon konfigürasyonunu temsil eden dört panelden

oluşmaktadır. Ayrıca çoklu doğrusal bağlantı içeren açıklayıcı değişkenlerin üretimi aşağıdaki gibi ifade edilebilir.

$$X \sim MV(\mu_x, \Sigma_x)$$

Burada, $\{\mu_i \sim U[0,1]\}_{i=1}^p$ ve $\Sigma_x = \begin{bmatrix} 1 & \rho \Sigma_{ij} \\ \rho \Sigma_{ij} & 1 \end{bmatrix}$ şeklinde olup $MV(\mu_x, \Sigma_x)$, μ_x ortalama

vektörüne ve Σ_x varyans kovaryans matrisine sahip çok değişkenli normal dağılıma sahip olduğunu gösteriyor. Çoklu doğrusal bağlantı üretilirken Σ_x köşegenleri bir ve kovaryansları korelasyon değerleri ile orantılı olacak şekilde dizayn edilmektedir. Çoklu doğrusal bağlantı içermeyen veri üretilirken ρ değeri sıfır veya çok küçük bir değer verilerek elde edilir ve veri üretimi aynı kalır. Şekil 3 incelendiğinde, veri setinin tamamı için ikili korelasyon değerleri ve ilişkilerin yoğunlukları Panel (a) ile ve bağımlı değişkene göre sınıfları içeren ikili grafikler ise Panel (b) ile gösterilmiştir. Panel (c), 0.85 korelasyon ile çoklu doğrusal bağlantı seviyeleri ve panel (d) 0.99 korelasyon ile çoklu doğrusal bağlantı seviyeleri için sırasıyla küçük ve büyük örneklem hacimlerine göre veri dağılımlarını göstermektedir. Bağımsız değişkenler arasında çoklu doğrusal bağlantı bulunmakta ve her bir değişkenin aynı dağılımdan üretildiği düşüncesiyle, temsilen sadece biri için saçılım grafiği verilmiştir. Grafikler incelendiğinde bağımsız değişkenler arasında güçlü bir ilişkinin olduğu görülmektedir.



Şekil 3. Çoklu doğrusal bağlantı için üretilen verilere ait tanımlayıcı grafikler

Tablo 7’de simülasyon çalışması için üretilen çoklu doğrusal bağlantılı veri setlerine ait koşul endeks değerleri yer almaktadır.

Tablo 7. Çoklu doğrusal bağlantılı üretilen veriler için koşul endeks değerleri

Simülasyon seti			Koşul Endeksi
n	ρ	p	
50	0.70	5	18.847
50	0.85	5	32.185
50	0.99	5	50.517
50	0.70	10	20.884
50	0.85	10	46.131
50	0.99	10	80.801
150	0.70	5	13.583
150	0.85	5	28.646
150	0.99	5	44.457
150	0.70	10	20.929
150	0.85	10	50.207
150	0.99	10	65.683

Tablo 7 incelendiğinde, veri setlerinde çoklu doğrusal bağlantı olduğu gözlenmektedir. Veri setindeki değişkenler arasındaki ilişki arttıkça çoklu doğrusal bağlantıda doğru orantılı olarak artmaktadır.

Sınıflandırma algoritmalarının genel performanslarına ait olası tüm simülasyon sonuçları Tablo 8’de verilmektedir.

Tablo 8. Çoklu doğrusal bağlantı simülasyon sonuçları

n	p	Ölçüt / Alg.	kNN	Naive Bayes	Lojistik Reg.	SVM	XGB	kNN	Naive Bayes	Lojistik Reg.	SVM	XGB	
50	5		<i>ÇDB yok</i>					$\rho = 0.70$					
		Doğruluk	0.611	0.653	0.858	0.650	0.635	0.573	0.635	0.777	0.510	0.610	
		Duyarlılık	0.936	0.913	0.876	0.938	0.944	0.913	0.878	0.892	0.972	0.930	
		Seçicilik	0.735	0.848	0.681	0.887	0.833	0.714	0.608	0.693	0.857	0.750	
		Kesinlik	0.812	0.689	0.722	0.876	0.804	0.800	0.666	0.697	0.857	0.782	
		F-Ölçütü	0.774	0.783	0.867	0.624	0.789	0.743	0.756	0.835	0.741	0.720	
			$\rho = 0.85$					$\rho = 0.99$					

150	10	Doğruluk	0.760	0.855	0.891	0.673	0.851	0.891	0.900	0.921	0.735	0.771	
		Duyarlılık	0.944	0.860	0.977	0.962	0.896	0.936	0.866	0.977	0.976	0.896	
		Seçicilik	0.925	0.726	0.946	0.829	0.831	0.900	0.817	0.950	0.950	0.816	
		Kesinlik	0.810	0.694	0.905	0.804	0.823	0.874	0.835	0.926	0.803	0.774	
		F-Ölçütü	0.852	0.857	0.934	0.817	0.873	0.913	0.883	0.949	0.855	0.834	
			<i>ÇDB yok</i>					$\rho = 0.70$					
	Doğruluk	0.646	0.596	0.745	0.583	0.540	0.496	0.641	0.718	0.606	0.676		
	Duyarlılık	0.744	0.770	0.938	0.900	0.716	0.882	0.884	0.926	0.914	0.817		
	Seçicilik	0.721	0.628	0.888	0.854	0.503	0.819	0.824	0.845	0.827	0.654		
	Kesinlik	0.687	0.606	0.787	0.790	0.508	0.786	0.803	0.758	0.781	0.650		
	F-Ölçütü	0.695	0.683	0.842	0.691	0.628	0.689	0.762	0.822	0.781	0.746		
			$\rho = 0.85$					$\rho = 0.99$					
	Doğruluk	0.618	0.751	0.855	0.526	0.678	0.415	0.618	0.791	0.520	0.470		
	Duyarlılık	0.966	0.888	0.940	0.903	0.862	0.880	0.892	0.868	0.805	0.894		
	Seçicilik	0.902	0.815	0.866	0.869	0.725	0.871	0.833	0.701	0.806	0.723		
	Kesinlik	0.820	0.675	0.848	0.815	0.675	0.793	0.654	0.720	0.789	0.719		
	F-Ölçütü	0.792	0.820	0.897	0.642	0.770	0.647	0.755	0.830	0.604	0.682		
	5	10		<i>ÇDB yok</i>					$\rho = 0.70$				
			Doğruluk	0.827	0.754	0.893	0.712	0.785	0.749	0.851	0.838	0.726	0.721
			Duyarlılık	0.926	0.956	0.951	0.928	0.922	0.938	0.893	0.962	0.943	0.932
Seçicilik			0.781	0.869	0.830	0.736	0.773	0.846	0.781	0.918	0.850	0.838	
Kesinlik			0.734	0.706	0.849	0.640	0.709	0.778	0.808	0.874	0.772	0.820	
F-Ölçütü		0.826	0.805	0.922	0.770	0.804	0.844	0.872	0.900	0.834	0.876		
			$\rho = 0.85$					$\rho = 0.99$					
Doğruluk		0.900	0.934	0.866	0.820	0.798	0.791	0.888	0.898	0.851	0.820		
Duyarlılık		0.952	0.906	0.962	0.963	0.933	0.956	0.859	0.953	0.960	0.940		
Seçicilik		0.894	0.812	0.923	0.919	0.843	0.898	0.746	0.888	0.916	0.881		
Kesinlik		0.892	0.867	0.890	0.854	0.814	0.837	0.804	0.889	0.878	0.843		
F-Ölçütü		0.926	0.920	0.914	0.892	0.865	0.873	0.873	0.925	0.906	0.880		
10			<i>ÇDB yok</i>					$\rho = 0.70$					
		Doğruluk	0.523	0.623	0.893	0.587	0.615	0.690	0.741	0.898	0.517	0.701	
		Duyarlılık	0.960	0.948	0.949	0.851	0.874	0.935	0.903	0.953	0.868	0.927	
	Seçicilik	0.916	0.892	0.874	0.795	0.701	0.848	0.815	0.909	0.776	0.851		

Kesinlik	0.627	0.700	0.881	0.760	0.634	0.758	0.769	0.902	0.753	0.749
F-Ölçütü	0.742	0.785	0.921	0.674	0.744	0.812	0.822	0.925	0.705	0.814
	$\rho = 0.85$					$\rho = 0.99$				
Doğruluk	0.647	0.844	0.892	0.671	0.713	0.727	0.834	0.884	0.642	0.837
Duyarlılık	0.941	0.867	0.944	0.991	0.901	0.898	0.834	0.943	0.939	0.866
Seçicilik	0.868	0.743	0.879	0.824	0.763	0.842	0.761	0.910	0.907	0.811
Kesinlik	0.729	0.783	0.881	0.786	0.734	0.772	0.792	0.894	0.558	0.820
F-Ölçütü	0.794	0.856	0.918	0.731	0.807	0.813	0.834	0.913	0.696	0.852

Tablo 8'deki simülasyon sonuçları $n=50$ ve $p=5$ çoklu doğrusal bağlantı olmayan durum için incelendiğinde; en yüksek doğruluk ve F-Ölçütü değerlerine sahip algoritma Lojistik Regresyon algoritmasıdır. Seçicilik ve kesinlik açısından SVM, duyarlılık açısından ise XGB en yüksek değere sahiptir. Çoklu doğrusal bağlantı varlığında ise; Lojistik Regresyon, Naive Bayes ve SVM algoritmalarının performansı çoklu doğrusal bağlantının artmasıyla birlikte genellikle arttığı gözlenmiştir. kNN ise çoklu doğrusal bağlantının artmasıyla birlikte bazı metriklerde performansını artırabilirken, bazılarında ise performanslarını düşürebildiği gözlenmiştir. Ayrıca ρ değeri 0.7'den 0.85'e çıkarılırken XGB performansı duyarlılık hariç diğer tüm metriklerde artarken 0.99'a çıkarıldığında ise duyarlılık için değişmezken diğer tüm metrikler için performanslarda düşüş gözlemlenmiştir. Böylece, çoklu doğrusal bağlantının bazı algoritmaların performansını olumlu ve bazılarını ise olumsuz etkilediği gözlenmiştir.

$n=50$ ve $p=10$ çoklu doğrusal bağlantı olmayan durum için incelendiğinde; tüm metriklerde en yüksek değere sahip algoritmanın Lojistik Regresyon olduğu ve en düşük değere sahip algoritmanın XGBoost algoritması olduğu söylenebilir. Çoklu Doğrusal Bağlantı varlığında ise: Lojistik Regresyon algoritması, çoğu metrikte en yüksek değeri vermektedir. Çoklu doğrusal bağlantı oranını gösteren ρ değeri 0.7'den 0.85'e çıkarıldığında kNN, Lojistik Regresyon ve XGBoost performanslarının tüm metriklere göre arttığı, Naive Bayes performansının ise Seçicilik ve Kesinlik metriğine göre düştüğü geri kalan diğer üç metriğe göre de arttığı görülmüştür. SVM algoritması ise Naive Bayes'in tam tersine bir davranış sergilemiştir. Ancak ρ değeri 0.85'ten 0.99'a çıkarıldığında duyarlılık ve seçicilik metriklerine göre Naive Bayes algoritmasının, duyarlılık ve kesinlik metriklerine göre ise XGBoost algoritmasının performansında hafif bir artış gözlenirken, geriye kalan metriklere göre ise bu iki algoritmanın performansında ciddi bir düşüş gözlenmiştir. Ayrıca kNN,

Lojistik Regresyon ve SVM algoritmaları tüm metriklerde farklı düzeylerde performans kaybına uğramıştır. Hem bağımsız değişken sayısındaki (p) artış hem de çoklu doğrusal bağlantı oranındaki (ρ) artış birlikte değerlendirildiğinde bu artışların algoritmaların performansı üzerinde genel olarak olumsuz etkiye sahip olduğu söylenebilir.

$n=150$ ve $p=5$ çoklu doğrusal bağlantı olmayan durum için incelendiğinde; Lojistik Regresyon Doğruluk, Kesinlik ve F-Ölçütü metriklerinde ve Naive Bayes algoritması ise Duyarlılık ve Seçicilik metriklerde en yüksek değerlere sahip algoritmalarıdır. Çoklu Doğrusal Bağlantı varlığında ise: Lojistik Regresyon birçok metrikte en yüksek değerlere sahip algoritmadır. Zira ρ , 0.70 olduğunda sadece Doğruluk metriğine göre Naive Bayes; 0.85 olduğunda Duyarlılık, Kesinlik ve F-Ölçütüne göre kNN, Doğruluk metriğine göre ise Naive Bayes; 0.99 olduğunda Duyarlılık ve Seçicilik metriklerine göre SVM en yüksek değerlere sahipken, geriye kalan tüm durumlarda Lojistik Regresyon en yüksek değerlere sahiptir. Çoklu doğrusal bağlantı arttıkça çoğu metriğe göre SVM algoritması performans artışı göstermektedir. ρ değeri 0.7'den 0.85'e çıkarıldığında Naive Bayes algoritması tüm metriklere göre performans artışı gösterirken 0.85'den 0.99'a çıkarıldığında ise tüm metriklere göre performans düşüşü göstermektedir. Naive Bayes algoritmasına çok yakın performans değişimi kNN algoritması içinde gözlenmiştir. Çoklu doğrusal bağlantı arttıkça XGBoost algoritması neredeyse tüm metriklerde daha kararlı bir duruş sergilemiş ve performansında önemli değişiklik gözlemlenmemiştir.

$n=150$ ve $p=10$ çoklu doğrusal bağlantı olmayan durum için incelendiğinde; Lojistik Regresyon doğruluk, kesinlik ve F-Ölçütü metriklerinde yüksek performans göstermiştir. kNN ise seçicilik ve duyarlılık metriklerinde en yüksek performansı göstermiştir. Çoklu doğrusal bağlantı varlığında ise: ρ , 0.85 olduğu durumda sadece Duyarlılık metriğine göre SVM, hem geriye kalan diğer metriklere göre hem de ρ , 0.70 ve 0.99 olduğu durumlarda tüm metriklere göre Lojistik Regresyon en yüksek değerlere sahip algoritma olduğu gözlemlenmiştir. Çoklu doğrusal bağlantı arttıkça kNN ve Naive Bayes algoritmalarının performanslarında düzensiz dalgalanmalar yaşanmış ve bu dalgalanmalar metrikten metriğe farklılık göstermiştir. ρ değeri 0.7'den 0.85'e çıkarıldığında SVM algoritması tüm metriklere göre performans artışı gösterirken, 0.85'den 0.99'a çıkarıldığında ise seçicilik hariç geriye kalan diğer tüm metriklere göre performans düşüşü göstermiştir. Çoklu doğrusal bağlantı arttıkça XGBoost algoritmasının; doğruluk metriğine göre performans artışı, duyarlılık metriğine göre performans düşüşü gösterdiği ve diğer üç metriğe göre ise performans artış ve düşüşlerinde bir kararsızlığın olduğu belirlenmiştir.

Simülasyon sonuçları için algoritmalar özelinde genel bir değerlendirme yapıldığında: Lojistik Regresyon, tüm n ve p değerleri için genel olarak en yüksek performans gösteren algoritma olmuştur. Özellikle doğruluk ve F-Ölçütü açısından oldukça başarılı sonuçlar vermiştir. XGBoost algoritmasının büyük örneklem hacminde genel olarak daha iyi sonuçlar verdiği gözlenmiştir. XGBoost algoritmasına ait bu sonuç, Tablo 3 ve Tablo 5'te verilen sonuçlarla genel olarak uyumsuzluk göstermektedir. kNN, genel olarak daha düşük performans göstermiş ve özellikle örneklem hacmi artırıldığında ($p=10$ olduğunda) performansı oldukça düşmüştür. Naive Bayes, çoklu doğrusal bağlantı varlığında özellikle doğruluk ve F-Ölçütü bakımından kNN'e göre genellikle daha iyi performans sergilemektedir. SVM algoritmasının performansı genel olarak Lojistik Regresyon ve XGBoost algoritmaları performansına yetişmezse de seçicilik açısından bu algoritmalarından aşağı olmadığı söylenebilir.

4. SONUÇ

Çoklu doğrusal bağlantıya sahip iki gerçek veri seti ve simülasyon çalışmaları ile gerçekleştirilen bu çalışmadan elde edilen sonuç ve öneriler aşağıda verilmiştir.

MAGIC gamma teleskop veri kümesi için, XGBoost algoritması ile SVM algoritmasının başarılı performans sergilediği, en düşük performansın ise Naive Bayes algoritması tarafından gösterildiği söylenebilir.

Kredi kartı veri setine ait sonuçlar incelendiğinde; XGBoost, SVM ve Lojistik Regresyon algoritmalarının başarılı performans gösterdiği, Naive Bayes algoritmasının düşük performans gösterdiği söylenebilir. Bu sonucun, MAGIC gamma teleskop veri seti ile uyumlu olduğu söylenebilir.

Simülasyon çalışmaları, kNN ve Lojistik Regresyon algoritmalarının çoklu doğrusal bağlantı karşısında farklı tepkiler verdiğini göstermiştir. Küçük örneklemler ve yüksek bağımsız değişken sayısında kNN algoritmasının performansı olumsuz etkilenirken, Lojistik Regresyon algoritması bu duruma karşı daha dirençlidir ve artan çoklu doğrusal bağlantı oranına bağlı olarak birçok metrikte iyileşme gösterir. Bu durum, Lojistik Regresyonun varsayımları arasında yer alan çoklu doğrusal bağlantı olmaması varsayımının ne kadar önemli olduğunu ve ihmal edilmemesi gerektiğini açıkça göstermektedir. Aksi takdirde yanıltıcı sonuçlar elde edilebilir. Örneklem hacmi artmaya başladığında ise XGBoost algoritması artan

çoklu doğrusal bağlantı ile birlikte başarılı sonuçlar vermektedir. Bu durum, XGBoost algoritması için gerçek verilerden elde edilen bulgularla uyumlu olduğunu göstermektedir.

Sonuç olarak, örneklem büyüklüğü yeterli seviyelere ulaştığında çoklu doğrusal bağlantıya sahip veri setlerinde sınıflandırma için XGBoost algoritmasının iyi bir alternatif olduğu söylenebilir. Derraz vd. (2023), çoklu doğrusal bağlantı durumunda topluluk öğrenme algoritmalarının geleneksel makine öğrenmesi yöntemlerinden daha başarılı performans sergilediğini belirtmişler. Georganos vd. (2018) ile McNamara vd. (2022), örneklem hacminin artmasıyla XGBoost algoritması performansının da arttığını ifade etmişlerdir. Çalışma ilgili çalışmalara ait sonuçlar ile benzerlik göstermektedir. Naive Bayes ile kNN algoritmasını çoklu doğrusal bağlantıdan olumsuz etkilendiği söylenebilir. Bu çalışmadan elde edilen sonuçlara benzer şekilde Dumancas ve Bello (2015) tarafından, çoklu doğrusal bağlantıdan en çok etkilenen algoritmanın Naive Bayes olduğu gösterilmiştir.

Gelecekteki çalışmalarda çoklu doğrusal bağlantı çözümü daha kapsamlı bir şekilde ele alınmalıdır. Bu çalışmanın sonuçları, kullanılan veri setleri ve algoritmalar ile sınırlandırılabilir. Daha genel sonuçlara varabilmek için farklı veri setleri ve algoritmalar üzerinde daha fazla araştırma yapılması gerekmektedir.

ETİK BEYAN

“Çoklu Doğrusal Bağlantı Olması Durumunda Veri Madenciliği Algoritmaları Performanslarının Karşılaştırılması” başlıklı çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş, toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

KAYNAKÇA

Alin, A. (2010), Multicollinearity, *Wiley Interdisciplinary Reviews Computational Statistics*, 2(3), 370–374.

Alpar, R. (2013), *Çok değişkenli istatistiksel yöntemler*, Detay Yayıncılık: Ankara, Türkiye.

Asselman, A., Khaldi, M. and Aammou, S. (2021), Enhancing the prediction of student performance based on the machine learning xgboost algorithm, *Interactive Learning Environments*, 1–20.

- Batista, G. E. A. P. A. and Monard, M. C. (2002), A study of k-nearest neighbour as an imputation method. In Abraham, A., Solar, J.R., Köppen, M. (Ed.), *Frontiers in artificial intelligence and applications*, 87, 251–260, IOS Press.
- Blommaert, A., Hens, N. and Beutels, P. (2014), Data mining for longitudinal data under multicollinearity and time dependence using penalized generalized estimating equations, *Computational Statistics & Data Analysis*, 71(0), 667–680.
- Burges, C. J. (1998), A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, 2(2), 121–167.
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. and Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189–215.
- Chan, J.-L., Leow, S., Bea, K., Cheng, W., Phoong, S., Hong, Z.-W. and Chen, Y. L. (2022), Mitigating the multicollinearity problem and its machine learning approach: A review, *Mathematics*, 10(8), 1283.
- Chen, T. and Guestrin, C. (2016), XGBoost: A scalable tree boosting system, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA.
- Cortes, C. and Vapnik, V. N. (1995), Support vector networks, *Machine Learning*, 20, 273–297.
- Cristianini, N. and Taylor, J. S. (2000), *An introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press: Cambridge, UK.
- Davidson, I. and Tayi, G. (2009), Data preparation using data quality matrices for classification mining, *European Journal of Operational Research*, 197(2), 764-772.
- Demir, Y. (2020), Çoklu doğrusal regresyon ve bazı cezalı tahmin yöntemlerinin incelenmesi. In S. Öztürk (Ed.), *Sosyal ve beşeri bilimlerde teori ve araştırmalar II*, 2, 261-276, Gece Akademi: Ankara.
- Derraz, R., Melissa Muharam, F., Nurulhuda, K., Ahmad Jaafar, N. and Keng Yap, N. (2023), Ensemble and single algorithm models to handle multicollinearity of UAV vegetation indices for predicting rice biomass, *Computers and Electronics in Agriculture*, 205, 107621.

