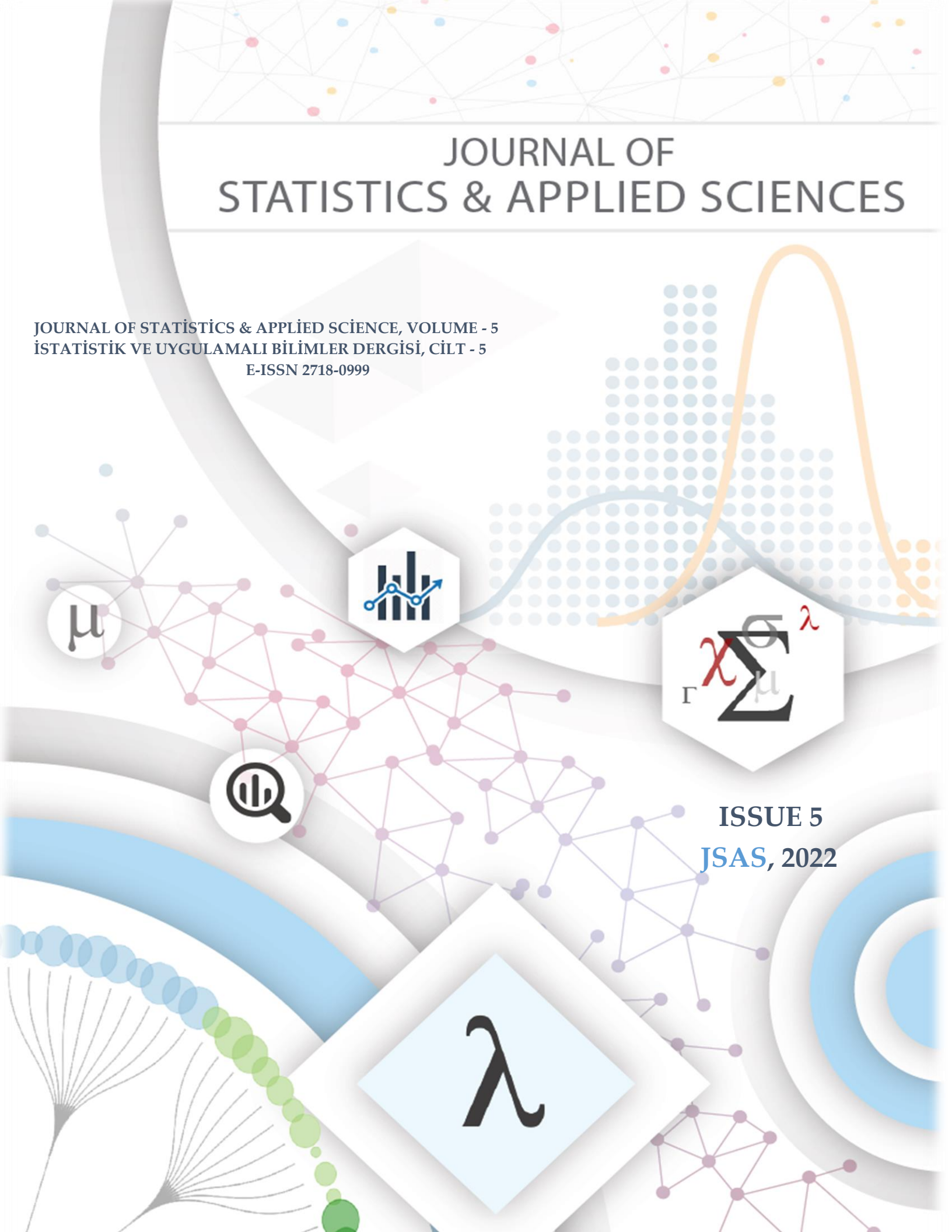


# JOURNAL OF STATISTICS & APPLIED SCIENCES

JOURNAL OF STATISTICS & APPLIED SCIENCE, VOLUME - 5  
İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ, CİLT - 5  
E-ISSN 2718-0999

ISSUE 5  
JSAS, 2022





Journal of Statistics & Applied Science, Issue -5  
ISSN 2718-0999

# -JOURNAL OF STATISTICS & APPLIED SCIENCE- -İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ-

Haziran 2022

Issue: 5

Sayı: 5

İmtiyaz Sahibi / Owner

Abdulkadir KESKİN

Baş Editör / Editor in- Chief

Dr.Abdulkadir ATALAN

Yayın Türü: 6 Aylık, Uluslararası, Hakemli

e-ISSN: 2718-0999

Dizinleme Bilgileri/ Abstracting and Indexing Services


İletişim / Contact

E-posta: [jsas.journal@gmail.com](mailto:jsas.journal@gmail.com)

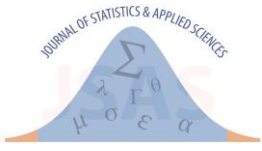
<https://dergipark.org.tr/tr/pub/jsas>

## BİLİM VE DANIŞMA KURULU

Prof. Dr. Serpil AKTAŞ ALTUNAY,	HACETTEPE ÜNİVERSİTESİ
Prof. Dr. Filiz KARAMAN,	YILDIZ TEKNİK ÜNİVERSİTESİ
Prof. Dr. Musa ŞAHİN,	YALOVA ÜNİVERSİTESİ
Prof. Dr. Ersoy ÖZ,	YILDIZ TEKNİK ÜNİVERSİTESİ
Prof. Dr. Dursun YENER,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Doç. Dr. Atif Ahmet EVREN,	YILDIZ TEKNİK ÜNİVERSİTESİ
Doç. Dr. Serpil KILIÇ DEPREN,	YILDIZ TEKNİK ÜNİVERSİTESİ
Doç. Dr. Abdulkadir ATALAN,	GAZİANTEP İSLAM BİLİM VE TEKNOLOJİ ÜNİVERSİTESİ
Doç. Dr. Akansel YALÇINKAYA,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Doç. Dr. Ömer BİLEN,	BURSA TEKNİK ÜNİVERSİTESİ
Doç. Dr. Serhat YÜKSEL,	İSTANBUL MEDİPOL ÜNİVERSİTESİ
Doç. Dr. Muhammed Fevzi ESEN,	SAĞLIK BİLİMLERİ ÜNİVERSİTESİ
Doç. Dr. Mertcan TAŞÇIOĞLU,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Doç. Dr. Yavuz ÖZDEMİR,	İSTANBUL SAĞLIK VE TEKNOLOJİ ÜNİVERSİTESİ
Dr. Kemal Gökhan NALBANT,	BEYKENT ÜNİVERSİTESİ
Dr. Şahika ÖZDEMİR,	İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
Dr. Mustafa DEMİRBİLEK,	GAZİANTEP İSLAM BİLİM ve TEKNOLOJİ ÜNİVERSİTESİ
Dr. Sevim ÖZULUKALE DEMİRBİLEK,	YOZGAT BOZOK ÜNİVERSİTESİ
Dr. Adilson SILVA,	CAPE VERDE ÜNİVERSİTESİ
Dr. Elif TUNA,	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Fatma Rabia AKTAŞ ŞENKARDEŞLER	BEYKOZ ÜNİVERSİTESİ
Dr. Erhan ÇENE,	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Tutku SEÇKİN ÇELİK,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Burak LEBLEBİCİOĞLU,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Hasan ŞAHİN,	BURSA TEKNİK ÜNİVERSİTESİ
Dr. Coşkun PARİM,	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. İrfan ERSİN,	İSTANBUL MEDİPOL ÜNİVERSİTESİ
Dr. Enes FİLİZ,	BALIKESİR ÜNİVERSİTESİ
Dr. Hasan Aykut KARABOĞA,	AMASYA ÜNİVERSİTESİ
Dr. Hasan, DİLBAŞ,	VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
Dr. Mohamed AHMED,	AL MADİNA HİGHER INSTITUTE FOR MANAGEMENT
Dr. Sinan ATAER,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Yavuz Selim HATİPOĞLU	BAYBURT ÜNİVERSİTESİ
Dr. Merve DURMUŞ,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Müge YETKİN ATAER,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Gözde NALBANT EFE,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. İrem BATIBAY TUNAYDIN,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Mustafa YAPAR,	RECEP TAYİP ERDOĞAN ÜNİVERSİTESİ
Dr. Burcu KUZUCU YAPAR,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Tarkan TALAN,	GAZİANTEP İSLAM BİLİM ve TEKNOLOJİ ÜNİVERSİTESİ
Dr. Çağatay TEKE,	BAYBURT ÜNİVERSİTESİ
Dr. Bayram DÜNDAR,	BARTIN ÜNİVERSİTESİ

## EDİTÖR KURULU

Abdulkadir ATALAN	Baş Editör	GAZİANTEP İSLAM BİLİM ve TEKNOLOJİ ÜNİVERSİTESİ
Abdulkadir KESKİN	Yardımcı Editör	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Mehmet Şamil GÜNEŞ	Yardımcı Editör	YILDIZ TEKNİK ÜNİVERSİTESİ
Enes FİLİZ	Uygulamalı İstatistik Alan Editörü	BALIKESİR ÜNİVERSİTESİ
Ahmet Atif EVREN	Teorik İstatistik Alan Editörü	YILDIZ TEKNİK ÜNİVERSİTESİ
Gökçe GÖK	Sosyal Bilimler Alan Editörü	ATATÜRK ÜNİVERSİTESİ
Burak LEBLEBİCİOĞLU	Sosyal Bilimler Alan Editörü	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Yasemin Ayaz ATALAN	Mühendislik Alan Editörü	YOZGAT BOZOK ÜNİVERSİTESİ
Bayram DÜNDAR	Mühendislik Alan Editörü	BARTIN ÜNİVERSİTESİ
Çağatay TEKE	Mühendislik Alan Editörü	BAYBURT ÜNİVERSİTESİ
Tarkan TALAN	Eğitim Bilimleri Alan Editörü	GAZİANTEP İSLAM BİLİM ve TEKNOLOJİ ÜNİVERSİTESİ
Robert RODGERS	Sağlık Bilimleri Alan Editörü	ALLEGHENY HEALTH NETWORK
Berat KARA	Yayın Editörü	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Abdurrahman KESKİN	Yayın Editörü	İSTANBUL ÜNİVERSİTESİ
Tutku SEÇKİN ÇELİK	Yabancı Dil Editörü	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Umut ÜNAL	Yabancı Dil Editörü	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Recep Uğurcan ŞAHİN	Mizanpaj Editörü	YALOVA ÜNİVERSİTESİ