- Dong, Z., Li, X., Luan, F., Ding, J. and Zhang, D. (2023), Point and interval prediction of the effective length of hot-rolled plates based on IBES-XGBoost, *Measurement*, 214(0), 112857.
- Dumancas, G. and Bello, G. (2015), Comparison of machine-learning techniques for handling multicollinearity in big data analytics and high performance data mining, The International Conference for High Performance Computing, Networking, Storage, and Analysis, Texas, USA.
- Garg, A. and Tai, K. (2013), Comparison of statistical and machine learning methods in modelling of data with multicollinearity, *International Journal of Modelling, Identification and Control*, 18(4), 295–312.
- Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M. and Wolff, E. (2018), Very high resolution object-based land use–land cover urban classification using extreme gradient boosting, *IEEE Geoscience and Remote Sensing*, 15(4), 607-611.
- Han, J., Kamber, M. and Pei, J. (2012), *Data mining concepts and techniques* (Third Edition). Morgan Kaufman Publishers: Massachusetts, USA.
- Harrington, P. (2012), *Machine learning in action*, Manning Publications: New York, USA.
- Hosmer, D. W., Lemeshov, S. and Sturdivant, R. X. (2013), *Applied logistic regression* (Third Edition). John Wiley & Sons, Inc: New Jersey, USA.
- Kartal, E. and Balaban, M. E. (2019), Destek vektör makineleri: teori ve R dili ile bir uygulama. In M. E. Balaban, E. Kartal (Eds.), *Veri madenciliği ve makine öğrenmesi temel kavramlar, algoritmalar, uygulamalar* (207-241), Çağlayan Kitapevi: İstanbul.
- Lewis, N. D. (2017), *Machine learning made easy with R: An intuitive step by step blueprint for beginners*, CreateSpace Independent Publishing Platform: Carolina, USA.
- Mason, C. H. and Perreault, W. D. (1991), Collinearity, power, and interpretation of multiple regression analysis, *Journal of Marketing Research*, 28(3), 268–280.
- McNamara, J. M., Green, R. F. and Olsson, O. (2006). Bayes' Theorem and its applications in animal behaviour, *Oikos*, 112(2), 243–251.
- McNamara, M. E., Zisser, M., Beevers, C. G. and Shumake, J. (2022), Not just “big” data: Importance of sample size, measurement error, and uninformative predictors for developing prognostic models for digital interventions, *Behaviour Research and*

Therapy, 153(0), 1-12.

- Mucherino, A., Papajorgji, P. J. and Paradalos, P. M. (2009), *Data mining in agriculture*, Springer: Dordrecht, Hollanda.
- Mulla, G. A. A., Demir, Y. and Hassan, M. (2021), Combination of PCA with SMOTE Oversampling for Classification of High-Dimensional Imbalanced Data, *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 10(3), 858–869.
- Obite, C. P., Olewuezi, N. P., Ugwuanyim, G. U. and Bartholomew, D. C. (2020), Multicollinearity effect in regression analysis: A feed forward artificial neural network approach, *Asian Journal of Probability and Statistics*, 6(1), 22-33.
- Öz, E. (2019), Destek vektör makineleri. In S. Alp, E. Öz (Ed.), *Makine öğreniminde sınıflandırma yöntemleri ve R uygulamaları* (67-189), Nobel Akademik Yayıncılık: Ankara.
- Rahman, M. M., Ghasemi, Y., Suley, E., Zhou, Y., Wang, S. and Rogers, J. (2021), Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features, *IRBM*, 42(4), 215-226.
- Roelofs, R., Shankar, V., Recht, B., Fridovich-Keil, S., Hardt, M., Miller, J. and Schmidt, L. (2019), A meta-analysis of overfitting in machine learning. *Advances in Neural Information Processing Systems*, 32.
- Senawi, A., Wei, H.-L. and Billings, S. A. (2017), A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking, *Pattern Recognition*, 67, 47-61.
- Silahtaroglu, G. (2013), *Veri Madenciliği Kavram ve Algoritmaları*, Papatya Yayınevi: İstanbul.
- Singh, R., Biswas, M. and Pal, M. (2022), Cloud detection using sentinel 2 imageries: A comparison of XGBoost, RF, SVM, and CNN algorithms. *Geocarto International*, 0(0), 1–32.
- Stoean, C., Stoean, R. (2014), *Evolutionary support vector machines and their application for classification*, Springer International Publishing: New York, USA.
- Uğuz, S. (2019), *Makine öğrenmesi teorik yönleri ve python uygulamaları* (1. Basım). Nobel Akademik Yayıncılık: Ankara.

- Urooj, B., Shah, M. A., Maple, C., Abbasi, M. K., Riasat, S. (2022), Malware detection: a framework for reverse engineered android applications through machine learning algorithms, *IEEE Access*, 10(6), 89031-89050.
- Yan, Z., Chen, H., Dong, X., Zhou, K. and Xu, Z. (2022), Research on prediction of multi-class theft crimes by an optimized decomposition and fusion method based on XGBoost, *Expert Systems with Applications*, 207, 117943.
- Ying, X. (2019), An overview of overfitting and its solutions, In *Journal of physics: Conference series*, 1168, 022022, IOP Publishing.
- Zhang, X., Liu, S. and Zheng, X. (2021), Stock Price Movement Prediction Based on a Deep Factorization Machine and the Attention Mechanism, *Mathematics*, 9(8), 800.
- Zhu, J., Ge, Z., Song, Z. and Gao, F. (2018), Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data, *Annual Reviews in Control*, 46(1), 107–133.

Yayın Geliş Tarihi (Submitted): 07/10/2023

Yayın Kabul Tarihi (Accepted): 10/05/2024

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

Please Cite As/Atıf için:

Uluskan, M. ve Şenli, H. D. (2024), YSA sınıflandırma modellerinde korelasyon-hipotez testi tabanlı filtreleme yoluyla girdi seçimi, *Nicel Bilimler Dergisi*, 6(1), 68-102. doi: 10.51541/nicel.1372774

YSA SINIFLANDIRMA MODELLERİNDE KORELASYON-HİPOTEZ TESTİ TABANLI FİLTRELEME YOLUYLA GİRDİ SEÇİMİ

Meryem Uluskan¹ ve Halil Derya Şenli²

ÖZ

Bu çalışmada başlıca amaç, yüksek miktardaki olası girdi değişken sayısını, bu değişkenler arasındaki korelasyonları göz önünde bulundurarak azaltarak sınıflandırma performansı yüksek Yapay Sinir Ağı (YSA) modelleri elde etmektir. Bunu gerçekleştirmek için 30 adet olası girdi değişkeni olan bir meme kanseri teşhis problemi ele alınmış ve önerilen korelasyon-hipotez testi tabanlı bir filtreleme yöntemi ile girdi değişken sayısı azaltılarak YSA modeli oluşturulmuştur. Önerilen modelin etkinliği farklı girdi değişken setlerini içeren altı YSA modeli ile karşılaştırılmıştır. Bu altı model, tüm girdi değişkenlerini içeren modelle, model tabanlı seçim yöntemlerinden aşamalı regresyon, ileri doğru seçim ve geriye doğru eleme yöntemleri ile seçilmiş girdi değişkenleriyle elde edilmiş olan modelleri kapsamaktadır. Modeller oluşturulurken veri seti farklı eğitim-test yüzdelerine bölünmüş ve gizli katmanda farklı nöron sayıları denenmiştir. Modellerin sınıflandırma performanslarını karşılaştırmak için doğruluk, duyarlılık, kesinlik ve F1-skoru ölçütleri kullanılmıştır. Sonuç olarak, önerilen korelasyon tabanlı filtreleme yöntemi ile seçilen dokuz girdi değişkenli modeller için doğruluk değeri 0,93-0,95 arasında bulunmuş olup bu değer belirgin şekilde iyidir. Duyarlılık değeri modelimiz için 0,85-0,88 aralığında ve yeterli düzeyde elde edilmiştir. Kesinlik değerinin önerilen modelimiz için 0,98-0,988 aralığında ve çok yüksek olduğu belirlenmiştir. Bu çalışmada önerilen modelin F1-skoru 0,907-0,931 arasında olup yeterince yüksek bir

¹Sorumlu yazar, Doç. Dr., Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye. ORCID ID: <https://orcid.org/0000-0003-1287-8286>

² Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye. ORCID ID: <https://orcid.org/0000-0001-6966-3388>

değere sahiptir. Karşılaştırılan modeller içinde önerilen dokuz girdi değişkenli modelin değişken sayısının en düşük olduğu, yani en sade model olduğu ve gizli katmanda sadece 10 nöronla bile iyi bir sınıflandırma performansına sahip olduğu göz önüne alındığında bu yöntemin özellikle model tabanlı yöntemlere kıyasla kısa sürede ve düşük maliyetlerle anlaşılır sınıflandırma modelleri oluşturmada verimli olacağı belirlenmiştir.

Anahtar Kelimeler: Girdi değişken seçimi, Filtreleme yöntemi, Yapay Sinir Ağları, Sınıflandırma, Korelasyon, Hipotez testleri.

INPUT SELECTION THROUGH CORRELATION-HYPOTHESIS TESTING BASED FILTERING IN ANN CLASSIFICATION MODELS

ABSTRACT

The main goal of this study is to obtain high performing Artificial Neural Network (ANN) models for classification by reducing the large number of potential input variables using correlations between these variables. To achieve this, a breast cancer diagnosis problem with 30 potential input variables was considered and an ANN model was created by reducing the number of input variables with a proposed correlation-hypothesis test-based filtering method. The effectiveness of the proposed model was compared with six ANN models containing different sets of input variables. These six models include the model containing all input variables and the models obtained with input variables selected by stepwise regression, forward selection and backward elimination methods, which are model-based selection methods. While creating the models, the data set was divided into different training-test percentages and different numbers of neurons were tried in the hidden layer. Accuracy, recall, precision and F1-score metrics were used to compare the classification performances of the models. As a result, the accuracy value for the models with nine input variables selected by the proposed correlation-based filtering method was found to be between 0.93-0.95, which is significantly high. The recall value for our model was obtained between 0.85-0.88 and was sufficient. The precision value for our proposed model was determined to be very high, in the range of 0.98-0.988. The F1-score of the model proposed in this study is between 0.907-0.931, which is high enough. Considering that the proposed model has the lowest number of variables among the compared models, that is, it is the simplest model, and has a good classification performance even with only 10 neurons in the hidden layer, this model can be

used for rapid, lean and efficient classification at low costs, especially compared to model-based models.

Keywords: Input variable selection, Filtering method, Artificial Neural Networks, Classification, Correlation, Hypothesis tests.

1. GİRİŞ

Sınıflandırma, en sık karşılaşılan karar verme faaliyetlerinden biridir. Bir birim veya bireyin, o birimle veya bireyle ilgili bir dizi gözlenen özelliğe dayalı olarak önceden tanımlanmış bir grup veya sınıfa atanması gerektiğinde, bir sınıflandırma problemi ortaya çıkmaktadır. Mühendislik araştırmalarında, farklı endüstrilerde faaliyet gösteren işletmelerde ve tıptaki birçok problem, sınıflandırma problemi olarak ele alınabilir. Örnekler arasında kredi puanlama, kalite kontrol, tıbbi teşhis, kalp atışı, görüntü ve konuşma tanıma sayılabilir (Chuang ve Huang, 2011; Uluskan, 2020; Ramani vd., 2020; Acharya vd., 2017; Ciregan vd., 2012).

Geleneksel istatistiksel sınıflandırma yöntemleri değişkenler arasındaki ilişkilerin matematiksel denklemler biçiminde ifade edilmesine dayanır ve bir dizi varsayım üzerinde çalışır. Örnek olarak, doğrusal regresyon yöntemi, bağımsız ve bağımlı değişkenler arasında doğrusal bir ilişki olması, gözlemlerin birbirinden bağımsız olması ve hatanın normal dağılması gerektiğini varsayar (Uluskan, 2020). İstatistiksel modeller yalnızca temeldeki varsayımlar karşılandığında iyi çalıştığı için bu modellerin başarılı şekilde uygulanabilmesi için kullanıcıların hem veri özellikleri hem de model yetenekleri hakkında iyi bir bilgiye sahip olması gerekir (Zhang, 2000). Öte yandan, veriye dayalı modellemenin dayandığı makine öğrenimi, kural tabanlı programlamaya dayanmadan verilerden öğrenebilen bir algoritmadır. Genel olarak makine öğrenimi algoritmaları, bu varsayımların çoğundan bağımsızdır (Uluskan, 2020).

Özellikle son yıllarda makine öğrenimi teknikleri çok farklı alanlarda kullanılmakta ve karar vericilere daha doğru sonuçlara ulaşmalarında yardımcı olmaktadır. Benzer şekilde sınıflandırma problemleri, makine öğrenimi teknikleri sayesinde kısa sürede uygun modeller oluşturularak çözülebilir ve bu modeller süreç sonunda geleceğe yönelik tahminlerde kullanılabilir. Son yıllarda teknolojinin gelişmesiyle birlikte makine öğrenimi tekniklerinden yapay sinir ağları (YSA) sıklıkla sınıflandırma problemlerinde kullanılmaya başlanmıştır (Ryu vd., 2007).

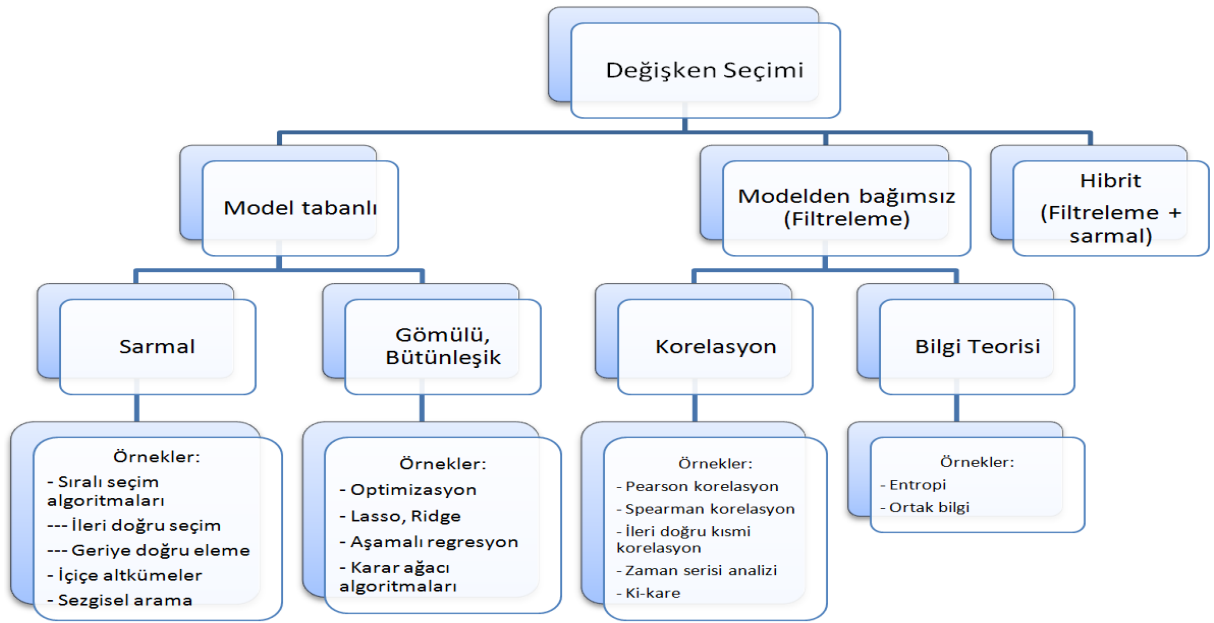
YSA, sınıflandırma için önemli bir araç olarak ortaya çıkmıştır. Son yıllarda gerçekleştirilen çalışmalar, YSA'nın çeşitli geleneksel sınıflandırma yöntemlerine etkili bir alternatif olduğunu ortaya koymuştur. Sınır ağları kendilerini veriye göre ayarlayabildikleri için kendi kendini uyarlayan yöntemlerdir. Ayrıca doğrusal olmayan karmaşık ilişkileri modellemede esnek yapıdadırlar. Son yıllarda YSA ile modelleme yapan kullanıcılar, girdi değişkenlerinin seçiminin gerekli olduğunun giderek daha fazla farkına varmaya başlamış ve bu amaç için farklı yöntemler kullanmışlardır. Girdi değişken seçimi yüksek sayıda potansiyel girdiye sahip modelleri sadeleştirerek kolay anlaşılabilir modeller elde etmek ve kaynakların etkin kullanımı açısından önemlidir. Literatürde girdi değişkenlerinin seçiminde model tabanlı yöntem uygulamalarındaki hesaplama süresi uzunluğu ve iş yükü fazlalığı gibi sebeplerden dolayı düşük maliyetli ve kısa sürede verimli sonuç alınabilecek yöntemlere gereksinim duyulmuştur. Bu durumlarda girdi değişken seçimi için genellikle filtreleme yöntemleri kullanılmıştır. Bu nedenlerle mevcut çalışmada olası girdi sayısı yüksek olan bir sınıflandırma problemi ele alınmış, bu problemde 30 adet olan olası girdi değişken sayısı önerilen korelasyon-hipotez testi tabanlı bir filtreleme yöntemi ile azaltılarak YSA modeli oluşturulmuştur. Önerilen modelin etkinliği, farklı girdi değişken setlerini içeren altı YSA modeli ile karşılaştırılmıştır. Bu altı model tüm girdi değişkenlerini içeren modelle, model tabanlı seçim yöntemlerinden aşamalı regresyonla, ileri doğru seçimle ve geriye doğru eleme yöntemleri ile seçilmiş girdi değişkenleriyle elde edilmiş olan modelleri kapsamaktadır. Performans karşılaştırmasında önerilen model için, tüm girdi değişkenlerini içeren modele göre ve model tabanlı seçim yöntemlerinden aşamalı regresyonla, ileri doğru seçimle ve geriye doğru eleme yöntemleri ile elde edilmiş olan modellere benzer oranlarda yüksek doğruluk, kesinlik, duyarlılık ve F1-skor değerleri elde edilmiştir.

2. BİLİMSEL YAZIN TARAMASI

2.1. Girdi Değişken / Öznitelik Seçimi ve Yöntemleri

Öznitelik olarak da adlandırılan girdi değişkenlerinin seçimi, oluşturulacak modellerin en uygun işlevsel formunun belirlenmesinde temel ve çok önemli bir husustur. Girdi değişkenlerinin seçimi, sınıflandırma, regresyon veya kümeleme gibi problemlerde model oluşturmak için en kullanışlı özelliklerin seçilmesi sürecidir (Solorio-Fernández vd., 2020). Seçim büyük ölçüde, model çıktısının uygun tahmincilerini belirlemek için mevcut veriler arasındaki ilişkilerin keşfedilmesine bağlıdır (Suzuki, 2011). Değişken seçiminde

boyutsallığın azaltılması için ilgisiz ve gereksiz değişkenler atılarak uygun değişkenler seçilmektedir (Remeseiro ve Bolon-Canedo, 2019). Dahası, girdi değişkenlerin seçimi yalnızca verilerin görselleştirilmesini ve anlaşılmasını kolaylaştıracak şekilde boyutsallığını azaltmakla kalmaz, aynı zamanda daha iyi genelleme yeteneğine sahip sade modeller elde edilmesini de sağlar (Solorio-Fernández vd., 2020). Böylece girdi boyutundaki azalma, öğrenme hızını ve model karmaşıklığını azaltarak veya genelleme yeteneğini ve sınıflandırma doğruluğunu artırarak performansı artırabilir. Uygun girdi değişkenlerinin seçimi aynı zamanda ölçüm maliyetini de azaltabilir ve problemin anlaşılmasını kolaylaştırabilir (Remeseiro ve Bolon-Canedo, 2019).



Şekil 1. Girdi değişken seçim yöntemleri

Son yıllarda YSA ile modelleme yapan kullanıcılar, girdi değişkenlerinin seçiminin gerekli olduğunun giderek daha fazla farkına varmaya başlamış ve bu amaç için farklı yöntemler kullanmışlardır. Literatürde değişken seçimi *model tabanlı* ve *modelden bağımsız* yöntemler (filtreleme yöntemleri) olmak üzere temelde ikiye ayrılmaktadır (May vd., 2011; Snieder vd., 2020). Model tabanlı yöntemler *sarmal* (wrapper) ve *gömülü* (embedded) olmak üzere iki gruba ayrılırken benzer şekilde modelden bağımsız filtreleme yöntemleri de *korelasyon* ve *bilgi teorisi* yöntemleri olmak üzere ikiye ayrılmaktadır. Son yıllarda filtreleme ve model tabanlı yöntemlerin birlikte kullanıldığı uygulamalar da *hibrit* yöntemler olarak isimlendirilmektedir (Jović vd., 2015; Solorio-Fernández vd., 2020). Literatür taraması sonunda özetlediğimiz değişken seçim teknikleri Şekil 1'de verilmiştir.

Model tabanlı sarmal ve gömülü yöntemlerin her ikisi de girdi değişken seçimini gerçekleştirmek için bir öğrenme yöntemi gerektirmektedir (Remeseiro ve Bolon-Canedo, 2019). *Sarmal yöntemler*, belirli bir kümeleme algoritmasının sonuçlarını kullanarak girdi değişkenlerden oluşan alt kümeleri değerlendirmektedir. Ancak, sarmal yöntemlerin yüksek hesaplama maliyetine sahip olmaları ve belirli bir kümeleme algoritmasıyla birlikte kullanımlarının sınırlı olması gibi dezavantajları vardır (Solorio-Fernández vd., 2020). Bu nedenle *gömülü yöntemler*, sarmal yöntemlerde oluşturulan farklı alt kümelerin yeniden sınıflandırılması için harcanan hesaplama süresini azaltmayı hedeflemektedir. Gömülü yöntemlerde temel yaklaşım, girdi değişken seçimini eğitim (training) sürecinin bir parçası olarak uygulamaktır (Chandrashekar ve Sahin, 2014). Gömülü yöntemler, girdi değişken seçimini modelleme algoritmasının uygulanması sırasında gerçekleştirmektedir. Dolayısıyla bu yöntemler algoritmanın içine entegre edilmekte, yani gömülmektedir (Jović vd., 2015). Sarmal ve gömülü yöntem uygulamalarındaki iş yükü fazlalığı ve hesaplama süresi uzunluğu gibi nedenlerden dolayı daha maliyet etkin ve daha kısa sürede verimli sonuç alınabilecek yöntemlere gereksinim duyulmuştur. Bu durumlarda girdi değişken seçimi için genellikle uzaklık, bilgi ve korelasyon gibi özellikler kullanılmıştır. Bu yöntemler filtreleme yöntemleri olarak adlandırılmıştır (Sezer ve Çakır, 2022).

Modelden bağımsız filtreleme yöntemleri önceden var olan bir modele dayanmamaktadır (Snieder vd., 2020). Filtreleme yöntemlerinde verinin kendisi aracılığıyla en ilgili girdi değişkenleri seçilmekte; yani girdi değişkenleri, ilgili değişkenlerin aranmasına rehberlik edebilecek herhangi bir kümeleme algoritması kullanılmadan, verinin kendine özgü özelliklerine göre değerlendirilmektedir (Solorio-Fernández vd., 2020). Bu şekilde filtreleme yöntemlerinin odak noktası verilerin genel özellikleri olduğundan bu yöntemler herhangi bir öğrenme yönteminden bağımsızdır. Tümevarım algoritmasından bağımsız olmaları nedeniyle hesaplama açısından maliyetli değildirler ve iyi bir genelleme kapasitesine sahiptirler (Remeseiro ve Bolon-Canedo, 2019). Ayrıca, filtreleme yöntemlerinin temel özellikleri hızları ve ölçeklenebilirlikleridir (Solorio-Fernández vd., 2020).

Son olarak, filtreleme ve sarmal yöntemlerin avantajlarını birleştirmek için *hibrit yöntemler* önerilmiştir. Bu yöntemler uygulanırken ilk önce, boyutsallığı azaltmak ve aday alt kümeler elde etmek için bir filtreleme yöntemi kullanılır. Daha sonra en iyi aday alt kümesini bulmak için bir sarmal yöntem kullanılır (Jović vd., 2015). Hibrit yöntemler, verimlilik (hesaplama çabası) ve etkililik arasında iyi bir denge sağlamaya çalışarak, her iki yaklaşımın niteliklerinden yararlanmaya çalışmaktadır (Solorio-Fernández vd., 2020). Bu şekilde hibrit

yöntemler hem sarmal yöntemlerin yüksek doğruluğuna, hem de filtreleme yöntemlerinin yüksek verimlilik özelliklerine sahip olur (Jović vd., 2015).

2.2. YSA Modellerinde Girdi Değişkenler Seçilirken Dikkat Edilecek Hususlar

Girdi değişkenlerini seçme faaliyetlerinde mevcut olan zorluklar şu şekilde özetlenebilir: (i) çok sayıda olası girdi değişken olması; (ii) olası girdi değişkenleri arasındaki korelasyonlar ve bu korelasyonlar sayesinde ortaya çıkan gereksiz girdi değişkenlerinin bulunması; ve (iii) tahmin gücü çok az olan veya hiç olmayan değişkenlerin bulunmasıdır (Suzuki, 2011).

YSA modelleri, açıklayıcı olmayan yetersiz girdi değişkenleri ile bilgi sağladığı halde diğer girdi değişkenleri ile yüksek korelasyonu olan gereksiz girdi değişkenlerinin dahil edilmesi nedeniyle gerekli olandan daha fazla girdi içerebilmektedir. YSA için en uygun girdi değişkenleri kümesini neyin oluşturduğunun tanımlanması, öncelikle girdi değişkenleri seçiminin model performansı üzerindeki etkisinin dikkate alınmasını gerektirir. Aşağıda bu temel hususlar özetlenmektedir (Suzuki, 2011):

İlgi düzeyi. YSA modellemelerinde sık görülen bir sorun, az sayıda değişkenin seçilmesi veya seçilen girdi değişkenlerinin yeteri kadar açıklayıcı olmamasıdır. Böyle bir durumda, çıktı seçilen girdi değişkenleri tarafından tam olarak açıklanamadığından, sonuçta performansı zayıf bir model elde edilmektedir.

Hesaplama zorluğu. Modele fazla sayıda girdi değişkeni eklemenin bir etkisi, YSA'nın hesaplama yükünü artırmasıdır. Bu durum, ağır eğitim hızının belirlenmesinde önemlidir.

Eğitim zorluğu. Gereksiz girdi değişkenlerinin modelde kullanılmasıyla bir YSA'nın eğitilmesi zorlaşabilmektedir. Böyle modellerde gereksiz parametreler ile hata arasındaki ilişkinin ortaya çıkarılması daha zordur. Gereksiz girdi değişkenleri modele gürültü ekleyerek öğrenme sürecini engellemektedir. Eğitim algoritması, çıktı değişkeni üzerinde hiçbir etkisi olmayan ağırlıkları ayarlamaya çalışarak gereksiz çalışabilir veya gürültü, önemli girdi-çıkıtı ilişkilerini maskeleyebilmektedir. Sonuç olarak, optimum hatanın belirlenebilmesi için eğitim algoritmasının çok kere yinelenmesi gerekebilir ve bu da modelin hesaplama yükünü artırmaktadır.

Sonuç olarak, istenilen bir girdi değişkeni, diğer girdi değişkenlerinden farklı (yani bağımsız) ve açıklayıcı olmalı, kısaca iyi bir tahminci olmalıdır. O halde optimal girdi değişkeni seti, açıklayıcı olmayan değişkenleri içermeyen, çıktı değişkeninin davranışını

tanımlamak için gereken en az girdi değişkenini içerecektir. Optimum girdi değişkenleri setinin tanımlanması, daha doğru, verimli, uygun maliyetli ve daha kolay yorumlanabilir YSA modelleriyle sonuçlanacaktır.

Bahsedilen zorluklar göz önüne alındığında mevcut çalışmamızda başlıca amaç, yüksek miktardaki olası girdi değişken sayısını, bu değişkenler arasındaki korelasyonlar göz önünde bulundurularak azaltılarak tahmin gücü yüksek YSA modelleri elde etmektir. Bunu yaparken de verimlilik ve harcanan zaman açısından avantajı olan ve fazla iş yükü gerektirmeyen *filtreleme yöntemlerinden korelasyon* kullanılarak girdi değişken seçimi yapılmıştır. Bu şekilde elde edilen sade modellerin etkinliği, tüm girdi değişkenlerini ve girdi değişkenlerin farklı alt kümelerini içeren YSA modellerinin performansları ile karşılaştırılmıştır. Modeller oluşturulurken veri seti farklı eğitim-test yüzdelerine bölünmüş ve gizli katmanda farklı nöron sayıları denenmiştir. Modellerin sınıflandırma performanslarını karşılaştırmak için *duyarlılık* (recall), *kesinlik* (precision) ve *F1-skoru* ölçütleri kullanılmıştır.

2.3. Meme Kanseri Teşhisinde YSA Uygulamaları

Tahmine dayalı modeller, teşhis ve prognostik faaliyetler için çeşitli tıbbi alanlarda kullanılmaktadır. Bu modeller, gerçek vakalardan elde edilen verileri oluşturan 'deneyim' üzerine kuruludur (Dreiseitl ve Ohno-Machado, 2002). Son yıllarda popülerliği artan YSA yöntemi birçok farklı alanda uygulanabilmektedir. Benzer şekilde, literatürde meme kanserini teşhis etmede de YSA uygulamaları görülmektedir.

Örnek olarak Chou (2004), YSA ve çok değişkenli uyarlanabilir regresyon eğrilerini (ÇDURE) bütünleştirerek meme kanseri teşhisi üzerinde çalışmıştır. ÇDURE yöntemiyle elde edilen değişkenler YSA için girdi olarak kullanılmıştır. Sonuç olarak hibrit sistemin geri yayımlı yapay sinir ağ türünden, diskriminant analizinden ve sadece ÇDURE yöntemiyle elde edilen sonuçlardan daha iyi sonuç verdiği bulunmuştur.

Jeres-Aragonés vd. (2003) YSA ile meme kanserinin tekrar nüksetmesini araştırmışlardır. 85 adet girdi değişkeni medikal uzmanlar tarafından 14'e kadar indirilebilmiştir. Yani derlenen 85 girdi değişkeninden 14'ünün kanseri etkilediği düşünülmüştür. Ardından bir karar ağacı ile 14 girdi değişkeni farklı alt kümelere ayrılıp yapay sinir ağı eğitilmiştir. Bu çalışmayla karar ağacı ile desteklenen YSA'nın farklı bireylerde doğru tedaviyi uygulamaya yardımcı olması amaçlanmıştır.

Zhang vd. (2021) iki farklı gelişmiş YSA olan grafik evrişim ağı (graph convolutional network) ve evrişimsel sinir ağını (convolutional neural networks) birleştirerek yeni bir metot elde etmeye çalışmışlardır. Bu metotta amaç meme mamografilerindeki kanserli olan mamografileri belirlemektir. Oluşturulan yeni YSA 332 resimlik yapay sinir ağında denenmiştir. %96,20±2,90 seviyesinde duyarlılık ve %96,10±1,60 seviyesinde ise doğruluk elde edilmiştir. YSA modeli 5 farklı modelle karşılaştırılmıştır. En iyi sonucu verenin, çalışmada oluşturulan YSA olduğu tespit edilmiştir.

Haddadnia vd. (2012) meme kanseri teşhisi için hastaların termal görüntülerinden alınan nitel ve nicel verileri kullanmışlardır. Bu bilgiler işlendikten sonra teşhis için gerekli en iyi parametreler seçilip teşhisin doğruluğunu arttırmak adına YSA ve genetik algoritmayla oluşturulan model sayesinde iyileştirme sağlanmaya çalışılmıştır. 200 kişiden alınan termal resimlerden 8 adet teşhise yardımcı parametre elde edilmiştir. Bu 8 parametre modellerde girdi değişken olarak kullanılmıştır.

Alshanbari vd.. (2021) evrişimsel sinir ağları ile meme kanserini teşhis etmeye çalışmışlardır. Ağ, 11 katmandan oluşmaktadır. Herkese açık olan BreakHis veri seti çalışmada kullanılmıştır. Sonuç olarak, eğitilen modelle %96 doğruluk oranına ulaşılmıştır.

Diğer taraftan, Irmak vd. (2021) YSA'yı geleneksel makine yöntemleriyle karşılaştırmışlardır. Çalışmada lojistik regresyon, rastgele orman ve k-en yakın komşuluk algoritmaları, destek vektör makineleri, XGboost ve YSA karşılaştırılmıştır. YSA kullanılarak elde edilen %99,36 doğruluk oranının tüm yöntemler arasında en yüksek olduğu tespit edilmiştir.

Bılgıç (2021) evrişimsel sinir ağları ile meme kanserini ve deri kanserini teşhis etmeye çalışmıştır. Daha sonra lojistik regresyon yöntemi ile verileri analiz etmiş ve başarı grafikleri oluşturarak deri kanseri ve meme kanserini karşılaştırmıştır.

Literatürde görüldüğü gibi farklı tip YSA modellerinin teşhiste yardımcı olmada sınıflandırma için kullanıldığı veya geleneksel makine öğrenme yöntemleri ile karşılaştırıldığı çalışmalar yapılmıştır. Ancak literatürde, çok sayıda olası girdi değişkeninin olduğu durumlarda teşhis için oluşturulacak modellerde, korelasyon tabanlı bir filtreleme yöntemi yoluyla değişken sayısının azaltılmasını öneren çalışmalara rastlanılmamıştır.

2.4. Girdi Değişken Seçimi Üzerine Yapılmış Çalışmalar

Girdi değişken seçimi üzerine çok farklı alanlarda ve farklı yöntemler kullanılarak çalışmalar gerçekleştirilmiştir. Örnek olarak, Emanet vd. (2021)'nin çalışmasında, literatürde önerilen farklı girdi değişken seçim yöntemleri ve makine öğrenimi teknikleri yardımıyla, girdi değişken seçiminin saldırı tespit sisteminin performansı üzerindeki etkisi incelenmiştir. Seçim için ki-kare testi, Spearman korelasyon katsayısı ve özyinelemeli değişken eleme yöntemleri uygulanmıştır. Elde edilen alt kümeler lojistik regresyon, karar ağacı, çok katmanlı algılayıcı, pasif-agresif ve gradyan artırma gibi makine öğrenimi yöntemleri ile sınıflandırılmış ve performans sonuçları karşılaştırılmıştır.

Sezer ve Çakır, (2022)'in bankacılık sektöründe sınıflandırma amaçlı değişken alt kümesi seçimi için gerçekleştirdikleri çalışmalarında, sezgisel arama yöntemlerinden ileri/geri yönlü seçim için BestFirst, sıralı arama için RankSearch algoritmaları seçilmiş ve elde edilen yedi farklı değişken alt kümesi için ve tüm değişkenleri içeren veri kümesi için ayrı ayrı bir karar ağacı algoritması ile sınıflandırma yapılarak sekiz farklı karar ağacı oluşturmuştur.

Diğer taraftan Solorio-Fernández vd. (2020) denetimsiz özellik seçim yöntemlerinin bir literatür taramasını sundukları çalışmalarında, bu yöntemlerin temel özelliklerini tanımlayarak bir sınıflandırmasını sunmuş ve yöntemlerin avantaj ve dezavantajlarını belirtmişlerdir. Benzer şekilde Chandrashekar ve Sahin (2014) filtreleme, sarmal ve gömülü girdi değişken seçim teknikleri üzerine bir anket çalışması yapmışlardır.

Snieder vd. (2020) YSA ile nehir taşkınları tahminlemesi üzerine yaptıkları çalışmalarında kısmi korelasyon, kısmi karşılıklı bilgi olmak üzere iki modelden bağımsız girdi değişken seçim yöntemi ile çalışmada geliştirdikleri iki yeni model tabanlı yöntemi karşılaştırmışlardır. Benzer şekilde Fernando vd. (2009)'nin su kaynakları değişkenlerinin modellenmesinde YSA kullandıkları çalışmalarında, kısmi karşılıklı bilgi girdi değişken seçim algoritmasını hesaplama verimliliğini artıracak şekilde değiştirilmiştir. Snelder vd. (2007) çevresel faktörlerin çok değişkenli sınıflandırmaları alanında gerçekleştirdikleri çalışmalarında girdi değişken seçimi için çevresel ve biyolojik matrisler arasındaki korelasyonu maksimuma çıkarma amacına sahip ileri adımlı regresyona benzeyen bir prosedür kullanmışlardır.

Son olarak sağlık alanında Gao vd. (2018) lojistik regresyon, rastgele orman, destek vektör makineleri ve XGboost gibi model tabanlı ve modelden bağımsız yöntemleri Parkinson hastalığı teşhisinde ve klinik sonuçların sınıflandırılması alanında uygulamışlardır.

Sonuç olarak girdi değişken seçimi için literatür incelendiğinde farklı alanlarda model tabanlı ve modelden bağımsız yöntemlerin kullanıldığı ve karşılaştırıldığı çalışmaların bulunduğu görülmektedir. Ancak sağlık alanında çok fazla çalışma karşımıza çıkmamaktadır. Diğer taraftan meme kanseri tahmini ile ilgili özellikli literatür incelendiğinde ilgili alandan girdi değişken seçimi ile ilgili az sayıda çalışma yapıldığı görülmektedir. Bu noktada mevcut çalışmada literatürdeki bu boşluğu doldurmak adına sağlık alanında, özellikli olarak meme kanseri teşhisi konusunda, modelden bağımsız korelasyon tabanlı bir girdi değişken seçim yöntemi önerilmiş ve etkinliği model tabanlı yöntemler dahil farklı modellerle karşılaştırılmıştır.

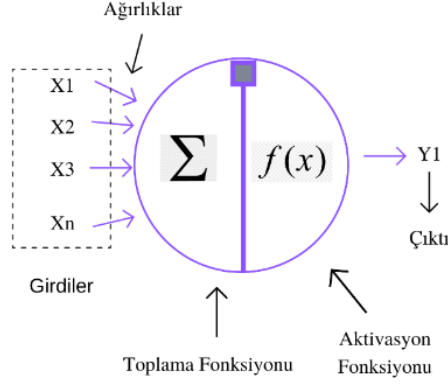
3. YÖNTEM

3.1. Yapay Sinir Ağları

Yapay sinir ağları (YSA) son yıllarda mühendislik, biyoloji, tıp, ekonomi gibi çeşitli alanlarda sınıflandırma, kümeleme, görüntü tanıma ve tahminleme gibi farklı uygulamalarda başarı ile kullanılan popüler teknolojilerden biri haline gelmiştir (Suzuki, 2011, Abiodun vd., 2018; Wu ve Feng, 2018). YSA'nın çok yönlülüğü, yüksek kapasitesinden ve öğrenme fonksiyonundan kaynaklanmaktadır (Suzuki, 2011). Bir makine öğrenimi türü olan YSA, kullanılabilirlik açısından geleneksel regresyon ve istatistiksel modellere iyi bir alternatiftir (Abiodun vd., 2018). YSA'nın temel özellikleri, karmaşık ve doğrusal olmayan girdi-çıkı ilişkilerini öğrenme, sıralı eğitim prosedürlerini kullanma ve kendilerini verilere uyarlama yeteneğine sahip olmalarıdır (Basu vd., 2010). Sadece sinir ağları olarak da adlandırılabilen YSA, beyindeki biyolojik sinir ağlarının yapısından ve işlevinden ilham alan matematiksel hesaplama modelleridir (Suzuki, 2011; Dongare vd., 2012).

Bir YSA, birbirine sinaptik ağırlıklar yoluyla bağlanan çok sayıda işlem birimlerinden, yani bir dizi yapay nörondan oluşur (Suzuki, 2011; Dongare vd., 2012). Bir sinir ağı genel olarak girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç katmana ayrılabilir. Dış ortamdan bilgi (veri) almaktan sorumlu girdi katmanı, gizli katmana verileri iletir. Gizli katman verilerin işlenmesinden sorumlu olan nöronlardan oluşmaktadır. Yine nöronlardan oluşan çıktı katmanı ise, son ağ çıktılarının üretilmesinden ve sunulmasından sorumludur (Uluskan, 2020). YSA'lar dışarıdan gelen bilgiyi toplama fonksiyonu ile toplar ve aktivasyon fonksiyonundan geçirerek çıktıyı oluşturur. Toplama fonksiyonu hücreye gelen net girdiyi hesaplayan fonksiyondur. Aktivasyon fonksiyonu ise gelen net girdiyi işleyerek hücrenin bu

girdiye karşı ne çıktı oluşturacağını belirler (Öztemel, 2003). Şekil 2’de basit bir YSA verilmiştir.



Şekil 2. Basit bir yapay sinir ağı hücresi

Geleneksel hesaplama yöntemlerinin aksine, YSA istenen girdi-çıkı ilişkilerini elde edebilmek için 'eğitilir'. Eğitim, yani öğrenme aşamasında, veri örnekleri ağa sunulur ve bir öğrenme algoritması kullanılarak parametreler ayarlanır (Liao ve Wen, 2007). Bu şekilde YSA, ağırlıkları ayarlayarak bir veri setini 'öğrenebilir' (Suzuki, 2011). Ele alınan probleme ilişkin mevcut bilgiye ve kullanıcının amacına göre, kullanılan öğrenme prosedürü 'denetimli', 'denetimsiz' veya her ikisi birden olabilir. Denetimli öğrenme prosedürü, bir çıktıyı gerektirir ve ağa sunulan girdi-çıkı çiftleri ile gerçekleştirilir (Liao ve Wen, 2007). Yani tüm örnekler için girdi ve o girdinin oluşturması gereken çıktılar ağa aktarılır. Ağ, verilen girdilere göre çıktıları oluşturabilmek adına ağırlıkları yeniler. Gerçek çıktılar ile ağın ürettiği çıktılar arasındaki hata hesaplanarak yeni ağırlıklar bu hata oranına göre düzenlenir (Alpaydin, 2020). Denetimsiz öğrenmede ise, ağ, çıktıya gerek duymaz ve bir maliyet fonksiyonuna dayalı olarak öğrenme gerçekleşir (Suzuki, 2011). Denetimsiz öğrenmede ağın eğitimine destek olan herhangi bir çıktı bulunmamaktadır. Sistem sadece girdilerle eğitilir ve buna göre bir çıktı oluşturmaya çalışır (Alpaydin, 2020). Yani denetimsiz öğrenmede ağ, örnek sınıfları arasında doğal olarak var olan ayrımı enbüyükleyebilmek için kendini aşamalı olarak ayarlamaya çalışır (Liao ve Wen, 2007).