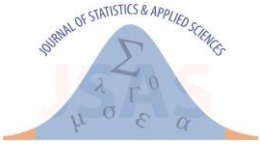
## İÇİNDEKİLER

### ARAŞTIRMA MAKALELERİ (RESEARCH ARTICLES)

<b>Genelleştirilmiş Doğrusal Modellerin İki Değişkenli Gauss Kopula'ya Dahil Edilmesi ve Bir Uygulama</b>	<b>1-9</b>
The Incorporation of Generalized Linear Models into Bivariate Gaussian Copula and An Application	
<b>Övgücan Karadağ ERDEMİR, Meral SUCU</b>	
<b>IT2 Bulanık ANP Metodolojisi Kullanılarak Beton Tesis Yeri Seçimi</b>	<b>10-16</b>
Selecting an Alternative Concrete Batching Plant Location Using IT2 Fuzzy ANP Methodology	
<b>Kemal Gökhan NALBANT, Şahika ÖZDEMİR, Yavuz ÖZDEMİR</b>	
<b>Medikal Sentetik Veri Üretimiyle Veri Dengelemesi</b>	<b>17-27</b>
Data Balancing with Synthetic Medical Data Generation	
<b>Ahmet DEVECİ, M. Fevzi ESEN</b>	
<b>Yeşil Hastanelerin Gelişimi İçin Stratejik Unsurların Belirlenmesi: Enerji Bazlı Faktörlere Yönelik DEMATEL Yöntemi ile Bir Analiz</b>	<b>28-37</b>
Identifying Strategic Elements for the Development of Green Hospitals: An Analysis of Energy Based Factors with the DEMATEL Method	
<b>Serhat YÜKSEL, Hasan DİNÇER, Büşra ÇELEBİ</b>	

### KİTAP İNCELEMELERİ (BOOK REVIEWS)

<b>SPSS ile İstatistik Rehberi Doç. Dr. İbrahim Demir İstanbul: Efe Akademi Yayınları, 2020, 1. Baskı,</b>	<b>38-40</b>
Statistics Guide with SPSS Assoc. Prof. Dr. İbrahim Demir İstanbul: Efe Akademi Publications, 2020, 1st Edition,	
<b>Hasan Aykut KARABOĞA</b>	



## ÖNSÖZ

Bilimsel ve akademik çalışmalar, geçmişten geleceğe insanlığa büyük katkı sağladığından gerekli önemi hak etmektedirler. İnsanlık ihtiyacının yanı sıra doğanın yapısını incelemek ve tartışmak için akademik branşlar ortaya çıkmıştır. Hemen hemen tüm akademik branşların ortak paydası olan istatistiksel uygulamalar ve teknikler, araştırmacılar tarafından büyük ölçüde akademik çalışmalarında kullanılmaktadırlar. **İstatistik ve Uygulamalı Bilimler Dergisi**, istatistik uygulamaları içeren kaliteli, literatüre katkı sağlayan ve okurları bilgilendiren akademik çalışmalara ev sahipliği yapmaktadır. **İstatistik ve Uygulamalı Bilimler** dergisi olarak, açık erişim imkanı sunan, yayın hayatının üçüncü yılının birinci sayısını (toplamda beşinci sayısını) bilim insanlarına, araştırmacılara, öğrencilere ve okurlara sunmanın mutluluğunu yaşamaktadır. Bu sayıda dört araştırma makalesi yanı sıra bir kitap inceleme/eleştiri makalesi yer almaktadır. Özellikle istatistik uygulamaları ile alakalı bilimsel çalışmalar yapan araştırmacılar için bu sayıyı oluşturan makaleler iyi bir referans olacaktır. Bu sebeple akademisyenlerin, araştırmacıların ve öğrencilerin çok büyük bir oranda fayda göreceği bu sayı ile kendi çalışmaları için farklı istatistiksel yaklaşım fikirleri oluşması bakımından literatüre önemli katkı sağlayacaklardır. Dergimizin bu sayısı ile okurlarımızın sunduğumuz bilimsel makalelerden istifade etmesini temenni ederim.

Dr. Abdulkadir Atalan



## PREFACE

Scientific and academic studies deserve the necessary importance as they make a great contribution to humanity from the past to the future. Academic branches have emerged to examine and discuss the structure of nature as well as human needs. Statistical applications and techniques, the common denominator of almost all academic branches, are primarily used by researchers in their academic studies. **Journal of Statistics and Applied Sciences** hosts academic studies that include statistical applications, contribute to the literature, and inform the readers. As the **Journal of Statistics and Applied Sciences**, it is happy to present the first issue (fifth issue in total) of the third year of its publication life, which offers open access, to scientists, researchers, students, and readers. This issue includes four research articles as well as a book review/criticism article. The articles that make up this issue will be a good reference, especially for researchers doing scientific studies on statistical applications. For this reason, with this issue of our journal, where academics, researchers, and students will benefit greatly, they will significantly contribute to the literature in terms of forming different statistical approach ideas for their own studies. I hope that our readers will benefit from the scientific articles we present.

Dr. Abdulkadir Atalan

Research Article

Received: date:11.21.2021  
Accepted: date:15.06.2022  
Published: date:30.06.2022

# The Incorporation of Generalized Linear Models into Bivariate Gaussian Copula and An Application

Övgücan Karadağ Erdemir <sup>1\*</sup> Meral Sucu<sup>2</sup>

<sup>1</sup>Hacettepe University, Department of Actuarial Science, Ankara, Turkey; ovgucan@hacettepe.edu.tr

<sup>2</sup> Hacettepe University, Department of Actuarial Science, Ankara, Turkey; msucu@hacettepe.edu.tr

Orcid: 0000-0002-4725-3588<sup>1</sup> Orcid: 0000-0002-7991-1792<sup>2</sup>

\* Corresponding Author, e-mail: ovgucan@hacettepe.edu.tr

**Abstract:** In non-life insurance mathematics, analyses and premium or reserve calculations are carried out in the presence of dependency between the claim variables in recent years. And, thus over- or underestimation of aggregate loss caused by the assumption of dependency between the claim severity and frequency are prevented. The Gaussian copula function, which is frequently used for dependency modeling, is integrated into the marginal generalized linear models to obtain a mixed copula-based regression model called "copula regression". In this study, a copula regression model is created using a bivariate Gaussian copula, Gamma and Poisson marginal generalized linear models for claim severity and frequency, respectively. An application is performed with a simulated data where there is a dependence between the claim severity and frequency using the R package "CopulaRegression". The importance of the modeling of dependency between claims is investigated by the comparison of the independent and dependent models and the results of application show that the copula regression model in which dependency is considered has lower relative mean square errors compared the independent marginal generalized linear models.

**Keywords:** Bivariate Gaussian Copula Function; Gamma GLM, Poisson GLM, Dependence, Mixed Copula Approach

## 1. Introduction

Insurance is a multivariate system created under certain conditions and the complex structures caused by the multivariate situations in life and non-life insurance calculations are generally ignored with some basic assumptions. One of the main assumptions used in calculations is that the random variables are independently and identically distributed. Although the independence is a very basic assumption, it is not a very realistic one and it can cause over-or underestimation [1]. The two main components of the non-life insurance mathematics are the claim severity which represents the monetary losses of claims and the claim frequency expressing the numbers of claims. The calculations such as pricing and reserve are carried out under the assumption of independence for many years, but in recent years, the dependency between variables is included to obtain more accurate and realistic estimation, pricing and reserve calculations. In the presence of the dependence, various dependency structures are encountered. Therefore, in studies where the dependency is taken into consideration, first of all, it is necessary to decide the dependence structure, and to include the dependency in the calculations by modeling.

In ratemaking studies, generally claim severity and frequency are modeled by generalized linear models (GLMs) separately, thereafter, expected values of claim severity and frequency are multiplied to calculate the aggregate loss under independence assumption [2, 3, 4]. To avoid the effects of independence assumption, some approaches were proposed to model the dependence. Copula which is the most used method for modeling dependence in financial and statistical studies is also introduced in actuarial studies. Firstly, copula was used to model dependency in contingent life insurance [5], and over time it was used for dependency modeling studies in non-life insurance mathematics [6, 7].

The dependence between claims can be modeled only with copula functions, or marginal GLMs can be included in the copula. Song [6] defined the mixed copula approach using GLMs and Gaussian copula

function and proposed Vector GLM (VGLM) in order to model the dependency between mixed variables. The mixed copula approach lays the groundwork to model dependency between mixed variables such as continuous claim severity and discrete claim frequency. Kastenmeier [8] established a joint regression model for claim severity and frequency using the mixed copula approach. Song et al. [7] modeled dependency between continuous, discrete and mixed variables by using a joint regression analysis. Kolev and Paiva [9] gave some results about regression models based on copulas. Mixed copula model was proposed by Czado et al. [1] to model dependency between claim severity and frequency using Gaussian copula. The usage of copula-based regression models for mixed variables in medical was examined by De Leon and Wu [10].