YSA gibi öğrenen sistemlerde eğitim, yöntem fark etmeksizin bazı kurallar altında gerçekleştirilir. Bu kurallar çevrimiçi (online) ve çevrimdışı (offline) olarak çalışmaktadır. Çevrimiçi öğrenme kuralları gerçek zamanlı çalışabilmektedir. Bu kurallara göre eğitim gören sistemler fonksiyonların işlemlerini yaparlarken öğrenmeye devam etmektedirler. Çevrimdışı

öğrenme kuralını kullanan sistemler ise kullanılmadan önce eğitilirler. Eğitildikten sonra kullanılmaya başlanan sistemler uygulanırken öğrenme gerçekleştirmezler. Eğer sistemin öğrenmesi gerekli olan yeni bilgiler varsa sistem kullanımdan çıkarılır, çevrimdışı olarak tekrar eğitilir. Eğitimin sonunda sistem tekrar kullanıma açık hale gelmektedir (Goodfellow vd., 2016). Bu çalışmamızda kullanılan YSA'lar denetimli öğrenme ile eğitilmiş çevrimdışı ağlardır. Ayrıca çalışmada oluşturulan YSA'lar çok katmanlı, geri yayımlı ve ileri beslemelidir.

3.2. Performans Değerlendirme Ölçütleri-Doğruluk, Kesinlik, Duyarlılık ve F1-Skoru

Bu çalışmada, önerilen korelasyon tabanlı filtreleme yöntemiyle girdi değişkenleri seçilmiş olan YSA modelinin etkinliği, tüm girdi değişkenlerini ve girdi değişkenlerin farklı alt kümelerini içeren altı farklı YSA modelinin performansları ile karşılaştırılmıştır. Modellerin sınıflandırma performanslarını karşılaştırmak için doğruluk (accuracy), duyarlılık (recall), kesinlik (precision) ve F1-skoru ölçütleri kullanılmıştır. Tablo 1'de görülen karmaşıklık matrisinin (karışıklık veya hata matrisi de denilmektedir) dört adet elemanı vardır. Bu elemanlar şunlardır:

Doğru Pozitifler (TP- True positive): "Pozitif" olarak doğru şekilde tahmin edilen örneklerin sayısıdır.

Yanlış Pozitifler (FP - False positive): Yanlışlıkla "pozitif" olarak tahmin edilen örneklerin sayısıdır.

Gerçek Negatifler (TN): "Negatif" olarak doğru şekilde tahmin edilen örneklerin sayısıdır.

Yanlış Negatifler (FN): Yanlışlıkla "negatif" olarak tahmin edilen örneklerin sayısıdır.

Tablo 1. Karmaşıklık Matrisi

		Gerçek	
		Pozitif	Negatif
Tahmin	Pozitif	TP	FP
	Negatif	FN	TN

Bu çalışmamızda "pozitif"ler "kötü huylu" hücreler olarak "negatif"ler de "iyi huylu" hücreler olarak tanımlanmıştır. O halde doğru sınıflandırılmış pozitif örnekler (TP), gerçekte kötü huylu hücrelerin kötü huylu olarak doğru tahmin edildiği örneklerdir. Doğru

sınıflandırılmış negatif örnekler (TN), gerçekte iyi huylu hücrelerin iyi huylu olarak tahmin edilen örneklerdir. Yanlış sınıflandırılmış pozitif örnek (FP), iyi huylu hücrelere kötü huylu denmesi anlamına gelmektedir. Son olarak, yanlış sınıflandırılmış negatif örnek (FN) kötü huylu olanlara iyi huylu denmesi anlamına gelmektedir.

Karmaşıklık matrisinin bileşenlerini kullanarak, sınıflandırma modellerini değerlendirmek için kullanılan doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-skoru gibi çeşitli ölçütler tanımlanabilir.

En kolay anlaşılabilen ölçütlerden biri olan *doğruluk*, Denklem 1'de verildiği gibi toplam doğru tahmin sayısının toplam tahmin sayısına oranıdır. "Yaptığımız tüm tahminlerden ne kadarı doğru çıktı?" sorusuna yanıt vermektedir. Bu çalışmada doğruluk ölçütü doğru yüzde oranı olarak tanımlanmıştır.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Geri çağırma olarak da adlandırılan *duyarlılık* ölçütü gerçek pozitiflerin doğru şekilde belirlenmesinin ölçüsüdür. Yani modelimizin ilgili verileri ne kadar doğru tanımlayabildiğinin bir ölçüsünü vermektedir. Bu nedenle duyarlılık hassasiyet veya gerçek pozitif oran olarak da tanımlanmaktadır. "Doğru olarak tahmin edilmesi gereken tüm veri noktalarından ne kadarını doğru olarak tahmin ettik?" sorusuna yanıt vermektedir. Duyarlılık Denklem 2'de verildiği şekilde hesaplanmaktadır.

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (2)$$

Kesinlik, Denklem 3'te verildiği gibi gerçek pozitiflerin tüm pozitiflere oranıdır. Kısaca gerçekte pozitif olanların tüm pozitif olarak tahminlenmişlere oranıdır. "Yaptığımız tüm pozitif tahminlerden ne kadarı doğru çıktı?" sorusuna yanıt vermektedir.

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (3)$$

Kesinlik ve duyarlılık tanımlarından da görülebileceği gibi bu ölçütler sıkı bir şekilde bağlantılıdır. Kısaca iki ölçüt arasında bir ödünleşme vardır. Bu nedenle kesinlik ve duyarlılık ölçütlerinin harmonik ortalaması olan ve Denklem 4'te verilen F1-skoru bu iki ölçütü birleştiren bir ölçüdür. Yapılan analiz açısından kesinlik ve duyarlılık ölçütlerinin ikisi de önemli olduğunda F1-skoru belirlenmelidir.

$$F1 \text{ skoru} = 2 * \frac{\text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (4)$$

3.3. Veri Seti

Çalışmada University of California Irvine Machine Learning Repository açık erişim internet sitesinden elde edilen ve 569 hastadan derlenmiş olan 30 girdi değişkenli bir veri seti kullanılmıştır (UCI, 2022). Veriler elde edilirken memedeki kistik bölgeden küçük bir hücre topluluğu, iğne aspirasyon biyopsisi yöntemiyle alınmış ve Olympus mikroskopuyla incelenmiştir. Ardından bu görüntüler mikroskobun üzerine takılmış olan bir kamerayla dijital ortama aktarılmıştır. Daha sonra seçilebilen hücrelerin sınırları çizilmiştir. Çizilen sınırlara “snake” adı verilmiştir. Şekil 3’te bu çizilen sınırlar görülebilir.



Şekil 3. Dijitalize edilmiş seçilebilir hücrelerin sınırları, snake

Bilgisayar görü sistemi bu hücrelerin yarıçap (radius), çevre uzunluğu (perimeter), hücrenin alanı (area), yoğunluk (compactness), yumuşaklık (smoothness), içbükeylik (concavity), içbükey noktalar (concavepoints), simetri (symmetry), fraktal boyut (fractal dimension) ve doku (texture) olmak üzere 10 adet özelliğini elde edebilmektedir. Her hücre kümesi için ortalama, maksimum ve standart hata değerleri hesaplanmıştır. Bu şöyle açıklanabilir: Bir hastadan meme hücresi topluluğu alınmıştır. Ardından bu hücre topluluğunun yukarıda verilen 10 adet değişkeni incelenmiştir. 10 değişkenden ilgili değişken için ortalama alınarak ortalama değer (ort) bulunmuştur. Hücre topluluğu içindeki ilgili değişken için en büyük değer, o değişkenin maksimum (maks - en kötü) değeridir. En büyük değere en kötü denilmesinin sebebi ilgili faktörlerin büyüdükçe kanser olasılığının artmasından yani daha kötüye gitmesinden dolayıdır. Daha sonra ise ilgili değişkenin standart hatası (sh) hesaplanmıştır. Böylece toplam 30 adet girdi değişken verisi, 569 hastadan elde edilmiştir. Bu 30 girdi değişkeni mevcut çalışmamızda Tablo 2'deki şekilde isimlendirilmiştir.

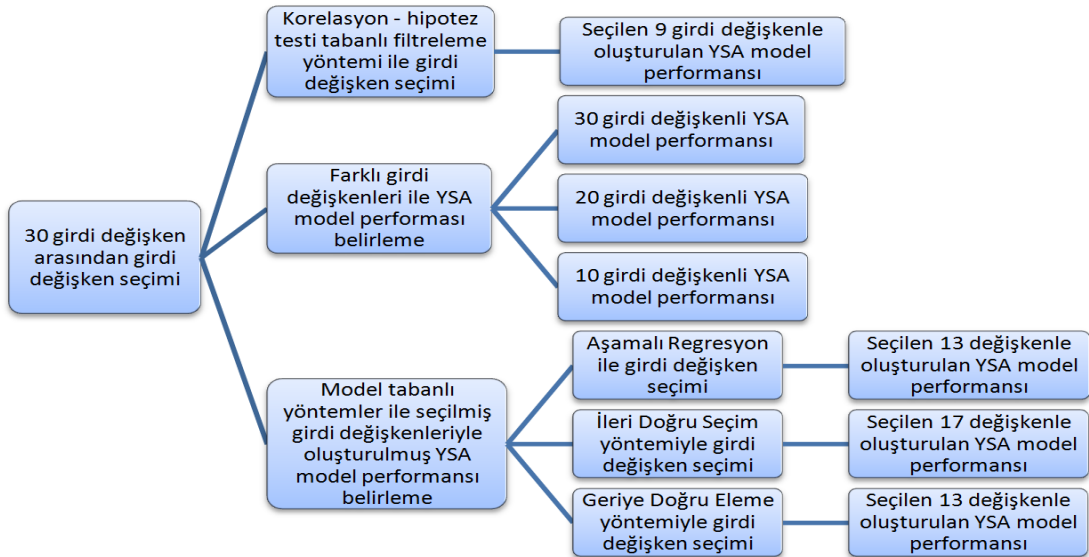
Kısaca veri seti 30x569 büyüklüğünde bir matristir. Veri setinde 357 tane iyi huylu (sağlıklı), 212 tane kötü huylu (kanseri) hücre örneği yer almaktadır.

Tablo 2. Girdi değişkenleri

Özellikler	İlgili girdi değişkenleri
Ortalama girdi değişkenleri	$X_1 = \text{yarıçap_ort}, X_2 = \text{çevre_ort}, X_3 = \text{doku_ort}, X_4 = \text{alan_ort}, X_5 = \text{yoğunluk_ort}, X_6 = \text{yumuşaklık_ort}, X_7 = \text{içbükeylik_ort}, X_8 = \text{içbükey_nkt_ortalama}, X_9 = \text{fraktal_boyut_ort}, X_{10} = \text{simetri_ort}$
Standart hata girdi değişkenleri	$X_{11} = \text{yarıçap_sh}, X_{12} = \text{çevre_sh}, X_{13} = \text{doku_sh}, X_{14} = \text{alan_sh}, X_{15} = \text{yoğunluk_sh}, X_{16} = \text{yumuşaklık_sh}, X_{17} = \text{içbükeylik_sh}, X_{18} = \text{içbükey_nkt_sh}, X_{19} = \text{fraktal_boyut_sh}, X_{20} = \text{simetri_sh}$
Maksimum değer girdi değişkenleri	$X_{21} = \text{yarıçap_maks}, X_{22} = \text{çevre_maks}, X_{23} = \text{doku_maks}, X_{24} = \text{alan_maks}, X_{25} = \text{yoğunluk_maks}, X_{26} = \text{yumuşaklık_maks}, X_{27} = \text{içbükeylik_maks}, X_{28} = \text{içbükey_nkt_maks}, X_{29} = \text{fraktal_boyut_maks}, X_{30} = \text{simetri_maks}$

4. ANALİZLER ve BULGULAR

Bu çalışmada başlıca amaç, yüksek miktardaki (30 adet) olası girdi değişken sayısını, bu değişkenler arasındaki korelasyonları göz önünde bulunduran bir yöntem önerisiyle azaltarak maliyet ve zaman açılarından etkin, sınıflandırma performansı yüksek YSA modeli elde etmektir. Bu nedenle mevcut çalışmada korelasyon-hipotez testi tabanlı bir filtreleme yöntemi önerilmiş ve bu yöntemle girdi değişkenleri seçilerek YSA modeli oluşturulmuştur. Bu modelin performansı sarmal ve gömülü yöntemlerle seçilmiş girdi değişkenlerini de içeren altı farklı YSA modeli ile karşılaştırılmıştır. Karşılaştırılan YSA modelleri ile çalışmanın akışı Şekil 4'te verilmiştir.

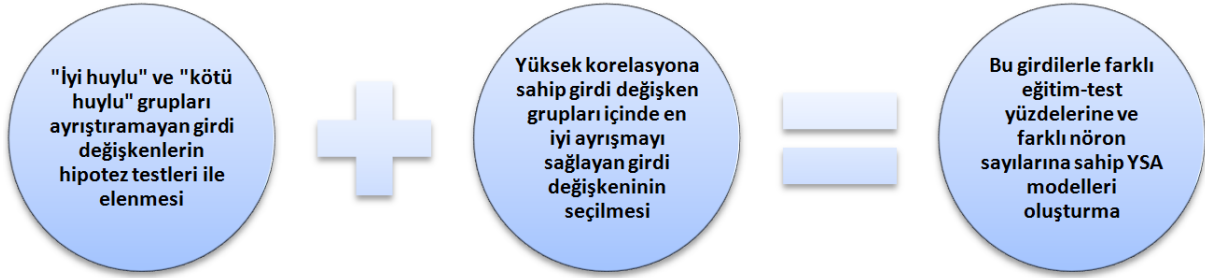


Şekil 4. Çalışma akışı

İzleyen bölümlerde sırayla, önerilen yöntem ile değişken seçimi anlatılmış ve seçilen dokuz değişken verilmiştir. Sonrasında model tabanlı yöntemlerden aşamalı regresyon, ileri doğru seçim ve geriye doğru eleme yöntemleriyle seçilen değişkenler verilmiştir. Son olarak, önerilen model dahil yedi farklı YSA modelinin performans incelemesi gerçekleştirilmiş ve sonuçlar detaylı olarak tartışılmıştır.

4.1. Önerilen Korelasyon-Hipotez Testi Tabanlı Filtreleme Yöntemiyle Girdi Değişken Seçimi

Önerilen yöntemde ilk olarak, 30 girdi değişken arasından çıktıyı en iyi açıklayanları belirleyebilmek adına, yani sınıflandırmayı en iyi yapanları belirleyebilmek için "iyi huylu" ve "kötü huylu" sınıflarını istatistiksel olarak anlamlı şekilde ayıramayan girdi değişkenleri hipotez testleri yardımıyla elenmiştir. Sonrasında, anlamlı girdi değişkenleri arasındaki korelasyonlar incelenerek en iyi ayrışmayı yapanlar adım adım seçilmiştir. Önerilen yöntemle girdi değişken seçimi Şekil 5'te özetlenmiş ve detayları aşağıda açıklanmıştır.



Şekil 5. Önerilen yöntemle girdi değişken seçimi

4.1.1. "İyi Huylu" ve "Kötü Huylu" Gruplarını Ayırtıramayan Girdi Değişkenlerin Hipotez Testleri İle Elenmesi

Öncelikli olarak 30 girdi değişkeni arasından, *iyi huylu* ve *kötü huylu* şeklinde etiketlenmiş olan çıktı değişkenini en iyi açıklayan değişkenleri belirleyebilmek için hipotez testleri gerçekleştirilmiştir. Burada iki ortalama farkı için z-testleri, her bir girdi değişken için kötü huylu ve iyi huylu grup ortalamaları arasında fark olup olmadığını belirleyecek şekilde gerçekleştirilmiştir. Hipotez testleri için normallik varsayımı test edilmiş ve normallik varsayımının $p - değeri > \alpha = 0,01$ değerleri ile sağlandığı görülmüştür. Bu testlerde kullanılan hipotezler aşağıdaki şekildedir:

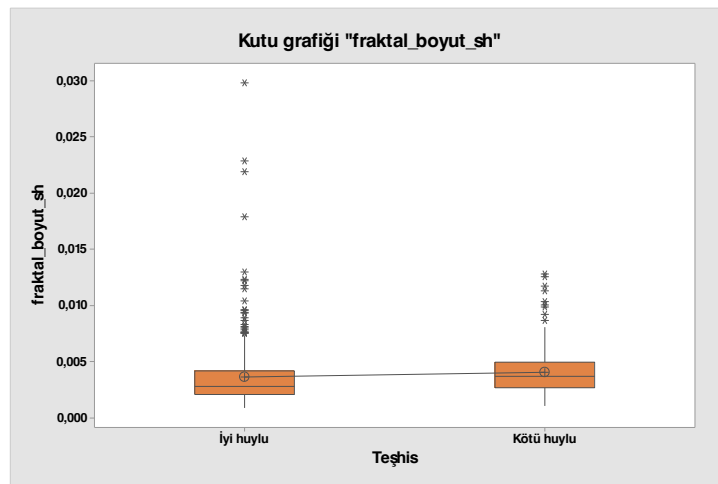
$$H_0: \mu_{kötü\ huylu} = \mu_{iyi\ huylu}$$

$$H_a: \mu_{kötü\ huylu} \neq \mu_{iyi\ huylu}$$

Hipotez testleri sonunda Tablo 3'te p-değerleri verilen girdi değişkenlerinin çıktığı açıklayamadığı, yani kötü huylu ve iyi huylu hücre gruplarını ayırtamadığı belirlenmiş ve bu girdi değişkenler elenmiştir. Eleme işlemi yapılırken $\alpha = 0,01$ olarak alınmıştır. Burada *fraktal_boyut_ort*, *doku_sh*, *yumuşaklık_sh*, *simetri_sh* ve *fraktal_boyut_sh* değişkenleri için p-değerlerinin $\alpha = 0,01$ değerinden büyük olduğu yani sınıflandırma açısından istatistiksel olarak anlamlı olmadıkları belirlenmiştir. Bu girdi değişkenleri arasından *fraktal_boyut_ort* değişkeninin p-değeri küçük olduğundan Şekil 6'da verilen kutu grafiği oluşturularak bu değişkenin iyi bir ayırma sağlamadığı teyit edilmiştir. Yani, bu değişken temelinde sınıflandırma yapıldığı zaman iyi huylu ve kötü huylu grupları için kutular büyük oranda çakışmaktadır. Bu nedenle *fraktal_boyut_ort* değişkeni dahil *doku_sh*, *yumuşaklık_sh*, *simetri_sh* ve *fraktal_boyut_sh* girdi değişkenleri elenmiştir.

Tablo 3. Anlamlı sınıflandırma yapamayan girdi değişkenleri

Girdi değişken	z-değeri	p-değeri
fraktal_boyut_ort	0,30	0,767
doku_sh	0,21	0,835
yumuşaklık_sh	1,62	0,105
simetri_sh	0,14	0,887
fraktal_boyut_sh	-2,04	0,042



Şekil 6. *fraktal_boyut_sh* değişkeni temelinde sınıflandırma için kutu grafikleri

4.1.2. Yüksek Korelasyona Sahip Girdi Değişken Grupları İçinde En İyi Ayırışmayı Sağlayan Girdi Değişkeninin Seçilmesi

Bu aşamada kalan girdi değişkenleri için Pearson korelasyon katsayıları hesaplanmıştır. Korelasyonları yüksek olan değişkenler için gruplar oluşturulmuş ve ilgili grup içinden en iyi ayırışmayı sağlayan girdi değişken yine hipotez testleri ile yardımıyla seçilmiştir. Birbiri ile korelasyonu yüksek olan girdi değişkenlerinin benzerlik göstermesi sebebiyle bu değişkenlerin bir arada kullanımı yapay sinir ağının eğitimine olumlu bir etki etmeyecektir. Ayrıca fazla sayıda girdi değişkenle yapılan eğitimler bazen ağın sınıflandırmasına olumsuz yansiyabilir (Arı ve Hanbay, 2019).

Korelasyon katsayısı 0,8'den büyük olan değişkenler arasında çok güçlü ilişki bulunmaktadır (LaMorte, 2021). Bu durum ise ciddi çoklu bağlantı (çoklu doğrusal bağlantı) problemi oluşturabilir (Calkins, 2005). Bu nedenle sırayla bir girdi değişkeni ile korelasyonları 0,8 ve üzeri olan girdi değişkenleri grup olarak alınmış, bu grup içinde iyi huylu/kötü huylu ayrışmasını en iyi yapan değişken seçilmiş ve diğerleri elenmiştir. Örnek olarak ilk önce *yarıçap_ort* değişkeni ile 0,8'den yüksek korelasyona sahip girdi değişkenleri belirlenmiştir (Tablo 4). Bu değişkenler *çevre_ort*, *alan_ort*, *yarıçap_maks*, *çevre_maks*, *alan_maks* ve *içbükey_noktalar_ort* değişkenleridir. Bu değişkenler için iyi huylu/kötü huylu grup ortalamaları farkını verecek şekilde z-testleri yapılmıştır. Sınıflandırma açısından anlamsız olan girdi değişkenleri ilk adımda elendiğinden geriye kalan bu anlamlı girdi değişkenleri ($p - değerleri = 0,000 < \alpha = 0,01$) arasından z-değeri en büyük olan değişken seçilmiş ve gruptaki diğer değişkenler elenmiştir. Kısaca bu grupta *çevre_maks* girdi değişkeni ($z - değeri = -25,33$) seçilmiştir.