Generally Gaussian bivariate copula function was used in mentioned studies. However, Krämer et al. [11] used the other parametric copulas such as Clayton, Gumbel and Frank copulas besides Gaussian copula and they referred the models contain GLMs and copula as the copula-based regression models. Krämer et al. [11] also modeled the dependency between claim components only using the copula and claims without GLMs and the approach is entitled as the copula-based models. Copula-based models are useful for modeling the dependence according to only the distribution of claim components without any explanatory variables.

In addition to, copula-based models and copula-based regression models, there are some other approaches for dependency modeling. Gschlößl and Czado [12] introduced a new approach to model dependency between claim severity and frequency by taking the claim frequency as an explanatory variable in the GLM modeling of total loss and Garrido et al. [13] also used same approach to model dependency in non-life insurance. A copula-based multivariate Tweedie regression model was proposed to model semi-continuous claims. [14]. A copula quantile regression approach was used to estimate the parameters of copula [15].

In this study, it is examined how the estimates will change if the dependency between the claim severity and frequency is considered instead of the independence assumption, which is frequently used in non-life insurance mathematics. Therefore, the dependency between the claim severity and frequency is modeled via copula regression model using the bivariate Gaussian copula function and the marginal gamma and Poisson GLMs. An application is performed with a data simulation where there is a dependence between the claim severity and frequency using the information of a real Turkey comprehensive insurance data with the R package "CopulaRegression". The importance of including dependency in calculations is investigated by the comparison of the independent marginal generalized linear model and copula regression model under the assumptions obtained with the real Turkey data. The study can be a guide for researchers who are interested in pricing studies involving dependency in Turkey.

The remainder of the paper is organized as follows. A general information about the bivariate Gaussian copula function and GLM, which are the components of the copula regression model is briefly given in the Section 2. Mixed copula approach and copula regression model are given in Section 3. In Section 4, using the R package "CopulaRegression" [16], joint cumulative distribution function (c.d.f.) and joint probability density function (p.d.f.) of claim severity and frequency are drawn and a simulation study is carried out. The concluding remarks are given in Section 5.

## 2. Methods

Copula-based regression models are obtained by the combination of a bivariate copula function and two marginal GLMs [1, 6, 8, 11] In this study, a bivariate Gaussian copula function is integrated with marginal gamma and Poisson GLMs to create a copula regression model. Definitions and properties of the bivariate Gaussian copula function, gamma GLM and Poisson GLM are given briefly as follows for a better understanding of copula regression model.

### 2.1. The Bivariate Gaussian Copula Function

Copula which is introduced by Sklar [17] is used to model dependence among variables in many disciplines such as economy, finance, econometric, statistics and actuarial science. Copula can be defined as a function which link a multivariate distribution function to their marginal distributions which have standard uniform distributions [18]. Using Sklar's Theorem, where  $\theta$  is the copula parameter, a  $\mathcal{C}(\cdot, \cdot | \theta)$  bivariate parametric copula can be defined by Equation (1) as follows,



$$F_{XY}(x, y|\theta) = C(F_X(x), F_Y(y)|\theta) \quad (1)$$

where  $F_X(x)$  and  $F_Y(y)$  are the marginal distribution functions. Copulas can be defined by Kendall's  $\tau = 4 \int_{[0,1]^2} C(x, y) dC(x, y) - 1 \in [-1,1]$  instead  $\theta$ , due to invariability under monotone transformations of marginal distributions [11].

Gaussian copula function is used in many studies due to the advantages of multivariate normal distribution [6, 19]. The relationship between  $\theta$  and Kendall's  $\tau$  for Gaussian copula is  $\tau = \frac{2}{\pi} \arcsin(\theta)$ . Let  $\Phi(\cdot)$  and  $\Phi_2(\cdot|\Gamma)$  denote the univariate and the bivariate standard normal distribution functions with  $\Gamma$  correlation matrix, respectively. A bivariate Gaussian copula  $C: I^2 \rightarrow I$  is given as follows,

$$C(u_1, u_2|\Gamma) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)|\Gamma) \quad (2)$$

$$C(u_1, u_2|\Gamma) = \frac{\partial}{\partial u_1} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-(\rho_{12})^2}} \exp\left\{-\frac{s^2-2\rho_{12}st+t^2}{2(1-(\rho_{12})^2)}\right\} ds dt$$

where  $(u_1, u_2) \in I^2$ ,  $i = 1, \dots, n$  is the number of observations and  $\Gamma = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}$  displays the correlation matrix.

## 2.2. Generalized Linear Model

GLM is a generalized form of a linear model which models the relationship between a response variable and explanatory variables. GLM consists of a density function of the response variable from the exponential family, a linear component and a link function [20]. In GLM, response variable  $y$  follows an exponential family distribution such as Poisson, binomial, negative binomial, normal, gamma, inverse Gaussian, etc. Linear component is an instrument that represents the relationship between the response variable and explanatory variables. The linear component for  $i^{\text{th}}$  observation  $\eta_i = \mathbf{X}'_i \boldsymbol{\beta} = \sum_{j=1}^p \beta_j x_{ij}$  where  $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$  and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$  for  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, p$ . The number of explanatory variables is  $p - 1$ , the first term of  $\mathbf{X}_i$  is model constant and equal to 1 for all observations. Monotone and differentiable link function links the expected value of the response variable and the linear component. According to distribution of the response variable, link function is determined as identity, logarithmic, power, square root and logit. For  $E(y_i) = \mu_i$ , the general form of the link function is  $g(\mu_i) = \eta_i$ . observation  $\eta_i = \mathbf{X}'_i \boldsymbol{\beta} = \sum_{j=1}^p \beta_j x_{ij}$  where  $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$  and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$  for  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, p$ . The number of explanatory variables is  $p - 1$ , the first term of  $\mathbf{X}_i$  is model constant and equal to 1 for all observations. Monotone and differentiable link function links the expected value of the response variable and the linear component. According to distribution of the response variable, link function is determined as identity, logarithmic, power, square root and logit. For  $E(y_i) = \mu_i$ , the general form of the link function is  $g(\mu_i) = \eta_i$ .

GLMs for gamma and Poisson distributed response variables are called as gamma and Poisson GLMs, respectively. Logarithmic link function is widely used for both GLMs. Let  $X \sim \text{Gamma}(\mu, \nu^2)$  and  $Y \sim \text{Poisson}(\lambda)$ , the p.d.f.s of  $X$  and  $Y$  are given in Equations (3) and (4) according to the mean parametrization. Here,  $\mu$  and  $\lambda$  are the mean parameters, while  $\nu^2$  is the dispersion parameter.

$$f_X(x|\mu, \nu^2) = \frac{1}{\Gamma(\frac{1}{\nu^2})} \left(\frac{1}{\mu\nu^2}\right)^{\frac{1}{\nu^2}} y^{\left(\frac{1}{\nu^2}\right)-1} \exp\left(-\frac{x}{\mu\nu^2}\right), \quad x \geq 0 \quad (3)$$

$$f_Y(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y = 0, 1, 2, 3, \dots \quad (4)$$

Gamma and Poisson GLMS with logarithmic link function are given by Equations (5) and (6), respectively as follows. Here,  $\mathbf{z}'_1 \in R^p$  and  $\mathbf{z}'_2 \in R^q$  are the explanatory variable vectors of the claim severity  $X$  with  $p$  parameters and the claim frequency  $Y$  with  $q$  parameters.  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are the parameter vectors of the random variables  $X$  and  $Y$ .  $\ln(e)$  is the offset term, where  $e$  is the exposure to risk.

$$g(x) = \ln(\mu) = \mathbf{z}'_1 \boldsymbol{\alpha} \quad (5)$$

$$g(y) = \ln(\lambda) = \ln(e) + \mathbf{z}'_2 \boldsymbol{\beta} \quad (6)$$

### 3. Copula Regression Model

A mixed copula-based regression model called ‘‘copula regression’’ can be obtained by a bivariate copula function and two marginal GLMs. For the purpose of modeling dependency between claim components, claim severity and frequency are modeled via gamma and Poisson GLMs, respectively. Then the marginal GLMs and a bivariate copula function are combined to define ‘‘copula regression model’’. Generally, Gaussian copula function is used in this combination due to the advantages of the Gaussian distribution [1, 6, 8]. However, Krämer et al. [11] used Clayton, Gumbel and Frank copulas besides Gaussian copula. The integration of bivariate Gaussian copula and the marginal GLMs is briefly summarized by Figure 1.



Figure 1. Copula regression model

Copula regression model is comprised of marginal GLMs and a bivariate Gaussian copula function as summarized by Figure 1 using the mixed copula approach proposed by Song [6]. The mixed copula approach is used to model the dependency between mixed variables such as continuous claim severity and discrete claim frequency. For a bivariate Gaussian copula function with the correlation matrix  $\boldsymbol{\Gamma}$ , copula regression model can be shown as follows by Equation (7).

$$C(\text{Gamma GLM}, \text{Poisson GLM} \mid \boldsymbol{\Gamma}) \quad (7)$$

The mixed copula approach, allows the usage of copula functions which, are used with only the continuous random variables, also together with discrete random variables. According to the Sklar’s Theorem [16], the joint c.d.f. of mixed variables  $X$  and  $Y$  is written by Equation (8). The joint p.d.f. of mixed variables  $X$  and  $Y$  is obtained by mixed copula approach [6] by Equation (9).