Tablo 4. Grup 1 için korelasyon ve z-değerleri

Grup 1			
Girdi değişkeni	yarıçap_ort	z-değeri	p-değeri
yarıçap_ort	1	-22,21	0,000
çevre_ort	0,998	-22,94	0,000
alan_ort	0,987	-19,64	0,000
yarıçap_maks	0,97	-24,83	0,000
çevre_maks	0,965	-25,33	0,000
alan_maks	0,941	-20,57	0,000
içbükey_noktalar_ort	0,823	-24,84	0,000

Tablo 5. Girdi deęişken grupları için korelasyon deęerleri ile z-deęerleri

Grup 2			
Girdi deęişkeni	doku_ort	z-deęeri	p-deęeri
doku_ort	1	-11,02	0,000
doku_maks	0,912	-12,26	0,000
Grup 3			
Girdi deęişkeni	yoęunluk_ort	z-deęeri	p-deęeri
yoęunluk_ort	1	-15,82	0,000
içbükeylik_ort	0,883	-20,33	0,000
yoęunluk_maks	0,866	-15,16	0,000
içbükeylik_maks	0,816	-19,60	0,000
içbükey_nkt_maks	0,816	-29,12	0,000
Grup 4			
	yumuşaklık_ort	z-deęeri	p-deęeri
yumuşaklık_ort	1	-9,30	0,000
yumuşaklık_maks	0,805	-10,82	0,000
Grup 5			
	yarıçap_sh	z-deęeri	p-deęeri
yarıçap_sh	1	-13,30	0,000
çevre_sh	0,973	-12,83	0,000
alan_sh	0,952	-12,16	0,000
Grup 6			
	yoęunluk_sh	z-deęeri	p-deęeri
yoęunluk_sh	1	-7,08	0,000
içbükeylik_sh	0,801	-6,92	0,000
Grup 7			
	çevre_maks	z-deęeri	p-deęeri
çevre_maks	1	-25,33	0,000
içbükey_nkt_maks	0,816	-29,12	0,000

Benzer şekilde tüm girdi değişkenleri arasındaki korelasyonlar incelenerek sırayla *doku-ort* değişkeni ile, *yoğunluk-ort* değişkeni ile, *yumuşaklık_ort* değişkeni ile, *yarıçap-sh* değişkeni ile, *yoğunluk_sh* değişkeni ile ve *çevre-maks* değişkeni ile 0,8 ve üzeri korelasyona sahip değişkenler belirlenerek gruplar oluşturulmuş ve her grup için en büyük z-değerine sahip değişken seçilmiş, diğer değişkenler elenmiştir. Tablo 5'te ilgili değişken grupları için korelasyon değerleri ile z-değerleri verilmiştir. Seçilmiş olan girdi değişkenleri Tablo 5'te *italik* olarak işaretlenmiştir.

Sonuç olarak önerilen yöntemle *simetri_ort*, *yarıçap_sh*, *yoğunluk_sh*, *içbükey_nkt_sh*, *doku_maks*, *yumuşaklık_maks*, *içbükey_nkt_maks*, *simetri_maks* ve *fraktal_boyut_maks* olmak üzere 9 adet girdi değişkeni seçilmiştir.

4.2. Model Tabanlı Yöntemlerle Değişken Seçimi

Aşamalı regresyon yöntemi ile değişken seçimi sonunda elde edilmiş olan 13 adet girdi değişkeni Tablo 6'daki ANOVA tablosunda ve ileri doğru seçim yöntemi ile elde edilmiş olan 17 adet girdi değişkeni Tablo 7'deki ANOVA tablosunda verilmiştir. Tablo 7'de bazı değişkenler için p-değerinin $\alpha = 0.05$ değerinden büyük olduğu görülmektedir. Bunun nedeni ileri doğru seçim yönteminde modele giren bir değişkenin tekrar çıkamaması kuralıdır.

Tablo 6. Aşamalı regresyon yöntemi ile değişken seçimi sonunda elde edilmiş olan 13 adet girdi değişkeni

Source	DF	SS	MS	F-Value	P-Value
Regression	13	102,461	7,88160	143,18	0,000
<i>yoğunluk_ort</i>	1	1,837	1,83650	33,36	0,000
<i>içbükey_nkt_ort</i>	1	0,739	0,73864	13,42	0,000
<i>yarıçap_sh</i>	1	0,823	0,82332	14,96	0,000
<i>alan_sh</i>	1	0,239	0,23853	4,33	0,038
<i>yumuşaklık_sh</i>	1	1,195	1,19496	21,71	0,000
<i>içbükeylik_sh</i>	1	0,840	0,83983	15,26	0,000
<i>yarıçap_maks</i>	1	2,813	2,81267	51,10	0,000
<i>doku_maks</i>	1	1,747	1,74723	31,74	0,000
<i>alan_maks</i>	1	1,546	1,54605	28,09	0,000

ıçbükeylik_maks	1	0,700	0,70012	12,72	0,000
ıçbükey_nkt_maks	1	0,260	0,26033	4,73	0,030
simetri_maks	1	0,687	0,68728	12,49	0,000
fraktal_boyut_mak	1	0,481	0,48082	8,73	0,003
Error	555	30,551	0,48082		
Total	568	133,012			

Tablo 7. İleri doğru seçim yöntemi ile deęişken seçimi sonunda elde edilmiş olan 17 adet girdi deęişkeni

Source	DF	SS	MS	F-Value	P-Value
Regression	17	102,802	6,04716	110,29	0,000
yarıçap_ort	1	0,090	0,08978	1,64	0,201
yoęunluk_ort	1	1,445	1,44546	26,36	0,000
ıçbükeylik_ort	1	0,142	0,14161	2,58	0,109
ıçbükey_nkt_ort	1	0,088	0,08784	1,60	0,206
yarıçap_sh	1	0,182	0,18248	3,33	0,069
alan_sh	1	0,040	0,04001	0,73	0,393
yumuşaklık_sh	1	1,050	1,05018	19,15	0,000
yoęunluk_sh	1	0,008	0,00846	0,15	0,695
ıçbükeylik_sh	1	0,672	0,67192	12,25	0,001
ıçbükey_nkt_sh	1	0,189	0,18882	3,44	0,064
yarıçap_maks	1	1,460	1,46007	26,63	0,000
doku_maks	1	1,578	1,57841	28,79	0,000
alan_maks	1	1,385	1,38464	25,25	0,000
ıçbükeylik_maks	1	0,236	0,23628	4,31	0,038
ıçbükey_nkt_maks	1	0,025	0,02543	0,46	0,496
simetri_maks	1	0,706	0,70575	12,87	0,000
fraktal_boyut_maks	1	0,527	0,52681	9,61	0,002
Error	551	30,211	0,05483		
Total	568	133,012			

Son olarak geriye doğru eleme yöntemi ile değişken seçimi sonunda elde edilmiş olan 13 adet girdi değişkeni Tablo 8'deki ANOVA tablosunda verilmiştir. Bu üç yöntemle elde edilen değişkenlerle YSA modelleri oluşturulup performansları önerilen korelasyon tabanlı yöntemin performansı ile karşılaştırılmıştır.

Tablo 8. Geriye doğru eleme yöntemi ile değişken seçimi sonunda elde edilmiş olan 13 adet girdi değişkeni

Source	DF	SS	MS	F-Value	P-Value
Regression	13	102,608	7,89292	144,08	0,000
yarıçap_ort	1	0,271	0,27104	4,95	0,027
yoğunluk_ort	1	2,033	2,03304	37,11	0,000
içbükey_nkt_ort	1	1,499	1,49871	27,36	0,000
yarıçap_sh	1	0,528	0,52800	9,64	0,002
yumuşaklık_sh	1	1,125	1,12547	20,54	0,000
içbükeylik_sh	1	0,906	0,90597	16,54	0,000
içbükey_nkt_sh	1	0,294	0,29398	5,37	0,021
yarıçap_maks	1	3,542	3,54226	64,66	0,000
doku_maks	1	1,688	1,68843	30,82	0,000
alan_maks	1	3,217	3,21682	58,72	0,000
içbükeylik_maks	1	1,128	1,12755	20,58	0,000
simetri_maks	1	0,724	0,72356	13,21	0,000
fraktal_boyut_maks	1	0,462	0,46180	8,43	0,004
Error	555	30,404	0,05478		
Total	568	133,012			

4.3. YSA Modellerinin Oluşturulup Performanslarının Karşılaştırılması ve Tartışma

İyi huylu/kötü huylu hücre sınıflandırmasında önerilen yöntemle elde edilmiş girdi değişkenleri kullanılarak oluşturulmuş YSA modelinin performansı altı farklı modelle karşılaştırılmıştır. Bu modellerden üç tanesi şu şekildedir: 1) tüm 30 girdi değişkeninin kullanıldığı model, 2) ortalama ve maksimum değerleri içeren 20 adet girdi değişkeninin kullanıldığı model, ve 3) sadece ortalama değerleri içeren 10 adet girdi değişkeninin kullanıldığı model. Diğer üç modeldeki girdi değişkenleri ise bu çalışmada önerilen korelasyon tabanlı filtreleme yönteminin *model tabanlı* yöntemlere göre performansını belirlemek için *sarmal* (wrapped) ve *gömülü* (embedded) yöntemler kullanılarak seçilmiştir.

Böylece son üç model sarmal yöntemlerden *ileri doğru seçim* (forward selection) ve *geriye doğru eleme* (backward elimination) yöntemleri ile, ve gömülü yöntemlerden *aşamalı regresyon* yöntemi ile seçilmiş girdi değişkenleri ile oluşturulmuştur. Bu son üç modelle elde edilmiş olan girdi değişkenleri Bölüm 4.2'de verilmiştir.

Tablo 9. Tüm modeller için doğru sınıflandırma yüzdeleri

Doğru Sınıflandırma Yüzdesi (Doğruluk)				
Girdi sayısı	Veri seti	10 Nöron	15 Nöron	20 Nöron
10	60-20-20	0,918	0,927	0,930
	70-15-15	0,922	0,930	0,923
	80-10-10	0,926	0,927	0,933
20	60-20-20	0,946	0,946	0,938
	70-15-15	0,945	0,948	0,941
	80-10-10	0,946	0,958	0,954
30	60-20-20	0,961	0,968	0,958
	70-15-15	0,970	0,963	0,966
	80-10-10	0,958	0,960	0,959
Önerilen yöntem (9 değişken)	60-20-20	0,950	0,939	0,936
	70-15-15	0,940	0,930	0,940
	80-10-10	0,940	0,947	0,940
Aşamalı regresyon (13 değişken)	60-20-20	0,957	0,962	0,965
	70-15-15	0,965	0,963	0,963
	80-10-10	0,970	0,968	0,963
İleri doğru seçim (17 değişken)	60-20-20	0,964	0,952	0,961
	70-15-15	0,966	0,963	0,962
	80-10-10	0,962	0,962	0,970
Geriye doğru eleme (13 değişken)	60-20-20	0,955	0,961	0,966
	70-15-15	0,957	0,957	0,962
	80-10-10	0,961	0,954	0,961

YSA modelleri oluşturulurken veri seti farklı eğitim-test yüzdelerine bölünmüş ve gizli katmanda farklı nöron sayıları denenmiştir. Kullanılan eğitim, geçerlilik ve test yüzdeleri 80-10-10, 70-15-15 ve 60-20-20 şeklindedir. Modellerde gizli katmandaki nöron sayısı olarak 10, 15 ve 20 nöron denenmiştir. Modeller 100'er defa çalıştırılmış ve ilgili performans ölçütü için ortalama değerler hesaplanmıştır. Modellerin sınıflandırma performanslarını karşılaştırmak için doğruluk veya doğru sınıflandırma oranı (accuracy), duyarlılık (recall), kesinlik (precision) ve F1-skoru ölçütleri kullanılmıştır. Tüm yedi model için doğru sınıflandırma oranları Tablo 9'da, duyarlılık (recall), kesinlik (precision) değerleri Tablo 10'da ve son olarak F1 skorları Tablo 11'da verilmiştir.

Tablo 9-11 incelendiği zaman genel olarak gizli katmandaki nöron sayısı arttıkça yani 10 nöron yerine 15 veya 20 nöron kullanıldığı zaman model performanslarının iyileştiği görülmektedir. Bu beklenen bir durumdur çünkü veri gizli katmanda daha fazla nöronla tahmin edilmeye çalışılmaktadır. Ancak nöron sayısını arttırmak modeli daha karmaşık bir yapıya büründürmektedir.

Diğer taraftan eğitim seti yüzdesini de arttırmak genel olarak model performanslarını iyileştirmektedir. Özellikle %60 eğitim verisi yerine daha büyük eğitim yüzdelerinde daha iyi sonuçların alındığı görülmektedir.

Doğru sınıflandırma yüzdeleri: Girdi değişkenleri açısından Tablo 9'da verilen doğru sınıflandırma yüzdeleri incelendiği zaman en yüksek sınıflandırma yüzdesinin 0,97 ile üç modele ait olduğu görülmektedir. Bu modellerden ilki 30 girdi değişkenli, veri seti 70-15-15 şeklinde bölünmüş ve gizli katmandaki nöron sayısı 10 olan model, diğerlerinin ise, aşamalı regresyonla elde edilmiş 13 değişkenli, veri seti 80-10-10 şeklinde bölünmüş ve gizli katmandaki nöron sayısı 10 olan model ile ileri doğru seçim yöntemi ile elde edilmiş 17 değişkenli, veri seti 80-10-10 şeklinde bölünmüş ve gizli katmandaki nöron sayısı 20 olan model olduğu görülmektedir. Burada 30 değişkenli model için sınıflandırma performansının yüksek olması girdi değişken sayısının fazla olmasına bağlanabilir. Çünkü modele girdi değişken eklendikçe modelin açıklanabilen oranı artmaktadır. Fakat önemli olan daha az girdi değişkeniyle iyi performanslar yakalayabilmektir. Diğer en iyi modelimiz ileri doğru seçim yöntemi ile elde edilmiş 17 girdi değişkenli, veri seti 80-10-10 şeklinde bölünmüş ve gizli katmandaki nöron sayısı 20 olan model demizdir. Bu modelin de performansının yüksek çıkması yine girdi değişken sayısı yüksekliği ile bağdaştırılabilir. Ayrıca bu modelde gizli

katmandaki nöron sayısı da yüksektir - 20 nöron. Bu nedenle bu modelde hem yüksek sayıda girdi değişken bulunmaktadır hem de model yapı olarak daha karmaşıktır.

En düşük doğru sınıflandırma yüzdesinin de 10 girdi değişkenli 60-20-20 şeklinde ayrılmış 10 nöronlu modele ait olduğu belirlenmiştir. Tablo 9'a bakıldığında genel olarak 10 girdi değişkenli YSA modelinin en düşük değerlere sahip olduğu görülmektedir. Böylece sadece ortalama değerleri içeren 10 adet girdi değişkeninin kullanıldığı modelin doğru yüzde performansının çok yüksek olmadığı söylenebilir. 30 girdi değişkenli, aşamalı regresyonla, ileri doğru seçimle ve geriye doğru eleme yöntemleriyle elde edilmiş modellerin doğru sınıflandırma performansları 0,95-0,97 arasındadır. Diğer taraftan bu çalışmada önerilen korelasyon tabanlı filtreleme yöntemi ile seçilen dokuz girdi değişkenli modelin performansı ise 0,93-0,95 arasında olup azımsanamayacak şekilde iyidir. Önerilen bu modelde girdi değişken sayısı (9 adet) karşılaştırılan diğer tüm modellere göre en düşüktür. Ayrıca önerilen model gizli katmanda sadece 10 nöronla 0,95 oranında doğru yüzde değeri elde edebilmiştir.

Duyarlılık, Kesinlik ve F1-Skor Değerleri: Duyarlılık ölçütü "Doğru olarak tahmin edilmesi gereken tüm veri noktalarından ne kadarını doğru olarak tahmin ettik?" sorusuna yanıt verir demiştik. Kanseri gibi ciddi bir hastalıkta duyarlılık önemlidir. Hastanın gerçekten kanserli olduğu ancak modelin bunu kansersiz olarak sınıflandırdığı düşünülürse bu bir sorun oluşturur. Bu sebeple duyarlılık değerinin olabildiğince yüksek tutulması oluşturulacak modeller için önemlidir. Tablo 10'da görüldüğü gibi duyarlılık değerleri 30 değişkenli model, aşamalı regresyon, ileri doğru seçim ve geriye doğru eleme yöntemleri ile elde edilen değişkenlerin kullanıldığı modellerde 0,89 - 0,935 arasında değişmektedir. En iyi model için 0,935 (Aşamalı regresyon ile seçim, 80-10-10 veri seti ve 10 nöron) olduğu görülmüştür. Diğer taraftan bu çalışmada önerilen yöntemle elde edilmiş 9 girdi değişkenli model için duyarlılık değerleri 0,845 ile 0,88 arasındadır. 13 girdi değişken kullanan en iyi modelin duyarlılık değerinin 0,935 olduğu düşünüldüğünde ve 30 girdi değişkeniyle eğitilen tam modelin bile duyarlılık değerlerinin 0,901-0,919 aralığında çıktığı göz önünde bulundurulduğunda, sadece 9 değişken ve gizli katmanda 10 nöronla eğitilmiş önerilen YSA modelinin duyarlılık performansının 0,88 olması iyi bir sonuçtur.

Tablo 10. Tüm modeller için duyarlılık ve kesinlik değerleri

Girdi sayısı	Veri seti	10 Nöron		15 Nöron		20 Nöron	
		Kesinlik	Duyarlılık	Kesinlik	Duyarlılık	Kesinlik	Duyarlılık
10	60-20-20	0,966	0,827	0,972	0,832	0,981	0,832
	70-15-15	0,977	0,820	0,973	0,842	0,969	0,825
	80-10-10	0,978	0,824	0,975	0,834	0,967	0,849
20	60-20-20	0,986	0,869	0,985	0,870	0,984	0,851
	70-15-15	0,984	0,866	0,971	0,880	0,978	0,862
	80-10-10	0,979	0,866	0,986	0,888	0,985	0,889
30	60-20-20	0,988	0,910	0,996	0,913	0,988	0,902
	70-15-15	0,997	0,919	0,992	0,908	0,981	0,927
	80-10-10	0,972	0,912	0,990	0,901	0,992	0,901
Önerilen yöntem (9 değişken)	60-20-20	0,988	0,880	0,981	0,858	0,983	0,845
	70-15-15	0,981	0,846	0,979	0,845	0,988	0,856
	80-10-10	0,980	0,863	0,989	0,867	0,982	0,868
Aşamalı regresyon (13 değişken)	60-20-20	0,988	0,899	0,989	0,912	0,984	0,923
	70-15-15	0,990	0,908	0,992	0,906	0,988	0,913
	80-10-10	0,986	0,935	0,988	0,925	0,984	0,911
İleri doğru seçim (17 değişken)	60-20-20	0,984	0,921	0,982	0,890	0,991	0,905
	70-15-15	0,988	0,915	0,991	0,906	0,991	0,907
	80-10-10	0,993	0,910	0,998	0,908	0,998	0,923
Geriye doğru eleme (13 değişken)	60-20-20	0,984	0,897	0,981	0,912	0,983	0,923
	70-15-15	0,988	0,897	0,967	0,913	0,989	0,907
	80-10-10	0,991	0,904	0,985	0,891	0,986	0,908

Benzer şekilde, kanser gibi hastalıklarda kötü huylu şekilde teşhis konulan hastaların da gerçekten kanserli olması tedavi başarısı açısından önemlidir. Diğer türlü yanlış tedavi uygulamaları yapılabilmektedir. Bu durum göz önüne alındığında teşhis konulurken kesinlik değeri de önem arz etmektedir. Bu noktada yine Tablo 10'da görüldüğü gibi önerilen modelimizin kesinlik değerlerinin 0,98-0,988 ile çok yüksek olduğunu görebilmekteyiz. En yüksek kesinlik değerine sahip modelin ileri doğru seçim yöntemiyle belirlenmiş 17 girdi değişkenli model olduğu ve değerinin 0,997 olduğu göz önünde bulundurulduğunda önerilen modelimizin kesinlik açısından performansının çok yüksek olduğu sonucuna varılabilir.

Tablo 11. Tüm modeller için F1-skorumları

F1 skorları				
Girdi değişken sayısı	Veri seti	10 Nöron	15 Nöron	20 Nöron
10	60-20-20	0,891	0,896	0,900
	70-15-15	0,892	0,903	0,891
	80-10-10	0,895	0,899	0,904
20	60-20-20	0,924	0,924	0,913
	70-15-15	0,921	0,923	0,916
	80-10-10	0,919	0,935	0,935
30	60-20-20	0,947	0,953	0,943
	70-15-15	0,956	0,948	0,953
	80-10-10	0,941	0,944	0,944
Önerilen yöntem (9 değişken)	60-20-20	0,931	0,915	0,909
	70-15-15	0,908	0,907	0,918
	80-10-10	0,922	0,918	0,908
Aşamalı regresyon (13 değişken)	60-20-20	0,942	0,949	0,953
	70-15-15	0,947	0,947	0,949
	80-10-10	0,960	0,956	0,946
İleri doğru seçim (17 değişken)	60-20-20	0,952	0,934	0,946
	70-15-15	0,950	0,947	0,947
	80-10-10	0,950	0,951	0,959
Geriye doğru eleme (13 değişken)	60-20-20	0,939	0,945	0,952
	70-15-15	0,941	0,939	0,946
	80-10-10	0,946	0,936	0,945

Kesinlik ve duyarlılık ölçütlerinin arasında bir ödünleşme olduğu için ve bu iki ölçüt de kanser teşhisinde önemli olduğundan bu iki ölçütü birleştiren F1-skoru aslında bakılması gereken en önemli ölçüdür. Eğitilen farklı modeller için F1-skorları Tablo 11'de verilmiştir. F1-skorları incelendiğinde en yüksek skora (0,96) sahip modelin aşamalı regresyonla seçilen 13 girdi değişkeniyle eğitilmiş model olduğu görülmektedir. Bu çalışmada önerilen yöntemin F1-skoru 0,931 olup yeterince yüksek bir değere sahiptir. 30 girdi değişkenli model, aşamalı regresyon, ileri doğru seçim ve geriye doğru eleme yöntemleriyle seçilmiş girdi değişkenleriyle eğitilmiş modellerde F1-skor aralığı 0,939 - 0,96 şeklindedir. 10 girdi değişkenli model için F1-skor değerlerinin 0,891 ile 0,904 arasında olduğu görülmektedir. Bu durumlar göz önüne alındığında sadece 9 girdi değişkenli modelimizin F1-skoru yönünden iyi bir performans gösterdiği sonucuna ulaşılabilir.