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = C(F_X(x), F_Y(y) \mid \boldsymbol{\Gamma}) \quad (8)$$

$$f_{XY}(x, y) = \frac{\partial}{\partial x} P(X \leq x, Y = y) \quad (9)$$

For a Gaussian bivariate function, using Radon-Nikodym derivative and the information of  $\frac{\partial}{\partial x} C(u_1, u_2) = C_1'(u_1, u_2 \mid \boldsymbol{\Gamma}) = \Phi\left(\frac{\Phi^{-1}(u_2) - \rho_{12}\Phi^{-1}(u_1)}{\sqrt{(1-\rho_{12}^2)}}\right) := D_{\rho_{12}}(u_1, u_2)$ , the p.d.f. of mixed variables  $X$  and  $Y$  can be expanded by Equation (10) as follows where  $F_X(x) = u_1$  and  $F_Y(y) = u_2$  [1, 6, 8].

$$\begin{aligned} f_{XY}(x, y) &= \frac{\partial}{\partial x} P(X \leq x, Y \leq y) - \frac{\partial}{\partial x} P(X \leq x, Y \leq y - 1) \\ &= \frac{\partial}{\partial x} C(F_X(x), F_Y(y) \mid \boldsymbol{\Gamma}) - \frac{\partial}{\partial x} C(F_X(x), F_Y(y - 1) \mid \boldsymbol{\Gamma}) \\ f_{XY}(x, y) &= f_X(x) [D_{\rho_{12}}(F_X(x), F_Y(y)) - D_{\rho_{12}}(F_X(x), F_Y(y - 1))] \end{aligned} \quad (10)$$

The parameter vector of a copula regression is  $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau)$  where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are the parameter vectors of gamma and Poisson GLMs and  $\tau$  is the Gaussian copula parameter. The parameter estimation is carried out by maximum-likelihood techniques.

Krämer et al. [11] were defined the relative MSE as  $MSE_{rel} := E\left(\frac{1}{k} \sum_{i=1}^k \left(\frac{y_i - \hat{y}_i}{y_i}\right)^2\right)$  for the parameter vector  $\boldsymbol{\gamma} \in R^k$ , where  $\hat{y}_i$  is the estimate of the parameter. To evaluate the performance of dependent model (copula regression) and independent model (marginal GLMs), the relative MSEs of the estimators

of parameters of GLMs  $\alpha \in \mathbf{R}^p$  and  $\beta \in \mathbf{R}^q$ , and also the relative MSE of the estimator of the parameter of Gaussian copula  $\tau$  is calculated.

#### 4. Application

Analyses are performed in R using the R package “CopulaRegression” [16] which presents a bivariate, copula-based model for the joint distribution of a pair continuous and discrete random variables. The R packages “MASS” [21] and “VineCopula” [22] are also used, since they work based on “CopulaRegression” package. The pair of continuous and discrete random variables is composed as a couple of claim severity and frequency. The claim severity and frequency are modeled by marginal gamma and Poisson GLMs, respectively. The marginal GLMs are linked by the bivariate Gaussian copula function to model dependence between claim severity and frequency.

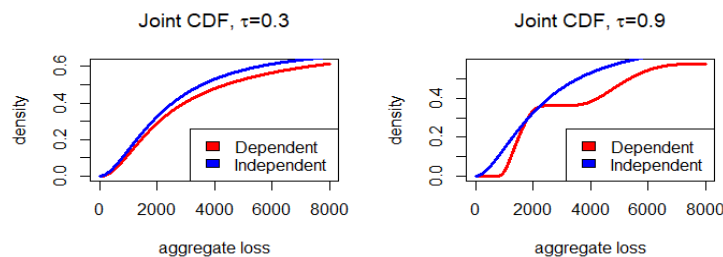
An insurance system is designed using the comprehensive insurance data in Erdemir and Sucu [23] to compare the copula regression and the independent models. The data taken from a Turkish non-life insurance company for year 2017 contains the information about the claim severity, claim frequency and some risk factors related the policyholders and the vehicles; such as age of policyholder (21,...,80), gender (male, female), type of vehicle (automobile, pickup, rent a car, taxi), age of vehicle (0,...,17), usage (private, leasing, commercial), residence (metropolis, little town), engine capacity of vehicle (small, medium, large), type of fuel (benzine, diesel) and status of the policy (new, renewal) of 2820 observations. The descriptive statistics of the comprehensive insurance data is given by Table 1.

**Table 1.** Descriptive statistics of comprehensive insurance data

Variables	Minimum	Maximum	Mean	Median	Variance
Claim Severity (X)	51.24	35477.00	1759.4936	2933.58735	8605934.753
Claim Frequency (Y)	1	4	1.7034	0.4678	0.2190

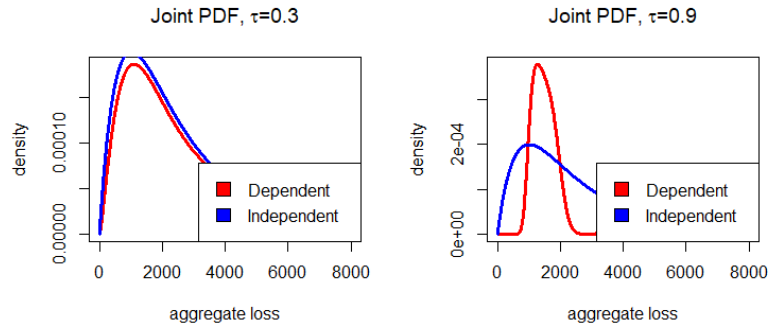
By the reason of, the package “CopulaRegression” is designed to generate its own dependent data, the information of the data is only used to make some assumptions of simulation study. The dispersion parameter of the gamma distribution is assumed constant as  $v^2=0.5$  for an easy calculation and the method of moments is used to determine the parameters of the gamma and Poisson distributions. The mean parameters of gamma and Poisson distributions are found as  $\mu=1760$  and  $\lambda=1.18$ , respectively [23].

First, the graphs of joint c.d.f. and p.d.f. of claim severity and frequency are plotted considering the dependency between the claim variables using the mixed copula approach and also drawn under the independence assumption to compare the dependent and independent models. “CopulaRegression” package use the Kendall’s  $\tau$  correlation coefficient instead of  $\rho$  Spearman correlation coefficient in some analysis. These two coefficients can be easily converted into each other through the copula parameter. The copula parameter is the correlation coefficient for the Gaussian copula function, hence Kendall’s  $\tau$  can be directly associated dependency with  $\tau = \frac{2}{\pi} \arcsin(\rho)$ . Kendall’s  $\tau$  is taken 0 for independent model, since  $\frac{2}{\pi} \arcsin(0) = 0$ . For the copula regression, the values of  $\tau$  are chosen as 0.3 and 0.9 for the low and high degrees, respectively. The graphs of the joint c.d.f. and p.d.f. are given in Figures 2 and 3. The effect of considering the dependency between claim components and also the effect of  $\tau$  on dependence modeling are investigated.



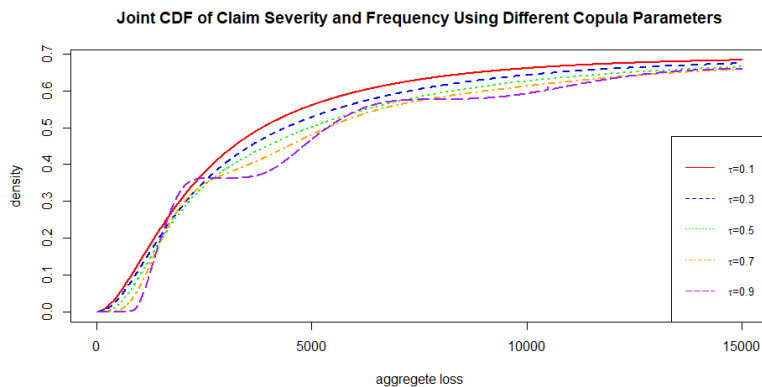
**Figure 2.** Joint c.d.f. of claim severity and frequency

Since the graphs given in Figure 2 are the graphs of cumulative distribution, it is an expected result to obtain increasing curves for both  $\tau$  values. The probability values for the dependent model which are displayed by red lines are smaller for both  $\tau$  values. The effect of dependence cannot be directly analyzed since the probabilities are expressed cumulatively, however, a little fluctuation is observed for highly dependent model when  $\tau=0.9$ .



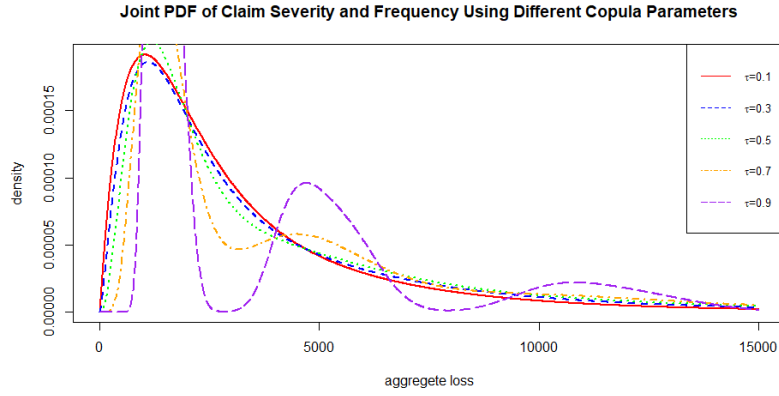
**Figure 3.** Joint p.d.f. of claim severity and frequency

Right skewness is observed in Figure 3, since gamma distribution assumption is used for claim severity. The effect of considering the dependency between claim components on aggregate loss is observed clearly and it is noticed that the p.d.f. is very sensitive to dependency for higher value of  $\tau$ . A hillier distribution is obtained with the dependent model. For  $\tau=0.3$ , that is, when the dependency level is low, both the graphs of c.d.f. and p.d.f. are similar in the two models according to the Figures 2 and 3. However, when the dependency level is high, the graphs differ for the dependent and independent models.