5. SONUÇLAR

Sınıflandırma, uygulamada sıklıkla karşılaşılan karar verme faaliyetlerinden biridir. Son yıllarda YSA modelleri sınıflandırma problemlerinde kullanılmaya başlanmıştır. Diğer taraftan girdi değişkenlerinin seçimi, oluşturulacak modellerin en uygun formunun belirlenmesinde çok önemli bir husustur. Bu nedenle, mevcut çalışmada başlıca amaç, yüksek miktardaki olası girdi değişken sayısını, bu değişkenler arasındaki korelasyonları göz önünde bulundurarak azaltarak sınıflandırma performansı yüksek YSA modelleri elde etmektir.

Kanser gibi hayati önemi olan hastalıkların teşhis aşamasında kullanılacak sınıflandırma modeli için performans ölçütü olarak "Yaptığımız tüm tahminlerden ne kadarı doğru çıktı" sorusuna cevap veren *doğruluk değeri* önemlidir. Bu noktada, önerilen korelasyon tabanlı filtreleme yöntemi ile seçilen dokuz girdi değişkenli modelimiz için doğruluk değeri 0,93-0,95 arasında olup belirgin şekilde iyidir. Diğer taraftan aşamalı regresyonla, ileri doğru seçimle ve geriye doğru eleme yöntemleriyle girdi değişkenleri seçilmiş modeller ile tüm girdi değişkenlerini içeren modelin doğru sınıflandırma performansları 0,95-0,97 arasındadır. Ortalama değerleri içeren 10 adet girdi değişkeninin kullanıldığı model için doğruluk değerleri 0,918-0,933 arasında olup önerilen modelimizin gerisinde performans sergilemiştir. Önerilen modelimizde girdi değişken sayısının (dokuz adet) karşılaştırılan diğer tüm modellere göre en düşük olduğu ve gizli katmanda sadece 10

nöronla 0,95 oranında doğru yüzde değeri elde edebildiği göz önünde bulundurulduğunda önerilen modelin doğruluk performansının yüksek olduğu söylenebilir.

Teşhis aşamasında kullanılacak sınıflandırma modelinin, doğruluk değerinin yanında *kesinlik* ve *duyarlılık değerleri* de önemlidir. Hastanın gerçekten kanserli olduğu ancak modelin bunu kansersiz olarak sınıflandırdığı düşünülürse bu bir sorun oluşturur. Bu sebeple duyarlılık değerinin olabildiğince yüksek tutulması oluşturulacak modeller için önemlidir. Benzer şekilde kanser gibi hastalıklarda kötü huylu şekilde teşhis konulan hastaların da gerçekten kanserli olması tedavi başarısı açısından önemlidir. Diğer türlü yanlış tedavi uygulamaları yapılabilmektedir. Bu durum göz önüne alındığında teşhis konulurken kesinlik değeri de önem arz etmektedir. "Yaptığımız tüm kötü huylu şeklindeki tahminlerden ne kadarı doğru çıktı?" sorusuna yanıt veren kesinlik değerinin önerilen modelimiz için 0,98-0,988 aralığında ve çok yüksek olduğu belirlenmiştir. Benzer şekilde "Doğru olarak tahmin edilmesi gereken tüm veri noktalarından ne kadarını doğru olarak tahmin ettik?" sorusuna yanıt veren duyarlılık değeri modelimiz için 0,88 bulunmuştur. Karşılaştırılan modeller için elde edilen performanslar incelendiğinde en yüksek 0,935 değerinin elde edilebildiği ve bu değer 13 girdi değişken kullanan modele ait olduğu düşünüldüğünde ve 30 girdi değişkeniyle eğitilen modelin bile duyarlılık değerlerinin 0,901-0,919 aralığında çıktığı göz önünde bulundurulduğunda, sadece 9 değişken ve gizli katmanda 10 nöronla eğitilmiş önerilen YSA modelimizin duyarlılık performansının 0,88 olması iyi bir sonuçtur.

Son olarak kesinlik ve duyarlılık ölçütleri arasında bir ödünleşme olduğu için, yani genel olarak biri artarken diğeri azaldığı için, bu iki ölçütün de önemli olduğu kanser teşhisi gibi durumlarda iki ölçütü birleştirerek ele alan *F1-skoru* belirlenmelidir. Bu çalışmada önerilen yöntemin F1-skoru 0,907-0,931 olup yeterince yüksek bir değere sahiptir. 30 girdi değişkenli model, aşamalı regresyon, ileri doğru seçim ve geriye doğru eleme yöntemleriyle seçilmiş girdi değişkenleriyle eğitilmiş modellerde F1-skor aralığı 0,939 - 0,96 şeklindedir. 10 girdi değişkenli model için F1-skor değerlerinin 0,891 ile 0,904 arasında olduğu görülmektedir. Bu durumlar göz önüne alındığında sadece 9 girdi değişkenli modelimizin F1-skoru yönünden iyi bir performans gösterdiği sonucuna ulaşılabilir.

Sonuç olarak bu çalışmada ele alınan sınıflandırma problemi için çok sayıdaki olası girdi değişken sayısı düşük maliyetli, hızlı ve kolay anlaşılır bir yöntemle azaltılarak yalın ve etkin bir YSA modeli elde edilmiştir. Literatürde girdi değişkenlerinin seçiminde *model tabanlı* yöntem uygulamalarındaki hesaplama süresi uzunluğu ve iş yükü fazlalığı gibi sebeplerden dolayı daha maliyet etkin ve daha kısa sürede verimli sonuç alınabilecek

yöntemlere gereksinim duyulmuştur. Bu durumlarda girdi değişken seçimi için genellikle filtreleme yöntemleri kullanılmıştır. Benzer şekilde bu çalışmada da korelasyon-hipotez testi tabanlı bir filtreleme yöntemi önerilmiş ve etkinliği farklı girdi değişken setlerini içeren YSA modelleri ile karşılaştırılmıştır. Modeller için performans karşılaştırmasında önerilen model için tüm girdi değişkenleri içeren modele göre ve model tabanlı seçim yöntemlerinden aşamalı regresyonla, ileri doğru seçimle ve geriye doğru eleme yöntemleri ile elde edilmiş olan modellerle benzer oranlarda ve yüksek doğruluk, kesinlik, duyarlılık ve F1-skor değerleri elde edilmiştir. Karşılaştırılan modeller içinde önerilen dokuz-değişkenli modelin değişken sayısının en düşük olduğu, yani en sade model olduğu ve gizli katmanda sadece 10 nöronla bile iyi bir sınıflandırma performansına sahip olduğu göz önüne alındığında bu modelin düşük maliyetle ve hızlı bir şekilde anlaşılır sınıflandırma modelleri elde etmede verimli olacağı belirlenmiştir.

Gelecek çalışmalarda, bu çalışmada önerilen yöntem, farklı alanlardaki sınıflandırma problemleri için farklı veri setleri kullanılarak uygulanabilir ve etkinliği farklı filtreleme veya model tabanlı yöntemlerle karşılaştırılabilir. Bu çalışmada hipotez testleri ile sadece ortalama değerler göz önünde bulundurarak ayırma durumu incelendiği için gelecek çalışmalarda korelasyon tabanlı farklı filtreleme yöntemleri önerilerek bu yöntemler model tabanlı yöntemlerle karşılaştırılabilir. Son olarak, sınıflandırma için bu çalışmada kullanılan modellere benzer YSA kombinasyonları en iyi parametreler seçilecek şekilde test edilebilir.

ETİK BEYAN

“YSA Sınıflandırma Modellerinde Korelasyon-Hipotez Testi Tabanlı Filtreleme Yoluyla Girdi Seçimi” başlıklı çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş, toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

KAYNAKÇA

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. and Arshad, H. (2018), State-of-the-art in artificial neural network applications: A survey, *Heliyon*, 4(11).

- Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A. and San Tan, R. (2017), A deep convolutional neural network model to classify heartbeats, *Computers in Biology and Medicine*, 89, 389-396.
- Alpaydin, E. (2020), *Introduction to Machine Learning*, MIT Press, Cambridge, Massachusetts, ABD.
- Alshanbari, E., Alamri, H., Alzahrani, W. and Alghamdi, M. (2021), Breast cancer classification using convolutional neural network, *International Journal of Computer Science and Network Security*, 21(6), 101-106.
- Arı, A. and Hanbay, D. (2019), Tumor detection in MR images of regional convolutional neural networks, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34(3), 1395-1408.
- Basu, J.K., Bhattacharyya, D. and Kim, T.H. (2010), Use of artificial neural network in pattern recognition, *International Journal of Software Engineering and Its Applications*, 4, 23–34.
- Bılgıç, B. (2021), Comparison of breast cancer and skin cancer diagnoses using deep learning method, 29, *Signal Processing and Communications Applications Conference (SIU)*, 9-11 Haziran, Istanbul, Türkiye.
- Calkins, K.G. (2005), *Correlation Coefficients*, Erişim adresi: <https://www.andrews.edu/~calkins/math/edrm611/edrm05.htm>
- Chandrashekar, G. ve Sahin, F. (2014), A survey on feature selection methods. *Computers & electrical engineering*, 40(1), 16-28.
- Ciregan, D., Meier, U. and Schmidhuber, J. (2012), Multi-column deep neural networks for image classification, *2012 IEEE Conference on computer vision and pattern recognition*, 16-21 Haziran, Providence, RI, ABD, 3642-3649.
- Chuang, C. L. and Huang, S. T. (2011), A hybrid neural network approach for credit scoring, *Expert Systems*, 28(2), 185-196.
- Chou, S.M., Lee, T.S., Shao, Y.E. and Chen, I.F. (2004), Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regressions lines, *Expert Systems with Applications*, 27(1), 133-142.

- Dongare, A.D., Kharde, R.R. and Kachare, A.D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2(1), 189-194.
- Dreiseitl, S. and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review, *Journal of Biomedical Informatics*, 35(5-6), 352-359.
- Emanet, S., Baydoğmuş, G.K. and Demir, Ö. (2021), Öznitelik seçme yöntemlerinin makine öğrenmesi tabanlı saldırı tespit sistemi performansına etkileri, *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 12(5), 743-755.
- Fernando, T.M.K.G., Maier, H.R. and Dandy, G.C. (2009), Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach, *Journal of Hydrology*, 367(3-4), 165-176.
- Gao, C., Sun, H., Wang, T., Tang, M., Bohnen, N.I., Müller, M.L., ... and Dinov, I.D. (2018), Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. *Scientific reports*, 8(1), 7129.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep learning*, MIT Press, Cambridge, Massachusetts, ABD.
- Haddadnia, J., Hashemian, M. and Hassanpour, K. (2012), Diagnosis of breast cancer using a combination of genetic algorithm and artificial neural network in medical infrared thermal imaging, *Iranian Journal of Medical Physics*, 9(4), 265-274.
- Irmak, M.C., Taş, M.B.H., Turan, S. and Haşiloğlu, A. (2021), Comparative breast cancer detection with artificial neural networks and machine learning methods, 29. *Signal Processing and Communications Applications Conference (SIU)*. 9-11 Haziran, İstanbul, Türkiye, 1-4
- Jerez-Aragonés, J.M., Gómez-Ruiz, J.A., Ramos-Jiménez, G., Muñoz-Pérez, J. and Alba-Conejo, E. (2003), A combineneural network and decision trees model for prognosis of breast cancer relapse, *Artificial Intelligence in Medicine*, 27(1), 45-63.
- LaMorte, W.W. (2021), *Correlation and Regression*, Boston University School of Public Health, Erişim adresi: <https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717->

QuantCore/PH717-Module9-Correlation-Regression/PH717-Module9-Correlation-Regression4.html.

- Liao, S.H. and Wen, C.H. (2007), Artificial neural networks classification and clustering of methodologies and applications–literature analysis from 1995 to 2005, *Expert Systems with applications*, 32(1), 1-11.
- May, R., Dandy, G. and Maier, H. (2011), Review of input variable selection methods for artificial neural networks, *Artificial neural networks-methodological advances and biomedical applications*, 10(1), 19-45.
- Öztemel, E. (2003), *Yapay Sinir Ağları*, Papatya Yayıncılık, İstanbul, Türkiye.
- Ramani, R., Devi, K.V. and Soundar, K.R. (2020), MapReduce-based big data framework using modified artificial neural network classifier for diabetic chronic disease prediction, *Soft Computing*, 24(21), 16335-16345.
- Remeseiro, B. and Bolon-Canedo, V. (2019), A review of feature selection methods in medical applications. *Computers in biology and medicine*, 112, 103375.
- Ryu, Y.U., Chandrasekaran, R. and Jacob, V.S. (2007), Breast cancer prediction using the isotonic separation technique, *European Journal of Operational Research*, 181(2), 842-854.
- Sezer, E. and Çakır, Ö. (2022), Sınıflandırma amaçlı değişken alt kümesi seçimi: bir bankacılık uygulaması, *Dicle Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12(24), 480-498.
- Snelder, T.H., Dey, K.L. and Leathwick, J.R. (2007), A procedure for making optimal selection of input variables for multivariate environmental classifications. *Conservation Biology*, 21(2), 365-375.
- Snieder, E., Shakir, R. and Khan, U.T. (2020), A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. *Journal of Hydrology*, 583, 124299.
- Solorio-Fernández, S., Carrasco-Ochoa, J.A. and Martínez-Trinidad, J.F. (2020), A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2), 907-948.
- Suzuki, K. (Ed.). (2011), *Artificial neural networks: methodological advances and biomedical applications*. BoD–Books on Demand.

UC Irvine Machine Learning Repository (2022), Erişim adresi:
<https://archive.ics.uci.edu/datasets>.

Uluskan, M. (2020), Artificial neural networks as a quality loss function for six sigma, *Total Quality Management and Business Excellence*, 31(15-16), 1811-1828.

Wu, Y.C. and Feng, J.W. (2018), Development and application of artificial neural network. *Wireless Personal Communications*, 102, 1645-1656.

Zhang, G.P. (2000), Neural networks for classification: A survey, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(4), 451-462.

Zhang, Y.D., Satapathy, S.C., Guttery, D.S., Górriz, J.M. and Wang, S.H. (2021), Improved breast cancer classification through combining graph convolutional network and convolutional neural network, *Information Processing and Management*, 58(2), 102439.

Zou, J., Han, Y. and So, S.S. (2009), Overview of artificial neural networks. *Artificial neural networks: methods and applications*, 14-22.

Yayın Geliş Tarihi (Submitted): 29/11/2023

Yayın Kabul Tarihi (Accepted): 17/01/2024

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

Please Cite As/Atıf için:

Baylan, P. and Demirel, N. (2024), Quantifying the impact of risk factors on direct compensation property damage in canadian automobile insurance, *Nicel Bilimler Dergisi*, 6(1), 103-127. doi: 10.51541/nicel.1397941

QUANTIFYING THE IMPACT OF RISK FACTORS ON DIRECT COMPENSATION PROPERTY DAMAGE IN CANADIAN AUTOMOBILE INSURANCE

Pervin Baylan¹ and Neslihan Demirel²

ABSTRACT

This study presents a statistical analysis assessing the impact of various risk factors on direct compensation property damage (DCPD) claims in private passenger vehicle accidents. Using automobile insurance data in Ontario, Canada for the decade years period between 2003 and 2012, a statistical model of property damage was explored via a generalized linear binary logit mixed model and considered the imbalance between the classes of insureds. The results indicate that several risk factors have a significant impact on the likelihood of DCPD claims, including usage, training, outstanding loss, and incurred loss. The effects of these risk factors were observed under the weights — the number of trials used to generate each success proportion — in the different classes of insureds. The generalized linear mixed models (GLMMs) analysis provides a powerful tool for quantifying the impact of risk factors on binary outcomes, which are called DCPD claims and property damage (PD) claims covered by third-party liability (TPL) insurance. These models can also inform insurance underwriting and policy design, focusing on identifying the most significant risk factors. The performance metrics calculated by considering the class imbalance in binary outcomes verify the resulting model's ability to accurately predict classes. The *F1* score, an evaluation metric to measure the performance of classification, was calculated as 0.934. In addition, *PR AUC*, which is the

¹ Corresponding Author, Research Assistant, Department of Statistics, Faculty of Science, Dokuz Eylul University, Izmir, Turkiye. ORCID ID: <https://orcid.org/0000-0003-2660-3814>

² Prof.Dr., Department of Statistics, Faculty of Science, Dokuz Eylul University, Izmir, Turkiye. ORCID ID: <https://orcid.org/0000-0002-5394-4721>

area under the Precision-Recall (PR) curve, was computed as 0.953. These high scores indicate that the resulting model performs well in the classification. The other metrics also support the classification accuracy of this model. The findings of the analysis can help insurers better understand the underlying drivers of property damages and develop more accurate and effective strategies for risk mitigation. Furthermore, this study highlights the importance of developing class-specific risk assessment models to account for the imbalance across different classes.

Keywords: Binary Logit Model, Direct Compensation Property Damage, Generalized Linear Mixed Model, Third-Party Liability Insurance, Unbalanced Panel Data.

KANADA OTOMOBİL SİGORTASINDA RİSK FAKTÖRLERİNİN DOĞRUDAN TAZMİN EDİLEN MADDİ HASAR ÜZERİNDEKİ ETKİSİNİN DEĞERLENDİRİLMESİ

ÖZ

Bu çalışma, özel binek araç kazalarında çeşitli risk faktörlerinin doğrudan tazmin edilen maddi hasarlar (Direct Compensation Property Damage - DCPD) üzerindeki etkisini değerlendiren istatistiksel bir analiz sunmaktadır. 2003 ile 2012 yılları arasındaki on yıllık döneme ait Ontario, Kanada'daki otomobil sigortası verileri kullanılarak, genelleştirilmiş doğrusal ikili logit karma model aracılığıyla maddi hasarın istatistiksel bir modeli araştırılmış ve sigortalıların sınıfları arasındaki dengesizlik dikkate alınmıştır. Sonuçlar, kullanım amacı, sürücü eğitimi, muallak hasar ve gerçekleşen hasar dahil olmak üzere çeşitli risk faktörlerinin DCPD hasarlarının olasılığı üzerinde önemli bir etkiye sahip olduğunu göstermektedir. Bu risk faktörlerinin etkileri, farklı sigortalı sınıflarındaki ağırlıklar — her bir başarı oranını oluşturmak için kullanılan deneme sayısı — altında gözlemlenmiştir. Genelleştirilmiş doğrusal karma modeller (GLMMs) analizi, risk faktörlerinin üçüncü şahıs sorumluluk (TPL) sigortası kapsamındaki DCPD hasarları ve maddi hasarlar (PD) olarak adlandırılan ikili sonuçlar üzerindeki etkisinin değerlendirilmesinde güçlü bir araçtır. Bu modeller, en önemli risk faktörlerini belirlemeye odaklanarak sigorta risk değerlendirmesine ve poliçe tasarımına da bilgi sağlayabilir. İkili sonuçlardaki sınıf dengesizliği dikkate alınarak hesaplanan performans ölçümleri, elde edilen modelin sınıfları doğru tahmin etme yeteneğini doğrulamaktadır. Sınıflandırma performansını ölçmeye yönelik değerlendirme ölçümü olan *F1* skoru 0,934 olarak hesaplanmıştır. Ayrıca, Kesinlik-Duyarlılık (Precision-Recall (PR))

eğrisinin altında kalan alan olan $PR AUC$ ise 0,953 olarak elde edilmiştir. Bu yüksek skorlar, elde edilen modelin sınıflandırmada iyi performans gösterdiğine işaret etmektedir. Diğer ölçümler de, bu modelin sınıflandırma doğruluğunu desteklemektedir. Analizin bulguları, sigortacıların maddi hasarların altında yatan nedenleri daha iyi anlamalarına ve risk azaltımı için daha doğru ve etkili stratejiler geliştirmelerine yardımcı olabilir. Ayrıca bu çalışma, farklı sınıflar arasındaki dengesizliği hesaba katmak için sınıfa özgü risk değerlendirme modellerinin geliştirilmesinin önemini vurgulamaktadır.

Anahtar Kelimeler: İkili Logit Model, Doğrudan Tazmin Edilen Maddi Hasar, Genelleştirilmiş Doğrusal Karma Model, Üçüncü Şahıs Sorumluluk Sigortası, Dengesiz Panel Veri

1. INTRODUCTION

Direct compensation property damage (DCPD) is a type of automobile insurance coverage that is designed to provide compensation to policyholders for damages to their vehicles caused by another driver in an accident. Under DCPD coverage, the policyholders' own insurer handles the claim and pays for the damages up to the limit of their coverage in cases where the accident was caused by another driver and was not their own fault; instead of seeking compensation from the other driver's insurance company. This coverage involves only property damage (PD) and not bodily injury claims occurring in a car accident; while enabling the repair of damage on the vehicle of the policyholders faster, without the delays and complications which might arise when dealing with another driver's insurer. Therefore, being an efficient and fair approach to insurance claims and vehicle repairs, DCPD coverage is available in several provinces in Canada, including Ontario, Quebec, Nova Scotia, New Brunswick and Prince Edward Island. If the policyholders are at fault for the accident, they will need to rely on other types of coverage, such as collision or liability insurance, to cover the cost of damages.