**Figure 4.** Joint c.d.f. of claim severity and frequency using different Kendall's  $\tau$

Five different values of the Kendall's  $\tau$  (0.1, 0.3, 0.5, 0.7, 0.9) are taken for Gaussian copula function and to analyze the effects of Kendall's  $\tau$  on joint c.d.f. and p.d.f., the graphs are redrawn by different  $\tau$  values, the graphs are given by Figure 4 and Figure 5 as follows. The effect of  $\tau$  cannot be directly observed in Figure 4 as in Figure 2, due to the cumulative structure. However, according to Figure 5, as the  $\tau$  value increases, the fluctuations in the probability value of the aggregate loss increase. The total loss is more sensitive to higher  $\tau$  values. It can be deduced that it is important to include dependency in the calculations.



**Figure 5.** Joint p.d.f. of claim severity and frequency using different Kendall's  $\tau$

Joint c.d.f. and p.d.f. are drawn independently of the possible explanatory variables. Explanatory variables can be included by the copula regression model using marginal GLMs. In the R package “CopulaRegression”, `copreg()` function fits a joint, bivariate regression model for a gamma GLM and a (zero-truncated) Poisson GLM. On the purpose of fitting model, the package simulates the joint regression data under some assumption about the distributions and copula function. Then using the simulated data, the copula regression model is created. An insurance system is designed using the comprehensive insurance data in Erdemir and Sucu [23] and a little simulation study with  $R=50$  trials is performed using the R package. The system with  $n=1000$  policy groups with only the automobile type of vehicle is considered and the groups contain insured with gamma-distributed claim severity with the parameters ( $\mu=1760$ ,  $\nu^2=0.5$ ) and Poisson distributed claim frequency with the parameter ( $\lambda=1.18$ ). It is assumed that all policy groups contain the same number of policyholders. The claim severity and frequency are modeled by gamma and Poisson GLMs, respectively. Gender, residence and engine power of vehicle are determined as the explanatory variables for GLM modeling and all explanatory variables are assumed as categorical variables. Gender (male-female) and residence (metropolis, little town) are two-category variables, hence they are modeled by only one dummy variables. On the other hand, since the engine capacity of vehicle (low, medium, high) is three-category variable, it is modeled with two dummy variables. Same explanatory variables are used for gamma and Poisson GLMs. Marginal GLMs are designed with an intercept term, hence the first column of the design matrices contains 1's as  $Z_1 := Z_2 := (1, z_{12}, \dots, z_{1n}) := (1, z_{22}, \dots, z_{2n}) \in R^{1000 \times 5}$ . The second and third columns are dummy variables corresponding to female and metropolis, respectively. The last two columns are the two dummy variables corresponding to low and high the engine capacity of vehicle. All dummy variables are generated randomly such as 0 or 1.

For the copula regression model, three different values are determined for  $\tau$  as 0.1, 0.5 and 0.9 to represent the low, moderate and high levels of dependency, while  $\tau$  is taken as 0 for independent model. Under these assumptions, the gamma distributed claim severity and the Poisson distributed claim frequency dependent on each other are generated by the R package. The dependent data changes according to the value of  $\tau$  in R, hence the values of relative MSE change for independent model for different  $\tau$  values.

The simulation study is performed for the comparison of the relative MSEs found with copula regression and independent model. A Monte Carlo simulation is carried out by  $R=50$  trials and  $\overline{MSE}_{rel}^{(r)} := E \left( \frac{1}{k} \sum_{i=1}^k \left( \frac{y_i - \hat{y}_i^{(r)}}{y_i} \right)^2 \right)$  is calculated in the  $r^{\text{th}}$  step. The mean of all simulations is obtained with  $\overline{MSE}_{rel} := \frac{\sum_{r=1}^R \overline{MSE}_{rel}^{(r)}}{R}$  and the mean relative MSEs of  $\hat{\alpha}$ ,  $\hat{\beta}$  and  $\tau$  are calculated and given in Table 1 as follows. The variance can be calculated with the formula  $\frac{\sum_{i=1}^R (MSE_{rel(i)} - \overline{MSE}_{rel})^2}{R(R-1)}$ .

**Table 1.** Relative MSEs of  $\hat{\alpha}$ ,  $\hat{\beta}$  and  $\tau$  with copula regression and independent models ( $\tau=0, 0.1, 0.5, 0.9$ )

	Copula Regression Model	Independent Model
$\overline{MSE}_{rel}(\hat{\alpha})$ ( $\tau=0.1$ )	<b>0.005721400</b>	0.00600650
$\overline{MSE}_{rel}(\hat{\alpha})$ ( $\tau=0.5$ )	<b>0.004993235</b>	0.00583436
$\overline{MSE}_{rel}(\hat{\alpha})$ ( $\tau=0.9$ )	<b>0.000778456</b>	0.00515738
$\overline{MSE}_{rel}(\hat{\beta})$ ( $\tau=0.1$ )	<b>0.018385800</b>	0.02009030
$\overline{MSE}_{rel}(\hat{\beta})$ ( $\tau=0.5$ )	0.001899537	<b>0.00184171</b>
$\overline{MSE}_{rel}(\hat{\beta})$ ( $\tau=0.9$ )	<b>0.004126288</b>	0.04266449
$\overline{MSE}_{rel}(\hat{\tau})$ ( $\tau=0.1$ )	<b>0.505447100</b>	0.54062490
$\overline{MSE}_{rel}(\hat{\tau})$ ( $\tau=0.5$ )	<b>2.14184E-05</b>	5.7223E-05
$\overline{MSE}_{rel}(\hat{\tau})$ ( $\tau=0.9$ )	<b>1.40001E-06</b>	0.00012083

Smaller values of relative MSE of parameter estimators are displayed bold in Table 1. It is noticed that, in the presence of dependence between of the claim severity and frequency, smaller errors are calculated using the copula regression model.

AIC values of copula regression and independent models are also calculated to compare the models using  $AIC = -2l(\hat{\gamma}) + 2df$  where  $l(\gamma) = \sum_{i=1}^n \ln f_{XY}(x, y)$  and the results of the comparison of AICs are given in Table 2.

**Table 2.** AIC values of copula regression and independent models

	Copula Regression Model	Independent Model
AIC ( $\tau=0.1$ )	<b>9946.862</b>	10021.906
AIC ( $\tau=0.5$ )	<b>9251.730</b>	9942.6440
AIC ( $\tau=0.9$ )	<b>6467.568</b>	10029.176

For all  $\tau$  values, in that for low, moderate and high degree of dependence, smaller AIC values are calculated for the copula regression models compared to the independent models.

## 5. Concluding Remarks

In non-life insurance mathematics, the dependence between claim severity and frequency has been modeled and included in the calculations in recent years. Copula-based models are converted into copula-based regression models with GLMs. The effects of the possible explanatory variables are also included in dependency modeling via copula-based models called “copula regression models”. Copula regression models can be obtained with a mixed copula approach for continuous and discrete variables. Due to the mixed copula approach, the copula function, which can only be used with continuous variables, can be used with both discrete and continuous variables. It provides a flexible calculation in the branch that includes both discrete and continuous variables such as non-life insurance mathematics.

In this study, the dependency between the claim severity and frequency is modeled via copula regression model using the bivariate Gaussian copula function and the marginal gamma and Poisson GLMs. The effects of considering the dependency between claim variables are investigated by the comparison of the independent model and the copula regression model. The joint c.d.f. and p.d.f of claim severity and frequency are plotted for both models considering different dependency degrees. The importance of modeling dependency claim components is observed clearly with especially the graph of p.d.f. of aggregate loss. In addition, an insurance system is designed under some assumptions using the information of a real Turkish comprehensive insurance data. The relative MSEs are calculated for copula regression and independent models using different  $\tau$  values. It is noticed that, copula regression models have smaller relative errors. Furthermore, AIC values are calculated for both models and the values support the result found with relative MSE values. In the light of these results, researchers studying on pricing or reserve in the non-life insurance mathematics can make more accurate calculations, including the dependence between the claim frequency and severity. With more accurate pricing policies, companies prevent problems with the ability to meet solvency margin. Since

the assumptions of application are based on real Turkish comprehensive data, this study can be a good guide for pricing studies for Turkey where the dependency between claim components is considered.

**Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “Conceptualization, Ö.K.E. and M.S.; methodology, Ö.K.E.; software, Ö.K.E.; validation, Ö.K.E. and M.S.; formal analysis, Ö.K.E.; investigation, Ö.K.E.; resources, Ö.K.E.; data curation, Ö.K.E.; writing—original draft preparation, Ö.K.E.; writing—review and editing, Ö.K.E. and M.S.; visualization, Ö.K.E. and M.S.; supervision, M.S. All authors have read and agreed to the published version of the manuscript.” Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

**Funding:** No financial resources were provided for this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- [1] C. Czado, R. Kastenmeier, E. C. Brechmann and A. Min, “A mixed copula model for insurance claims and claim sizes”, *Scand. Actuar. J.*, vol. 4, pp. 278-305, 2012.
- [2] Y. K. Tse. “Nonlife actuarial models: theory, methods and evaluation”, Cambridge University Press, 2009.
- [3] E. Ohlsson, B. Johansson, “Non-life insurance pricing with generalized linear models”, Springer, 174, 2010.
- [4] M. David, “Automobile insurance pricing with generalized linear models”, *Proceedings in GV- The 3rd Global Virtual Conference*, 6-10 April, 2015.
- [5] E. W. Frees and E. A. Valdez, “Understanding relationships using copulas”, *N. Am. Actuar. J.*, vol. 2, pp. 1-25, 1998.
- [6] P. X. K. Song, “Correlated data analysis: modeling, analytics, and applications”, Springer Science & Business Media, 2007.
- [7] P. X. K. Song, M. Li and Y. Yuan, “Joint regression analysis of correlated data using Gaussian copulas”, *Biometrics*, vol. 65(1), pp. 60-68, 2009.
- [8] R. Kastenmeier, “Joint regression analysis of insurance claims and claim sizes”, Diploma Thesis, Technische Universität München, Mathematical Sciences, 2008.
- [9] N. Kolev and D. Pavia, “Copula-based regression models: A survey”. *J Stat Plan Inference*, vol. 139(11), pp. 3847-3856, 2009.
- [10] A. R. De Leon and B. Wu, “Copula-based regression models for a bivariate mixed discrete and continuous outcome”, *Stat Med*, vol. 30(2), pp. 175-185, 2011.
- [11] N. Krämer, E. C. Brechmann, D. Silvestrini and C. Czado, “Total loss estimation using copula-based regression models”, *Insur Math Econ*, vol. 53(3), pp. 829-839, 2013.
- [12] S. Gschlößl and C. Czado, “Spatial modelling of claim frequency and claim size in non-life insurance” *Scand*, vol. 3, pp. 202-225, 2007.
- [13] J. Garrido, C. Genest and J. Schulz, “Generalized linear models for dependent frequency and severity of insurance claims”, *Insur Math Econ*, vol. 70, pp. 205-215, 2016.
- [14] P. Shi, “Insurance ratemaking using a copula-based multivariate Tweedie model”, *Scand. Actuar. J.*, vol. 3, pp. 198-215, 2016.
- [15] A. T. Payandeh Najafabadi, M. Qazvini, “A GLM approach to estimating copula models”, *Comm. Statist. Simulation Comput.*, vol. 44 (6), pp. 1641-1656, 2015.
- [16] N. Krämer, D. Silvestrini and M. N. Krämer, Package ‘CopulaRegression’, 2013.
- [17] A. Sklar, “Fonctions de répartition à n dimensions et leurs marges”, *Publications de l’Institut de Statistique de L’Université de Paris*, vol. 8, pp. 229-231, 1959.
- [18] R. B. Nelsen, “An introduction to copulas”, Springer Science & Business Media, 2006.
- [19] D. Brigo, A. Pallavicini and R. Torresetti, “Credit Models and The Crisis: A Journey Into Cdos, Copulas”, *Correlations And Dynamic Models*, John Wiley & Sons, 2010.
- [20] P. McCullagh and J. A. Nelder, “Generalized Linear Models”, CRC press, 37, 1989.
- [21] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth and M. B. Ripley, Package ‘mass’, *Cran R*, 2013.
- [22] U. Schepsmeier, J. Stoeber, E. C. Brechmann, B. Graeler, T. Nagler and T. Erhardt, “VineCopula: Statistical Inference of Vine Copulas”, R package version 1, 2012.
- [23] Ö. K. Erdemir and M. Sucu, “A comparative study on modeling of dependency between claim severity and frequency”, *J. Stat.: Stat and Actuar. Sci.*, vol. 13(1), pp. 18-29, 2020.

Research Article

Received: date:21.03.2022

Accepted: date:28.03.2022

Published: date:30.06.2022

# Selecting an Alternative Concrete Batching Plant Location Using IT2 Fuzzy ANP Methodology

Kemal Gokhan Nalbant<sup>1</sup>, Sahika Özdemir<sup>2</sup> Yavuz Özdemir<sup>3,\*</sup>

<sup>1</sup>Department of Software Engineering, Beykent University, Istanbul, Turkey; kemalnalbant@beykent.edu.tr

<sup>2</sup>Department of Interior Architecture and Environmental Design, Istanbul Sabahattin Zaim University, Istanbul, Turkey; sahika.ozdemir@izu.edu.tr

<sup>3</sup>Department of Industrial Engineering, Istanbul Health and Technology University, Istanbul, Turkey; yavuz.ozdemir@istun.edu.tr

Orcid: 0000-0002-5065-2504<sup>1</sup> Orcid: 0000-0002-5762-1962<sup>2</sup> Orcid: 0000-0001-6821-9867<sup>3</sup>

\* Correspondence Author, yavuz.ozdemir@istun.edu.tr

**Abstract:** Ready-Mixed Concrete started to be used after the 1970s, and its importance has gradually increased in recent years. The rapid development of the construction industry in our country in recent years has contributed to the rapid development of Ready-Mixed Concrete plants. Because of this reason, in this study, the concrete plant's location is determined using the Interval Type-2 Fuzzy Analytic Network (IT2 FANP) methodology. As a result of this methodology, the ranking of the alternative concrete plant's location is found as Basaksehir, Ikitelli, and Bagcilar, respectively. To the authors' knowledge, this is the first study in the literature that searches for an alternative concrete batching plant location using Interval Type-2 (IT2) Fuzzy Multi-Criteria Decision Making (MCDM) techniques. The problem may be examined using various MCDM methodologies, and further solutions for selecting alternate concrete plant locations can be researched in the future. The value of this study is that it fills in a gap in the existing literature and serves as a foundation for future research in this sector.

**Keywords:** Ready-Mixed Concrete; concrete batching plant; location selection; decision making; IT2 FANP; Multi-Criteria Decision Making (MCDM)

## 1. Introduction

Although Ready-Mixed Concrete was first used in our country in the 1970s, the construction started in the early 1990s, and the building became widespread in the early 2000s. Today, manual concrete casting is almost nonexistent. Ready-Mixed Concrete technology has become so widespread that nearly every district has at least one concrete plant.

In addition to the widespread use of ready-mixed concrete, the production of concrete in our country has increased in recent years in parallel with the development of the construction industry. Production dimensions in the sector have increased; since 2009, it has become the country with the maximum production in Europe.

Concrete, which has a vast usage area today, has become the most used material after water. Concrete is the essential building material that is economical, safe, durable, and does not require much maintenance. Concrete is chosen with a very high percentage of more than 99% in-house selection in our country [1].

The reasons for the rapid increase in ready-mixed concrete construction in recent years are:

- Growth in the construction industry in recent years,
- The increase in the supply of housing (Mass Housing and Public Partnership Administration's (TOKI's) house constructions throughout the country has been effective), and there are significant projects where ready-mixed concrete is used extensively,
- Selecting the most reinforced concrete structures in Turkey,
- Ready-mixed concrete is cheap, economical, and easy to use,
- It facilitates the supply of ready-mixed concrete thanks to the concrete batching plants opened all over Turkey.



A concrete batching plant is a system for the production of fresh concrete, consisting of storage parts, combining and blending aggregate, cement, water, and additives when necessary, mixing, obtaining new concrete, and discharging the fresh concrete into the mobile concrete mixer with a trans mixer [2]. The primary element that distinguishes ready-mixed concrete from the concrete prepared at the construction site by hand or mixed with a concrete mixer is the production of ready-mixed concrete in modern establishments with computer control. The qualities that ready mixed concrete users will look for in ready mixed concrete are included in the Turkish Standardization Institute, TS EN 206 [3].

The first scientific study on ready mixed concrete in Turkey was made by Öztekin et al. [4]. In this study, fresh concrete temperature, slump, and compressive strength were investigated to compare the quality of ready-mixed concrete at the plant exit and on-site delivery. As a result of the research, it has been stated that the transportation time, air temperature, and humidity of fresh concrete are adequate for its workability (consistency) [4-6].

The choice of establishment location is one of the most basic and critical strategic decisions that should be considered during the establishment phase. Making this decision wrong significantly reduces future efficiency and competitiveness from a micro point of view. Because the supply of inputs to be used by the enterprise and the effective marketing of the outputs will keep some problems (production costs, transportation expenses, marketing, administrative expenses, etc. high, problems experienced in the supply of work energy of the desired quality) on the agenda. From a macro point of view, it is seen that it prevents the effective use of the country's resources. The selection of the establishment location is made depending on many factors. Due to the nature of this selection, these factors have both quantitative and qualitative characteristics. Persons and managers in the position of decision-makers need a decision support system that will examine these factors in a very complex structure and number, regardless of the limits of the human brain, to make these decisions healthily.

IT2 Fuzzy Set (FS) is a special version of the generalized Type-2 (T2) FS. Many literature studies were done to develop IT2 FS [7-12]. First, to solve a problem, it is necessary to decide. Many areas are faced with decision-making problems. Issues that have multiple criteria and choose one of the alternatives are included in MCDM problems. Many studies using MCDM techniques can be found in the literature [12-25].

IT2 FANP method can also be used to solve MCDM problems. Senturk et al. [26] proposed a new approach, a Fuzzy Analytic Network (FANP) method with interval type-2 fuzzy sets. Wu and Liu [27] proposed the FANP method with IT2 FSs to evaluate the Enterprise Technology Innovation Ability (ETIA). Senturk et al. [28] developed a new IT2 FANP methodology to model a Third-party Logistics (3PL) company selection problem. Ozdemir et al. [29] proposed a new hybrid model based on Interval Type-2 Fuzzy Analytic Network Process (IT2 FANP) and Interval Type-2 Fuzzy Technique for Order Preference by Similarity to an Ideal Solution (IT2 TOPSIS) for the evaluation of store plan alternatives produced with rule-based design method. Ozdemir et al. [30] prioritized store plan alternatives using the Interval Type-2 Fuzzy Analytic Network (IT2 FANP) methodology, and the best alternative for store plans was selected.