One of the major problems facing actuaries in third-party liability (TPL) insurance is the building of an accurate mathematical model to calculate insurance premiums. This is because it is essential to strike a balance between charging premiums which are affordable for policyholders and generating enough revenue to cover the costs of potential claims and provide a profit for the insurer. To develop an accurate mathematical model, actuaries should consider various risk factors that might influence the likelihood and cost of claims. Accurate

assessment of risk factors is a complex process which involves analyzing historical claims data. The improper models built in the analysis of the historical claims data lead to the premiums being determined lower than they should be, and thus increase the risk of sector failure. Overall, the accurate assessment of risk factors and the development of predictive models that estimate the likelihood of an insured event are crucial components for insurers in the automobile insurance sector, in terms of effectively managing their risk and providing their policyholders with affordable coverage. Actuaries typically use statistical models to calculate insurance premiums; considering the estimated risk of an insured and the potential cost of a claim. By using statistical models to price insurance premiums that reflect the true risk of potential claims, actuaries can help insurers to provide affordable coverage to policyholders; while also ensuring the long-term stability and success of the insurance industry.

Various problems in actuarial science rely on the creation of a mathematical model that can be used in premium pricing. The accurate calculation of premiums for compulsory TPL insurance is particularly important because this type of insurance has a significant impact on the non-life premium income of insurers. By improving the premium evaluation for this line of business, the potential financial losses of the insurance sector can be prevented. DCPD is a mandatory component of automobile insurance in Ontario and is included in all basic auto policies along with TPL insurance. Therefore, it has a considerable share of the yearly non-life premium income. Quantifying the impact of risk factors on the likelihood of DCPD claims versus PD claims covered by TPL insurance can help insurers make more informed decisions about insurance underwriting and policy design. By taking these risk factors into account, the actuaries can calculate insurance premiums appropriate for the level of risk being assumed by the insurer, so that identifying the most significant risk factors leads to a more efficient and effective insurance market.

The use of generalized linear mixed models (GLMMs) in actuarial science allows for the incorporation of risk factors into the premium pricing process, improving the accuracy of insurance premiums and reducing the risk of financial losses for insurers. Most actuarial pricing techniques in use today are based on the generalized linear model research of Nelder and Wedderburn (1972) and McCullagh and Nelder (1989). Over the last 30 years, generalized linear models (GLMs) have been one of the most commonly used statistical tools for modeling actuarial data in actuarial work. In an actuarial context, Haberman and Renshaw (1996) provide an overview of the applications of GLMs in actuarial science and show that

GLMs are not limited to models for automobile insurance premiums. Embrechts and Wüthrich (2022) in the case of non-life insurance demonstrate how combining traditional statistical methods, such as GLMs with neural networks, improves comprehension and interpretation of actuarial data.

Many actuarial problems have a data structure that includes repeated measurements, especially panel data, which are characterized by a tendency to correlate repeated observations on a group of subjects over time. This correlation between observations on the same subject leads to extra difficulties during the analysis. Since the assumption of independence is not fulfilled in GLMs due to this correlation, GLMMs, which are extensions of GLMs, can be used for correlated data. Statistical techniques are considered for modeling panel data within the framework of GLMs in Antonio and Beirlant (2007). They also discuss the advantages of the GLM approach and represent the usage of GLMMs in actuarial mathematics. Miao (2018), using a hierarchical generalized linear model, shows that GLMMs can more effectively reflect the differences between distinct risk individuals as well as the heterogeneity and correlation of risk individual loss over multiple insurance periods.

The GLMM approach has been frequently used to model actuarial data and provides a useful approach in the analysis of unbalanced panel data. This approach procures extra flexibility in estimating the model and helps eliminate the extra complexity resulting from the internal correlation of each subject. Yau et al. (2003) consider the application of the GLMM approach to the analysis of repeated claim frequency data in motor insurance. All of these mentioned features also make GLMMs a powerful tool for identifying risk factors. Antonio and Valdez (2012) present a risk classification based on GLMs in insurance. Garrido et al. (2016) explore how the assumption that claim counts and amounts are independent in non-life insurance can be relaxed via GLMs while incorporating rating factors into the model.

The motivation of this study is to create a statistical model within the framework of GLMs that identifies the impact of the most important risk factors on DCPD claims in private passenger vehicle accidents. For this purpose, the paper is structured as follows. Section 2 describes the methodological framework used in this study. Each subsection of this part mentions the basic concepts of GLMMs, the structure of automobile insurance data provided by a Canadian insurance company, statistical analysis of binary outcomes such as DCPD claims and PD claims covered by TPL insurance, and how risk factors are identified. In addition, the performance metrics used in this study are explained in detail. Section 3 presents the results of the model developed for estimating the likelihood of DCPD claims. Section 4

introduces the main conclusions of this study. The acronyms used in this study are listed in Table 1.

Table 1. A list of acronyms and definitions used throughout the paper

Acronym	Definition
DCPD	Direct Compensation Property Damage
PD	Property Damage
TPL	Third-Party Liability
GLMM	Generalized Linear Mixed Model
GLM	Generalized Linear Model
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
LRT	Likelihood Ratio Test
SN	Sensitivity
SP	Specificity
P	Precision
ACC	Accuracy
BA	Balanced Accuracy
AUC	Area Under the ROC Curve
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ROC	Receiver Operating Characteristic
PR	Precision-Recall
PR AUC	Area Under the PR Curve

2. MATERIAL AND METHODS

2.1. Generalized Linear Mixed Models

A logistic regression model that can be viewed as a GLM is generally used to model binary or more than two categories under the assumption of independence. However, in many actuarial problems, observations on the same subject over time are often correlated. In these circumstances, the logistic GLM might not be appropriate to model repeated observations due to the structure of correlation between observations of the same subject. GLMs are extended to GLMMs by including random effects in the linear estimator that determine the inherent

correlation between observations on the same subject. Thus, the random effect also accounts for unobserved heterogeneity between subjects due to unobserved characteristics.

GLMM provides a more flexible approach in terms of normality and homoscedasticity assumptions since it is extended to distributions from the exponential family. In addition, in GLMM, the additive effect of independent variables is modeled on a transformation of the mean (Antonio and Beirlant, 2007).

Here, the model is extended to include random effects since the focus will be on longitudinal design, which is repeated observations on a group of subjects over time. We consider a model where the conditional distribution of \mathbf{y} , a vector of the outcome variable y_{ij} , given the random effects, follows a binomial distribution such that the property damage type of the i th subject in time j . A GLMM for binary data with logit-link, which is the link function $g(\mu_{ij})$ determining how the mean is related to the independent variables \mathbf{x} , is written in the form:

$$g(\mu_{ij}) = \text{logit}(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i, \quad i = 1, \dots, n, \quad j = 1, \dots, t_i \quad (1)$$

where $\boldsymbol{\beta}$ ($p \times 1$) is a vector of fixed effect parameters; \mathbf{b}_i ($q \times 1$) is a vector of random effects which represent the influence of subject i on its repeated observations, having dimension n ; \mathbf{x}_{ij} ($p \times 1$) is a vector of independent variables associated with the ij th observation; and \mathbf{z}_{ij} ($q \times 1$) is a vector of variables having random effects (Antonio and Beirlant, 2007). GLMM utilizes the logit-link for the analysis of dichotomous data, namely

$$g(\mu_{ij}) = \text{logit}(\mu_{ij}) = \log \left[\frac{\mu_{ij}}{1-\mu_{ij}} \right] \quad (2)$$

where μ_{ij} is the probability of an event on subject i in time j . Here, the conditional expectation equals the conditional probability of a response given the random effects and covariate values, i.e.,

$$\mu_{ij} = E(y_{ij}|\mathbf{b}_i, \mathbf{x}_{ij}) = P(y_{ij} = 1|\mathbf{b}_i, \mathbf{x}_{ij}) \quad (3)$$

(Hedeker, 2005). Assuming that the random effects are mutually independent and identically distributed completes the specification of the GLMM. Furthermore, the correlation between

observations on the same subject occurs since they share the same random effects b_i (Antonio and Beirlant, 2007).

For more information on the theory and application of GLMs, see McCullagh and Nelder (1989), De Jong and Heller (2008), Kaas et al. (2008), Frees (2010), and Ohlsson and Johansson (2010).

2.2. Other Traditional Methods

The random parameter approach has been the most widely used to account for unobserved heterogeneity. Alternative approaches for addressing heterogeneity and panel effects include grouped random parameter (Fountas et al., 2019; Pantangi et al., 2019), correlated random parameter (Balusu et al., 2018; Fountas et al., 2019; Tran et al., 2015), bivariate/multivariate random parameter (Barua et al., 2015, 2016; Dong et al., 2014; Gong et al., 2022; Pantangi et al., 2019), and mixed generalized models (Anarkooli et al., 2017; Balusu et al., 2018; Chen et al., 2018; Eluru et al., 2008). The random parameter model, under the concept of hierarchical modeling, is also the most widely used technique (Bakhshi and Ahmed, 2021; Fountas and Anastasopoulos, 2017; Kim et al., 2017; Lord and Mannering, 2010). Mannering et al. (2016) summarize the methodological approaches accounting for unobserved heterogeneity.

Because the data structure consisting of unbalanced repeated measures and panel data can be problematic to analyze, GLMMs are suitable for this purpose. In the GLMM context, in addition to determining the structure of correlation between observations on the same subject, the random effects also consider heterogeneity among subjects due to unobserved features (Antonio and Beirlant, 2007). Since the mixed-effects logistic regression model is the most popular GLMM, the data are analyzed by means of the logistic GLMM in this study (Hedeker, 2005). In order to model the risk factors having a major impact on the likelihood of DCPD claims and to account for unobserved heterogeneity when addressing these risk factors, a random parameter approach is used in the framework of GLMMs in this study. This model is used to analyze binary data from TPL insurance, specifically DCPD and PD, by assuming the same random parameter mean and variance for all observations. In addition, a GLMM with independent, homoscedastic residual errors is specified for the likelihood of DCPD claims from the Canadian automobile insurance dataset in this study.

2.3. Data Description

Data about only private passenger automobiles are provided from the automobile portfolio of an active insurance company in Canada. The dataset includes insurance information about a total of 1,946 observations for 1,397 policies that have been in the portfolio for ten complete years, each of which consists of the claim experience for several rating factors and a given calendar year. The data do not contain insurance details for the policy year in which no claim was filed.

Table 2. Variables in the dataset

Variable	Definition
Age	Age of policyholder at the time of claim
Territory	Residential area 0 (Urban), 1 (Rural)
Usage	Vehicle usage 0 (Work/Business), 1 (Pleasure)
Time	Accident year $j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$, corresponding to values of 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011 and 2012, respectively
Class	Code of class 0 (Vehicle used for pleasure or having vehicle usage restrictions for commuting to work one way, and driver is 25 years of age or over), 1 (Vehicle used for pleasure and business or not having vehicle usage restrictions for commuting to work one way, and driver is 25 years of age or over), 2 (Vehicle not having vehicle usage restrictions, and driver is under 21 years of age), 3 (Vehicle not having vehicle usage restrictions, and driver is under 25 years of age, but not under 21 years of age)
Driver record	Number of claims-free years for each policy (in the last 6 years) 0 (No claims-free years), 1 (One claims-free year), 2, 3, 4, 5, 6
Claims history	Number of claims the risk has had in the last 6 years before the policy was rated 0 (Number of chargeable claims is zero), 1, 2 (Number of chargeable claims is two or more)
Claims-free years	Number of years since the risk had a claim 0 (Zero year), 1, 2, 3, 4, 5, 6, 7, 8, 9 (Nine or more years)
Experience	Number of years the driver has been licensed 0 (Zero year), 1, 2, 3, 4, 5, 6, 7, 8, 9 (Nine or more years)
Training	Driving education provided to all candidates 0 (Drivers have taken the course in Ontario), 1 (Drivers have taken the course, but maybe a different jurisdictionally specific one), 2 (Drivers have not taken the course)
Gender	0 (Female), 1 (Male)
Outstanding loss	Loss reported to the insurer but is still in the process of settlement
Incurred loss	Amount actually paid in loss during a specified time

The analysis is performed to the company's liability insurance claim experience for 2003–2012. The data comprise outstanding loss ($x^{(12)}$), which only includes zero and positive claim amounts, incurred loss ($x^{(13)}$), which only includes positive claim amounts, and several rating factors for each policy that consist of age ($x^{(1)}$), territory ($x^{(2)}$), usage ($x^{(3)}$), time ($x^{(4)}$), class ($x^{(5)}$), driver record ($x^{(6)}$), claims history ($x^{(7)}$), claims-free years ($x^{(8)}$), experience ($x^{(9)}$), training ($x^{(10)}$), and gender ($x^{(11)}$). Table 2 gives detailed information about the rating factors of the policy.

In the following analysis, territory ($x^{(2)}$), usage ($x^{(3)}$), class ($x^{(5)}$), training ($x^{(10)}$), and gender ($x^{(11)}$) are treated as factor covariates while age ($x^{(1)}$), time ($x^{(4)}$), driver record ($x^{(6)}$), claims history ($x^{(7)}$), claims-free years ($x^{(8)}$), experience ($x^{(9)}$), outstanding loss ($x^{(12)}$), and incurred loss ($x^{(13)}$) are treated as continuous covariates in the model.

Driver characteristics also involve the date of birth of the policyholders, while the claim profiles include information on the type of coverage regarding property damage, such as 0 (PD covered by liability insurance) and 1 (DCPD), policy effective and expiry date, claim identification number, and accident date.

The model is fitted using the claims for the years 2003–2008, and its predictive ability is evaluated using the claims from 2009–2012. The data for 2003–2008 consist of 1,169 observations on 942 policies for 179 brokers, and each observation includes the claim experience at the individual policy level. Of the 1,169 observations, 88 (7.5%) have PD covered by liability insurance and 1,081 (92.5%) have DCPD. These observations are summarized as shown in Table 3.

Table 3. Summary statistics of the data

Variable	Mean	Std.Dev.	Minimum	Maximum
Age	45.05	13.33	18.42	85.10
Time	3.78	1.54	1.00	6.00
Driver record	5.54	1.33	0.00	6.00
Claims history	0.08	0.28	0.00	2.00
Claims-free years	7.69	2.60	0.00	9.00
Experience	8.37	1.78	0.00	9.00
Outstanding loss	620.5	1356.19	0.00	5550.00
Incurred loss	3625.20	4129.49	26.84	43539.90

The analysis herein focuses on estimating the model using the property damages that occurred during each individual year to examine the likelihood of DCPD claims versus PD claims covered by TPL insurance. Table 4 presents the mean of the outstanding and incurred losses used in the forthcoming estimations for each of the six years.

Table 4. Mean of outstanding and incurred loss distribution by years

Year	Outstanding Loss	Incurred Loss
2003	1172.18	2502.82
2004	307.43	3498.63
2005	603.57	3612.98
2006	431.30	3735.09
2007	814.31	3695.84
2008	778.56	3966.12

To optimize the merits of the variables in the model, a transformation is applied to both outstanding and incurred losses. The Yeo-Johnson transformation handles both positive and negative values, whereas the Box-Cox transformation only handles positive values. Because outstanding loss only includes zero and positive claim amounts, the Yeo-Johnson transformation is made for outstanding loss. Incurred loss, on the other hand, only includes positive claim amounts. Therefore, the Box-Cox transformation is applied for incurred loss.

In the insurance portfolio, these observations are handled as separate classes. The frequency table of the classes is given in Table 5.

Table 5. Frequency table of the class

Variable	Group	Number of Observations	Percent (%)
Class	0	1,002	85.71
	1	126	10.78
	2	31	2.65
	3	10	0.86
Total		1,169	100.00

Of the 1,169 observations, 1,002 (85.71%) include the drivers who are 25 years of age or over and use their vehicle for pleasure or have vehicle usage restrictions for commuting to work one way, 126 (10.78%) consist of drivers who are 25 years of age or over and use their vehicle for pleasure and business or not have vehicle usage restrictions for commuting to

work one way, 31 (2.65%) contain drivers who are under 21 years of age and not have vehicle usage restrictions, and 10 (0.86%) comprise drivers who are under 25 years of age, but not under 21 years of age and not have vehicle usage restrictions. Because the observations in the data are not distributed in a balanced way among the categories of class ($x^{(5)}$) from the factor covariates, the weights on class ($x^{(5)}$), which are the number of trials, are used to generate each success proportion. As a result, since the dataset is unbalanced, using weights allows us to consider the relative importance of various possible target values and to better fit the model.

Among the rating variables in the dataset, claims-free years ($x^{(8)}$) and experience ($x^{(9)}$) are highly correlated. The models are built considering the correlation between these variables and then compared to one another to determine the best model. In the following analysis, the best fitted model is presented. Table 6 shows the correlation between the independent variables in this fitted model.

Table 6. Correlation matrix of independent variables in the model

	$x^{(3)}$	$x^{(4)}$	$x^{(10)}$	$x^{(11)}$	$x^{(12)}$	$x^{(13)}$
$x^{(3)}$	1.000	-0.012 ^c	0.007 ^a	0.033 ^b	0.016 ^c	-0.027 ^c
$x^{(4)}$	-0.012 ^c	1.000	0.059 ^{c*}	-0.018 ^c	0.064 ^d	0.106 ^d
$x^{(10)}$	0.007 ^a	0.059 ^{c*}	1.000	0.037 ^a	-0.058 ^{c*}	0.030 ^{c*}
$x^{(11)}$	0.033 ^b	-0.018 ^c	0.037 ^a	1.000	0.016 ^c	-0.014 ^c
$x^{(12)}$	0.016 ^c	0.064 ^d	-0.058 ^{c*}	0.016 ^c	1.000	0.051 ^d
$x^{(13)}$	-0.027 ^c	0.106 ^d	0.030 ^{c*}	-0.014 ^c	0.051 ^d	1.000

* The greatest correlation between the discrete or continuous variable and all possible pairs of levels of the nominal variable

^a Goodman and Kruskal's Lambda

^b Phi coefficient

^c Point-biserial correlation coefficient

^d Spearman correlation coefficient

(Khamis, 2008)

As a result, the model presented below does not exhibit any multicollinearity issue. Within this model, 650 (55.6%) of the 1,169 observations use the vehicle for work and business, while 519 (44.4%) use it for pleasure. 20 (1.7%) of the observations include the drivers who have taken the course in Ontario, whereas 1,130 (96.7%) consist of those who have taken it in another jurisdiction. 19 (1.6%) of the observations also comprise the drivers

who have not taken the course. Female drivers make up 524 (44.8%) of the observations, while male drivers add up to 645 (55.2%).

2.4. Fitted Model

A random intercept effect model is a type of GLMM that allows for the inclusion of individual-specific random effects in addition to more general risk factors. This model can help to account for unobserved heterogeneity in the data, which can have a significant impact on the likelihood of claims. By incorporating random intercepts into the model, the effect of unobserved heterogeneity can be accounted for, resulting in more accurate estimates of risk and more appropriate insurance premiums.

This study aims to determine how the most significant risk factors affect DCPD claims under TPL insurance. Two categories are addressed to model the property damage coverage type following a traffic accident: DCPD or PD covered by liability insurance. The GLMM described in Section 2.1 is fitted using the `glmer` function in R with logit-link.

Using GLMM analysis for the subject-specific random intercept effect model, the best-fitting random intercept effect model is specified as follows:

$$g(\mu_{ijk}) = \beta_0 + \beta_1 x_{ijk}^{(3)} + \beta_2 x_{ijk}^{(4)} + \beta_3 x_{ijk}^{(10)} + \beta_4 x_{ijk}^{(11)} + \beta_5 x_{ijk}^{(12)} + \beta_6 x_{ijk}^{(13)} + b_{0k},$$
$$i = 1, \dots, n, \quad j = 1, \dots, t_i, \quad k = 1, \dots, m \quad (4)$$

where n is the total number of different policies; m is the total number of different brokers; t_i is the number of repeated observations in policy i . t_i is the same for all policies in balanced panel data, but conversely, the panel data structure here is unbalanced. In addition, μ_{ijk} is the probability of a claim on policy i ($i = 1, \dots, 942$) at time j ($j = 1, \dots, 6$) for broker k ($k = 1, \dots, 179$).

In the fixed-effects part of the model, the parameters β_0 , β_1 , and β_2 define an overall intercept, the change in the expected log odds of DCPD claims for vehicle usage, and the change caused by a one-year change in time, for a given the random intercept, respectively. The change in the expected log odds of DCPD claims for driving education and gender are expressed in parameters β_3 and β_4 , for a given the random intercept, respectively. Additionally, β_5 and β_6 describe how the expected log odds of DCPD claims have changed due to a unit increase in both outstanding loss and incurred loss for a given the random intercept.

In the random-effects part of the model, the term b_{0k} in Equation (4) denotes a broker-specific random intercept. The random intercept b_{0k} is a subject-specific deviation from the fixed intercept β_0 . The results of the panel data generalized linear binary logit mixed model are summarized in Table 7.