Studies related to concrete batching plants in the literature were examined. There are very few studies using MCDM techniques in this area. Simsek et al. [31] used a TOPSIS-based Taguchi optimization to determine optimal mixture proportions of the high-strength self-compacting concrete. Simsek and İc [32] made a fuzzy Failure Mode and Effect Analysis (FMEA) application to reduce the risk level in a ready-mixed concrete plant. Their results showed that the fuzzy FMEA methodology effectively identified and eliminated potential failure modes at the ready-mixed concrete plant.

When the literature was examined, no study could be found on a concrete batching plant using IT2 fuzzy MCDM techniques. Literature studies using fuzzy and IT2 fuzzy MCDM methods related to the construction sector were examined. There are very few studies in the literature in this sector. Abdelgawad and Fayek [33] aimed to extend the application of FMEA to risk management in the construction industry. They combined Fuzzy FMEA and Fuzzy Analytical Hierarchy Process (FAHP) for the construction industry. Moreover, studies on the construction sector using IT2 fuzzy MCDM techniques have been found in the literature. Debnath and Biswas [34] proposed a method for assessing

the risk of the workers at construction sites using an interval type-2 fuzzy analytic hierarchy process. When the literature is examined, there is no study in which the IT2 FANP method is used in the construction industry field.

IT2 fuzzy set has been used to resolve MCDM problems in the published research. Kou et al. [35] evaluated Fintech-based investments of European banking services using an application of an original methodology that takes into consideration interval type-2 (IT2) fuzzy decision-making trial and evaluation laboratory and IT2 fuzzy TOPSIS models. This was done in order to determine the value of the investments. Bera et al. [36] offered two unique MCDM strategies in an interval type-2 fuzzy (IT2F) environment. These techniques can handle uncertain subjective and objective elements concurrently, which is necessary for selecting efficient suppliers to be used in real-world applications. Technique for order preference by similarity to the ideal solution (TOPSIS) and multi-objective optimization on the basis of ratio analysis (MOORA) methods are used in the IT2F environment. Zhou et al. [37] investigated the factors that contribute to the success of the loan application procedure for large-scale energy projects undertaken by financial institutions. In this scenario, a three-stage analysis is used as the appropriate method. When it comes to the topic of financing energy projects, the IT2 fuzzy DEMATEL and IT2 fuzzy QUALIFLEX approach are two methods that are used. Liu et al. [38] analyzed various renewable energy sources and came up with the best options for usage in blockchain-based financial transactions. The significant levels of these criteria are determined with the assistance of the interval type-2 (IT2F) decision making trial and evaluation laboratory (DEMATEL)-analytical hierarchy process (ANP) (DANP) approach. Additionally, the fuzzy IT2 VIKOR technique has been considered.

The originality of this study is to use IT2 FANP, one of the fuzzy multi-criteria decision-making methods, to determine the location of the concrete plant. Because, as can be seen, no study related to the concrete batching plant using this method has been found in the literature. It has been concluded that this method is not used even in the construction industry field. The advantage of this study is to enrich this aspect, which is seen as missing in the literature, and to be a pioneer for studies to be made in this field.

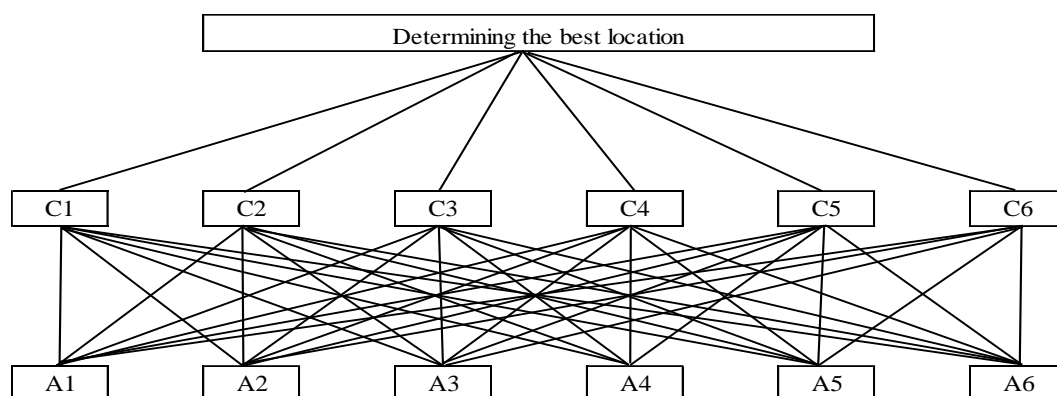
The ANP technique enables the consideration of many primary criteria and sub-criteria, the calculation of the weights of each of these criteria as the choice is being made, pairwise comparisons of criteria, and links between criteria. As a result of the fact that in paired matrices, fuzzy ANP may be used for the computation of the weights of criterion. Additionally, interval type-2 fuzzy sets include fuzziness in the membership functions of the set. Consequently, the interval type-2 fuzzy set technique is more sensitive when modeling uncertainty [26]. The type-2 fuzzy technique allows uncertainty determination by including fuzziness for the membership functions. This contrasts with the fuzzy AHP and fuzzy ANP approaches, which have been the subject of numerous articles in the past. Because they are simpler to calculate with, interval type-2 fuzzy sets are the ones that are recommended [28].

Interval type-2 fuzzy sets (IT2 FSs) can handle linguistic uncertainty variables more flexibly and precisely than type-1 fuzzy sets with their second fuzzy membership functions [27]. So, IT2 FANP was chosen for this study. You may get more detailed information on the procedures involved in this approach from [26, 28, 29]. In [26], an interval type-2 FANP technique is created, and it is given in the academic literature for the first time. Specialists evaluated the criteria following the linguistic scale FANP as type-2 fuzzy numbers. Interval type-2 fuzzy sets and arithmetic operations between trapezoidal interval type-2 fuzzy sets are first explained. An interval type-2 fuzzy ANP method for modeling vagueness originating from both the linguistic variables of experts and membership functions is discussed [26, 28].

In the second section, an IT2 FANP method for prioritizing location alternatives for the concrete plant is made. Finally, the "Conclusion" section concludes the article and discusses the comparison of outcomes and future research directions.

## 2. Problem Definition and a Numerical Case Study

The MCDM problem is selecting an alternative concrete batching plant location. In this study, the IT2 FANP methodology was used to determine the concrete plant's location. In this section, we apply the IT2 FANP technique found in the study [26]. Three decision-makers determined decision criteria and alternatives for the problem in Figure 1. 6 criteria were chosen and shown in Table 1. These criteria are chosen; C1 as Proximity to Market (Construction Site), C2 as Compliance with Legal Conditions, C3 as Competition Conditions, C4 as Proximity to the Center, C5 as Traffic Conditions, and C6 as Proximity to Logistics Centers.



**Figure 1.** Hierarchy of the selection problem.

**Table 1.** Criteria of problem.

Criteria	
C1	Proximity to Market (Construction Site)
C2	Compliance with Legal Conditions
C3	Competition Conditions
C4	Proximity to the Center
C5	Traffic Conditions
C6	Proximity to Logistics Centers

The location alternatives are A1, A2, A3, A4, A5, and A6. These alternatives are chosen; A1 as Basaksehir, A2 as Gungoren, A3 as Bagcilar, A4 as Ikitelli, A5 as Esenler, and A6 as Zeytinburnu. These are shown in Table 2.

**Table 2.** Location Alternatives.

Alternative	Explanation
A1	Basaksehir
A2	Gungoren
A3	Bagcilar
A4	Ikitelli
A5	Esenler
A6	Zeytinburnu

IT2 fuzzy scales can be seen in Table 3 [39]. Also, these terms can be used in IT2 FANP. Following establishing a network, an expert compared the main criteria using the type-2 fuzzy scale of the linguistic variables shown in Table 3.

**Table 3.** IT2 fuzzy scales.

Low/high Levels	Linguistic Terms	Trapezoidal IT2 fuzzy scales
F	Fairly Strong	(3,4,6,7;1,1) (3.2,4.2,5.8,6.8;0.8,0.8)
E	Exactly Equal	(1,1,1,1;1,1) (1,1,1,1;1,1)
A	Absolutely Strong	(7,8,9,9;1,1) (7.2,8.2,8.8,9;0.8,0.8)
S	Slightly Strong	(1,2,4,5;1,1) (1.2,2.2,3.8,4.8;0.8,0.8)
V	Very Strong	(5,6,8,9;1,1) (5.2,6.2,7.8,8.8;0.8,0.8)
F	Fairly Strong	(3,4,6,7;1,1) (3.2,4.2,5.8,6.8;0.8,0.8)

Decision-makers make comparisons to solve the problem using the IT2 FANP method. A fuzzy pairwise comparison matrix between criteria can be seen in Table 4. Da, Db, and Dc denote the comparisons of decision-maker-A, decision-maker-B, and decision-maker-C in Table 4.