Table 7. Generalized linear binary logit mixed model estimation results

Parameter	Variable	Estimated Coefficients	Std. Error	z-value	Pr(> z)	Exp(β)
β_0	Intercept	- 4.691	0.694	- 6.761	< 0.001 ***	0.009
β_1	Usage1	- 1.083	0.268	- 4.038	< 0.001 ***	0.339
β_2	Time	0.157	0.085	1.844	0.065 `	1.170
β_3	Training1	1.215	0.464	2.621	0.009 **	3.372
	Training2	3.411	1.235	2.763	0.006 **	30.279
β_4	Gender1	- 0.463	0.275	- 1.680	0.093 `	0.629
β_5	Outstanding loss	0.187	0.063	2.948	0.003 **	1.205
β_6	Incurred loss	0.946	0.075	12.665	< 0.001 ***	2.576
b_{0k}	Random parameter					
	Std. dev. of broker	0.632				
	- 2 Log-likelihood	462.0				
	AIC	480.1				
	BIC	525.7				

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

2.5. Performance Metrics

In this study, the GLMM approach is applied to unbalanced panel data to determine which factors have a significant impact on the likelihood of DCPD claims that policyholders will make next year. To inform model selection, the Akaike information criterion (*AIC*) and likelihood ratio test (*LRT*) are used. If the number of observations (N) is large enough, $k < (N/40)$, *AIC* is defined as

$$AIC = -2 \ln(\hat{L}) + 2k \quad (5)$$

where k represents the number of estimated parameters in the fitted model and $\ln(\hat{L})$ is the maximum log-likelihood value (Portet, 2020). In this study, Equation (5) is used to calculate the AIC value since $k = 9$ is smaller than $N/40 = 29.225$ for $N = 1169$, and the model with a lower AIC value is preferred.

The reference model, which includes weights (the number of trials used to generate each success proportion) in the different classes of insureds, is compared to the nested model, which is reduced to a model without weights, using likelihood ratio tests to determine which is statistically preferable. The likelihood ratio test is shown in Equation (6).

$$LRT = 2\{\log\text{Lik}(\text{reference}) - \log\text{Lik}(\text{nested})\} \quad (6)$$

where $\log\text{Lik}(\text{reference})$ and $\log\text{Lik}(\text{nested})$ are the log-likelihood of the generalized linear mixed model with weights (under the alternative hypothesis) and the generalized linear mixed model without weights (under the null hypothesis) for the same dataset, respectively. With degrees of freedom equal to the difference in the number of parameters between the two models, the test statistic is a chi-square distribution (Pai and Walch, 2020). The chi-square value of the test is 64.058 with one degree of freedom. The corresponding p -value is $Pr(\chi_1^2 > 64.058)$. From the chi-square table, we can conclude that $Pr(\chi_1^2 > 7.88) = 0.005$ and hence the p -value is significantly lower than 0.0025. The model under the alternative hypothesis is chosen since the p -value is much less than 0.05. In other words, the random-effects model with weights is preferred because it significantly differs from the random-effects model without weights.

The evaluation metrics used in this analysis include measures of sensitivity (recall) (SN), specificity (SP), precision (P), accuracy (ACC), balanced accuracy (BA), $F1$ score, and area under the curve (AUC) to assess the performance of each model and to determine which model is most effective for predicting the likelihood of DCPD claims. These measures are defined based on a confusion matrix, as shown in Table 8 (Hossin and Sulaiman, 2015).

Table 8. Confusion matrix for the binary classification

	Prediction	
Actual	DCPD	PD
DCPD	TP	FN
PD	FP	TN

In this confusion matrix, TP (true positive) and TN (true negative) denote the number of positive (classifying the claim as DCPD) and negative (classifying the claim as PD covered by liability insurance) claims that are correctly classified, respectively. Additionally, FP (false positive) and FN (false negative) represent the number of positive and negative claims that are incorrectly classified, respectively. In other words, TP and TN indicate DCPD claims correctly identified as DCPD and PD claims correctly identified as PD, respectively. FP stands for PD claims incorrectly identified as DCPD, whereas FN implies DCPD claims incorrectly identified as PD. The performance evaluation metrics used in this analysis are generated as shown in Equations (7) – (12).

$$SN = \frac{TP}{TP+FN} \quad (7)$$

$$SP = \frac{TN}{TN+FP} \quad (8)$$

$$P = \frac{TP}{TP+FP} \quad (9)$$

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \quad (10)$$

$$BA = \frac{SN+SP}{2} \quad (11)$$

$$F1 \text{ score} = \frac{2TP}{2TP+FP+FN} \quad (12)$$

Precision and recall are employed as the evaluation metrics in this study since the developed model aims to predict 1 as accurately as feasible and to identify as many actual 1 as possible. In classification issues, accuracy is one of the most frequently used evaluation metrics. It is helpful when the target class is well-balanced but not a suitable option when the classes are unbalanced. This study assesses the target classes that are to be applied to a severely unbalanced dataset in which positives greatly outnumber negatives. Balanced accuracy is chosen as a performance measure in this analysis because it is a better metric when dealing with imbalanced data, and it also accounts for both positive and negative classes and avoids data imbalances that could be misleading. Additionally, the $F1$ score is a commonly employed evaluation metric to measure the performance of binary classification and outperforms accuracy in enhancing the target classes for binary classification problems. Therefore, it is used in this analysis as a performance measure rather than an accuracy measure.

Another evaluation metric is the Receiver Operating Characteristic (ROC) curve, which assesses the predictive performance of the fitted model. The ROC curve is a plot of the true positive rate (SN) versus the false positive rate ($1-SP$), which shows how the number of correctly classified positive instances varies with the number of incorrectly classified negatives when evaluating binary decision problems. The ROC curve captures the trade-off between these performance measure parameters for different possible thresholds. The resulting score known as AUC is the area under the ROC curve and illustrates the model's ability to accurately predict classes. A higher score indicates a higher probability of making correct predictions and can be viewed as a measure of accuracy (Davis and Goadrich, 2006).

A Precision-Recall (PR) curve, on the other hand, evaluates the fraction of true positives among positive predictions. By offering valuable insights into the effectiveness of the classification model in capturing and correctly labeling minority class instances, the PR curve can provide an accurate prediction of future classification performance. The PR curve outperforms the ROC curve in terms of both information and power when dealing with binary classes on unbalanced datasets (Saito and Rehmsmeier, 2015). Due to class imbalance in this analysis, presenting results by considering only the ROC curve could be misleading about the reliability of classification performance. In this study, as well as the ROC curve, the PR curve is also considered to evaluate the classification performance because the PR curve can explicitly reveal claim differences in imbalanced cases. The resulting score known as $PR AUC$ is the area under the PR curve and emphasizes the performance of the model for predicting the positive class. A high $PR AUC$ means that the model performs better in predicting the positive class. These performance assessment measures are acquired as presented in Table 9.

Table 9. Performance evaluation metrics

Sensitivity	Specificity	Precision	Accuracy
0.906	0.647	0.964	0.883
Balanced Accuracy	F1 Score	AUC	PR AUC
0.776	0.934	0.776	0.953

The fitted model's $F1$ score of 0.934, which is regarded as a very good value, indicates that it can both capture positive classes and accurately predict the classes it does capture. Regarding the balanced accuracy, it has a value of 0.776, indicating that the fitted model performs well at predicting whether policyholders will make DCPD claims. Due to the imbalanced classes in this analysis, the balanced accuracy gives us a more realistic picture of how well the model classifies both groups correctly. To evaluate the predictive performance of the fitted model, the ROC and PR curves are plotted as shown in Figure 1.

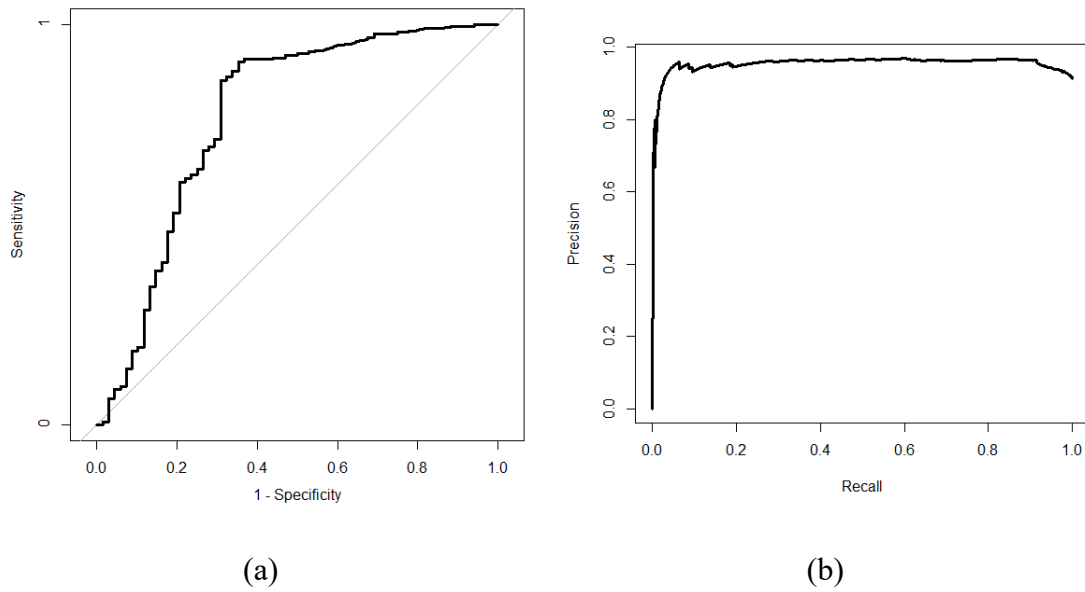


Figure 1. Predictive performance of the fitted model: (a) the ROC curve and (b) the PR curve

For the fitted model using different probability thresholds, the ROC curve highlights the trade-off between the true positive rate and the false positive rate. The fitted model provides a good fit to the data according to the computed AUC of 0.776. For the fitted model employing different probability thresholds, the PR curve highlights the trade-off between the true positive rate and the positive predictive value. Compared to the ROC curve, the PR curve is preferable to the ROC curve for imbalanced datasets. Due to the class imbalance in this analysis, $PR AUC$, calculated as 0.953, describes that the fitted model performed very well in predicting the positive class.

3. RESULTS

This paper describes a generalized linear binary logit mixed model considering the imbalance between the classes of policyholders using automobile insurance data. This model

assesses the impact of various risk factors on DCPD claims in private passenger vehicle accidents. The risk factors having a significant impact on the likelihood of DCPD claims are the independent variables named “usage”, “time”, “training”, “gender”, “outstanding loss”, and “incurred loss” estimated in unbalanced longitudinal data.

Gender and time are included in the model even if they are thought to be ineffective. However, these two variables are significant at the 0.10 level, as shown in Table 7. The results of these variables indicate that female drivers are 1.59 times more likely to make a DCPD claim than male drivers, and that the risk of a DCPD claim occurring is 1.17 times higher when time increases by 1 year.

As for other significant variables, usage, training, outstanding loss, and incurred loss have a significant effect on the likelihood of DCPD claims. For policyholders who use their vehicles for work or business, the risk of making a DCPD claim is 2.95 times greater than for those who use them for pleasure. Since drivers who commute to work or use the vehicle for business are far more likely to be in traffic than those who drive for pleasure, this result is meaningful and the vehicle usage has a quite significant effect on DCPD claims.

Driver training is of vital importance in preventing traffic accidents. Even if most drivers in Ontario have taken courses, some have not taken any training. Given that Ontario is one of the provinces with the highest number of immigrants, many drivers have taken driver training in various jurisdictions, whereas some have taken it in Ontario. According to the results of the training variable in the model, policyholders who have taken the driver training in a separate jurisdiction are 3.37 times more likely to make a DCPD claim than those who have taken it in Ontario; whereas policyholders who have not taken courses are 30.28 times more likely to make a DCPD claim than those with driver training in Ontario. These results indicate that drivers who have taken courses in a different jurisdiction or have not taken any training pose a risk in traffic and support the importance of driver training in preventing traffic accidents.

For insurers to manage their claims liabilities, determine appropriate premium rates, and evaluate their overall financial circumstances, outstanding loss and incurred loss are crucial. The claim reported to the insurance company but has not yet been paid is known as an outstanding loss. This claim is an estimate of the insurer's future financial obligations. Incurred loss, also called paid loss, is the actual loss that the insurance company has paid or became obligated to pay during a specific period. The results of these two variables in the model demonstrate that the risk of a DCPD claim occurring is 1.21 times higher when the outstand-

ing loss increases by \$1 and that the risk of a DCPD claim occurring is 2.58 times greater when the incurred loss increases by \$1.

DCPD claims are one of the most common types of damage insurance companies incur. DCPD coverage under TPL insurance provides compensation to policyholders for damages by the policyholders' own insurer in cases where the accident was caused by another driver and was not their own fault. It can indeed be advantageous to consider these rating factors which significantly affect the likelihood of DCPD claims for evaluating insurance premiums and enhancing the financial stability of an insurance company. By incorporating these factors into the premium evaluation process, insurers can more accurately estimate the risk associated with each policyholder and price premiums accordingly.

It is recommended that the above rating factors having a significant impact on the likelihood of DCPD claims be considered in the premium evaluation since it is believed to help the financial stability of the insurance company. The financial stability of the company could potentially be affected if the insurance company pays more compensation than it collects in premiums.

4. CONCLUSIONS

The purpose of this study is to develop a statistical model that identifies the impact of the most important risk factors on DCPD claims under TPL insurance in private passenger vehicle accidents in Ontario, Canada. GLMMs are approaches that are constantly used to model actuarial data and provide an advantage in the analysis of unbalanced panel data. This approach eliminates the extra complexity resulting from the internal correlation of each policy. Therefore, the developed model in this study analyzes the likelihood of DCPD claims in the context of a generalized linear binary logit mixed model by dealing with unbalanced panel data, and also, the imbalance between the classes of insureds is considered in this model.

As a type of data application, the data in this study include many factors associated with the driver and claim characteristics found critical to the likelihood of DCPD claims. The estimation results from the model demonstrate that the broker, which is a time-varying factor, has a significant influence on the likelihood of DCPD claims as a random parameter. In addition, rating factors such as usage patterns, driver training, outstanding loss and incurred loss have been found to correlate with the likelihood of DCPD claims as fixed effects. Observing the effects of these risk factors under the weights in the different classes of policyholders high-

lights the importance of developing class-specific risk assessment models. Moreover, by considering the performance evaluation metrics in detail, this study ensures a comprehensive assessment that accounts for the potential challenges posed by imbalanced datasets and provides a more reliable interpretation of the results.

Taking these factors into account during premium evaluation helps insurers maintain financial stability by ensuring that premiums are adequately priced based on the associated risks. This, in turn, helps the company avoid potential financial instability caused by underpricing policies or facing a higher volume of claims than anticipated.

Ultimately, incorporating rating factors that have a significant impact on the likelihood of DCPD claims in premium evaluation promotes a fair and sustainable insurance pricing strategy, benefiting both the insurance company and its policyholders.

ACKNOWLEDGMENTS

In addition to the editor and the anonymous referees for their constructive suggestions, the authors would like to thank Dr. Jeffrey S. Pai from the University of Manitoba and the Canadian insurance company for providing the data used for conducting the analysis. This work was partially supported by the Scientific and Technological Research Council of Turkey (TUBITAK 2214-A).

ETHICAL DECLARATION

In the writing process of the study titled “Quantifying the impact of risk factors on direct compensation property damage in Canadian automobile insurance”, there were followed the scientific, ethical and the citation rules; was not made any falsification on the collected data and this study was not sent to any other academic media for evaluation.

REFERENCES

Anarkooli, A. J., Hosseinpour, M. and Kardar, A. (2017), Investigation of factors affecting the injury severity of single-vehicle rollover crashes: A random-effects generalized ordered probit model, *Accident Analysis and Prevention*, 106, 399-410.

- Antonio, K. and Beirlant, J. (2007), Actuarial statistics with generalized linear mixed models, *Insurance: Mathematics and Economics*, 40(1), 58-76.
- Antonio, K. and Valdez, E. A. (2012), Statistical concepts of a priori and a posteriori risk classification in insurance, *AStA Advances in Statistical Analysis*, 96, 187-224.
- Bakhshi, A. K. and Ahmed, M. M. (2021), Practical advantage of crossed random intercepts under Bayesian hierarchical modeling to tackle unobserved heterogeneity in clustering critical versus non-critical crashes, *Accident Analysis and Prevention*, 149, 105855.
- Balusu, S. K., Pinjari, A. R., Mannering, F. L. and Eluru, N. (2018), Non-decreasing threshold variances in mixed generalized ordered response models: A negative correlations approach to variance reduction, *Analytic Methods in Accident Research*, 20, 46-67.
- Barua, S., El-Basyouny, K. and Islam, M. T. (2015), Effects of spatial correlation in random parameters collision count-data models, *Analytic Methods in Accident Research*, 5, 28-42.
- Barua, S., El-Basyouny, K. and Islam, M. T. (2016), Multivariate random parameters collision count data models with spatial heterogeneity, *Analytic Methods in Accident Research*, 9, 1-15.
- Chen, F., Chen, S. and Ma, X. (2018), Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data, *Journal of Safety Research*, 65, 153-159.
- Davis, J. and Goadrich, M. (2006), The relationship between Precision-Recall and ROC curves, In: *Proceedings of the 23rd International Conference on Machine Learning – ICML '06*, 233-240.
- De Jong, P. and Heller, G. Z. (2008), *Generalized Linear Models for Insurance Data*, In: International Series on Actuarial Science, Cambridge University Press.
- Dong, C., Clarke, D. B., Yan, X., Khattak, A. and Huang, B. (2014), Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections, *Accident Analysis and Prevention*, 70, 320-329.

- Eluru, N., Bhat, C. R. and Hensher, D. A. (2008), A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes, *Accident Analysis and Prevention*, 40(3), 1033-1054.
- Embrechts, P. and Wüthrich, M. V. (2022), Recent challenges in actuarial science, *Annual Review of Statistics and Its Application*, 9, 119-140.
- Frees, E. W. (2010), *Regression Modeling with Actuarial and Financial Applications*, In: International Series on Actuarial Science, Cambridge University Press.
- Fountas, G. and Anastopoulos, P. C. (2017), A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities, *Analytic Methods in Accident Research*, 15, 1-16.
- Fountas, G., Pantangi, S. S., Hulme, K. F. and Anastopoulos, P. C. (2019), The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: A correlated grouped random parameters bivariate probit approach, *Analytic Methods in Accident Research*, 22, 100091.
- Garrido, J., Genest, C. and Schulz, J. (2016), Generalized linear models for dependent frequency and severity of insurance claims, *Insurance: Mathematics and Economics*, 70, 205-215.
- Gong, H., Fu, T., Sun, Y., Guo, Z., Cong, L., Hu, W. and Ling, Z. (2022), Two-vehicle driver-injury severity: A multivariate random parameters logit approach, *Analytic Methods in Accident Research*, 33, 100190.
- Haberman, S. and Renshaw, A. E. (1996), Generalized linear models and actuarial science, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 45(4), 407-436.
- Hedeker, D. (2005), Generalized linear mixed models, In: B. Everitt, D. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science*, John Wiley & Sons, New York, 729-738.
- Hossin, M. and Sulaiman, M. N. (2015), A review on evaluation metrics for data classification evaluations, *International Journal of Data Mining and Knowledge Management Process*, 5(2), 1-11.
- Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008), *Modern Actuarial Risk Theory: Using R*, Second Edition, Springer Berlin, Heidelberg.

- Khamis, H. (2008), Measures of association: How to choose?, *Journal of Diagnostic Medical Sonography*, 24(3), 155-162.
- Kim, M., Kho, S. Y. and Kim, D. K. (2017), Hierarchical ordered model for injury severity of pedestrian crashes in South Korea, *Journal of Safety Research*, 61, 33-40.
- Lord, D. and Mannering, F. (2010), The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives, *Transportation Research Part A: Policy and Practice*, 44(5), 291-305.
- Mannering, F. L., Shankar, V. and Bhat, C. R. (2016), Unobserved heterogeneity and the statistical analysis of highway accident data, *Analytic Methods in Accident Research*, 11, 1-16.
- McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, In: Monographs on Statistics and Applied Probability 37, Second Edition, Chapman and Hall, London, New York.
- Miao, G. M. (2018), Application of hierarchical model in non-life insurance actuarial science, *Modern Economy*, 9(3), 393-399.
- Nelder, J. A. and Wedderburn, R. W. M. (1972), Generalized linear models, *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370-384.
- Ohlsson, E. and Johansson, B. (2010), *Non-life Insurance Pricing with Generalized Linear Models*, In: EAA Series Textbook, Springer Berlin, Heidelberg.
- Pai, J. S. and Walch, A. H. (2020), *ACTEX Study Manual for Exam MAS-II*, ACTEX Learning/SRBooks, Inc., Greenland, NH.
- Pantangi, S. S., Fountas, G., Sarwar, M. T., Anastasopoulos, P. C., Blatt, A., Majka, K., Pierowicz, J. and Mohan, S. B. (2019), A preliminary investigation of the effectiveness of high visibility enforcement programs using naturalistic driving study data: A grouped random parameters approach, *Analytic Methods in Accident Research*, 21, 1-12.
- Portet, S. (2020), A primer on model selection using the Akaike Information Criterion, *Infectious Disease Modelling*, 5, 111-128.

- Saito, T. and Rehmsmeier, M. (2015), The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PloS One*, 10(3), e0118432.
- Tran, V., Liu, D., Pradhan, A. K., Li, K., Bingham, C. R., Simons-Morton, B. G. and Albert, P. S. (2015), Assessing risk-taking in a driving simulator study: Modeling longitudinal semi-continuous driving data using a two-part regression model with correlated random effects, *Analytic Methods in Accident Research*, 5, 17-27.
- Yau, K., Yip, K. and Yuen, H. K. (2003), Modelling repeated insurance claim frequency data using the generalized linear mixed model, *Journal of Applied Statistics*, 30(8), 857-865.