**Table 4.** IT2 fuzzy pairwise comparison matrix.

	C1			C2			C3			C4			C5			C6		
	Da	Db	Dc	Da	Db	Dc	Da	Db	Dc	Da	Db	Dc	Da	Db	Dc	Da	Db	Dc
C1	E	E	E	F	S	F	E	S	1/S	S	E	S	E	S	1/S	S	F	F
C2	1/F	1/S	1/F	E	E	E	1/F	1/S	1/F	1/S	1/S	1/S	1/F	1/S	1/F	1/S	S	E
C3	E	1/S	S	F	S	F	E	E	E	S	1/S	F	E	E	E	S	S	F
C4	1/S	E	1/S	S	S	S	1/S	S	1/F	E	E	E	1/S	S	1/F	1/S	F	S
C5	E	1/S	S	F	S	F	E	E	E	S	1/S	F	E	E	E	S	S	F
C6	1/S	1/F	1/F	S	1/S	E	1/S	1/S	1/F	S	1/F	1/S	1/S	1/S	1/F	E	E	E

The geometric mean of the criteria is calculated as in Table 5. According to the Table 5, the ranking of the criteria is found as “C1>C3=C5>C4>C6>C2”.

**Table 5.** The geometric mean of the criteria.

	Geometric mean
C1 (Proximity to Market (Construction Site))	0.69
C2 (Compliance with Legal Conditions)	0.31
C3 (Competition Conditions)	0.64
C4 (Proximity to the Center)	0.43
C5 (Traffic Conditions)	0.64
C6 (Proximity to Logistics Centers)	0.32

Then fuzzy weights for criteria are obtained. Alternatives' local weights are found, and their fuzzy weights are aggregated in Table 6.

**Table 6.** Alternatives' fuzzy weights.

A1												
0	0	1	4	1	1	0	0	1	3	0.7	0.69	
A2												
0	0	0	2	1	1	0	0	0	1	0.7	0.7	
A3												
0	0	1	4	1	1	0	0	1	3	0.7	0.7	
A4												
0	0.1	0.8	4	1	1	0	0.1	0.6	2.7	0.7	0.7	
A5												
0	0	0.3	1.6	1	1	0	0	0.2	1.1	0.69	0.69	
A6												
0	0	0	2	1	1	0	0	0	2	0.69	0.69	

Then, T2 fuzzy numbers are defuzzified using the DTraT method [39], as seen in Table 7. Defuzzification of the alternative weights is carried out for both the inner and the outer dependencies, and the results of this defuzzification are shown in Table 7 below.

**Table 7.** Outcomes of the application using IT2 FANP method.

	Fuzzy weights	Normalized values
A1	1.12	24.34%
A2	0.47	10.20%
A3	0.98	21.29%
A4	1.02	22.13%

A5	0.39	8.57%
A6	0.62	13.48%

Alternatives are ranked and fuzzy weights are found as 1.12, 0.47, 0.98, 1.02, 0.39 and 0.62 in Table 7. Moreover, normalized values are 24.34, 10.20, 21.29, 22.13, 8.57 and 13.48 in Table 7. When the outcomes are analyzed, the ranking is found as "A1>A4>A3>A6>A2>A5". Furthermore, it would be said that choosing A1 is the most relevant result.

### 3. Conclusions

Concrete, which has a wide usage area in today's construction sector, is the most used material after water. In addition, the use of ready-mixed concrete has increased considerably in recent years.

It is imperative that the feasible selection of the location of the concrete facilities required to produce ready-mixed concrete used in this sector. The selection of establishment location is one of the most basic and critical strategic decisions that should be considered during the establishment phase of a business. In this study, the rankings of the alternative locations were found using the IT2 FANP method according to the criteria and alternative locations determined by the experts. Consequently, using the IT2 FANP methodology has obtained the most appropriate outcome as A1 (Basaksehir) and ranking as "A1 (Basaksehir) > A4 (Ikitelli) > A3 (Bagcilar) > A6 (Zeytinburnu) > A2 (Gungoren) > A5 (Esenler)". The fuzzy weights of A1, A4, and A3 alternatives are high. The fuzzy weights of the A6 and A2 alternatives are medium. The fuzzy weight of the A5 alternative is very low. Alternative A1 (Basaksehir) is the best option for the concrete batching plant site when the decision model considers the dependencies and feedback between the different criteria.

Concerning future research, the problem can be evaluated with other MCDM methods, and more solutions can be studied for selecting alternative concrete plant locations. In addition, intelligent software can be developed to calculate solutions to these problems automatically.

**Yazar Katkıları:** Bu çalışmada tüm yazarlar eşit oranda katkı sağlamıştır.

**Finansman:** Bu araştırma dışarıdan fon almadı.

**Çıkar çatışmaları:** Yazarlar çıkar çatışması beyan etmemektedir.

### References

- [1] Türkiye İstatistik Kurumu , [www.tuik.gov.tr](http://www.tuik.gov.tr)
- [2] TS 12165, 1997. İnşaat Makinaları – Beton Santrali, Türk Standartları Enstitüsü, Ankara.
- [3] [www.thbb.org](http://www.thbb.org), 2007. Türkiye Hazır Beton Birliği web sitesi
- [4] Öztekin, E., Sümer, M., "Hazır betonda santral çıkışı ile şantiye teslim kalitelerinin karşılaştırılması incelenmesi", 3. Ulusal Beton Kongresi, İstanbul, (1994).
- [5] Çil, İ., "Hazır betonda santral çıkışı ile şantiye teslimi kalitelerinin karşılaştırılması incelenmesi ve İstatiksel kalite kontrol", Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, (2000).
- [6] THBB, Dünyada ve Türkiye'de Hazır Beton, Sektöre ilişkin veriler, THBB yayını, İstanbul, (2003).
- [7] Karnik, N.N., Mendel, J.M. and Liang, Q. (1999) Type-2 fuzzy logic systems, *IEEE Transactions on Fuzzy Systems*, 7(6): 643-658.
- [8] Mendel, J.M., John R.I., and F. Liu, F. (2006). Interval type-2 fuzzy logic systems made simple, *IEEE Transactions on Fuzzy Systems*, 14(6): 808-821.
- [9] Chen, S.M. and Lee, L.W. (2010). Fuzzy multiple attributes group decision-making based on the ranking values and the arithmetic operations of interval type-2 fuzzy sets, *Expert Systems With Applications*, 37(1)(2010): 824-833.
- [10] Sola, H.B., Fernandez, J., Hagrass, H., Herrera, F., Pagola, M., and Barrenechea, E. (2014). Interval type-2 fuzzy sets are generalization of interval-valued fuzzy sets: Toward a wider view on their relationship, *IEEE Transactions on Fuzzy Systems*, 23(5): 1876-1882.
- [11] Boran, F.E. and Akay, D. (2013). A generic method for the evaluation of interval type-2 fuzzy linguistic summaries, *IEEE Transactions On Cybernetics*, 44(9): 1632-1645.
- [12] B. Leblebicioğlu and A. Keskin , "Evaluation of Supplier Selection Criteria with Fuzzy DEMATEL Method: An Application on the Pharmacy Industry", *Avrupa Bilim ve Teknoloji Dergisi*, no. 23, pp. 236-242, Apr. 2021, doi:10.31590/ejosat.864116
- [13] Kumar, A., Sah, B., Singh, A. R., Deng, Y., He, X., Kumar, P., & Bansal, R. C. (2017). A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renewable and Sustainable Energy Reviews*, 69, 596-609.
- [14] Ozdemir, Y., Nalbant, K. G., & Basligil, H. (2017). Evaluation of personnel selection criteria using Consistent Fuzzy Preference Relations. *International Journal of Management Science*, 4(6), 76-81.
- [15] Mousavi-Nasab, S. H., & Sotoudeh-Anvari, A. (2017). A comprehensive MCDM-based approach using TOPSIS, COPRAS and DEA as an auxiliary tool for material selection problems. *Materials & Design*, 121, 237-253.
- [16] Özdemir, Y., Nalbant, K. G., & Başlıgil, H. (2018). Personnel selection for promotion using an integrated fuzzy analytic hierarchy process-grey relational analysis methodology: a real case study. *Anadolu University Journal of Science and Technology A-Applied Sciences and Engineering*, 19(2), 278-292.

- 
- [17] Yu, X., Zhang, S., Liao, X., & Qi, X. (2018). ELECTRE methods in prioritized MCDM environment. *Information Sciences*, 424, 301-316.
- [18] Özdemir, Y., & Nalbant, K. G. (2018). A real personnel selection problem using the generalized choquet integral methodology. *Business & Management Studies: An International Journal*, 6(2), 694-716.
- [19] Nalbant, K. G., & Ozdemir, Y. (2018). Personnel selection using fuzzy VIKOR methodology. *International Journal of Management Science*, 5(2), 10-17.
- [20] Lin, M., Huang, C., & Xu, Z. (2020). MULTIMOORA based MCDM model for site selection of car sharing station under picture fuzzy environment. *Sustainable cities and society*, 53, 101873.
- [21] Ozdemir, Y., & Nalbant, K. G. (2020). Personnel selection for promotion using an integrated consistent fuzzy preference relations-fuzzy analytic hierarchy process methodology: A real case study. *Asian Journal of Interdisciplinary Research*, 3(1), 219-236.
- [22] Chowdhury, P., & Paul, S. K. (2020). Applications of MCDM methods in research on corporate sustainability: A systematic literature review. *Management of Environmental Quality: An International Journal*.
- [23] Ozdemir, S., Nalbant, K.G. and Ozdemir, Y. (2021). "Evaluating Campus Components According To The Inclusive Design Principles Using An Integrated Consistent Fuzzy Preference Relations And Grey Relational Analysis Methodology", *Computational Research Progress in Applied Science & Engineering, CRPASE: Transactions of Civil and Environmental Engineering*, 7(3): 1–5, Article ID: 2363, (2021).
- [24] Hezam, I. M., Nayeem, M. K., Foul, A., & Alrasheedi, A. F. (2021). COVID-19 Vaccine: A neutrosophic MCDM approach for determining the priority groups. *Results in physics*, 20, 103654.
- [25] Nalbant, K.G., Ozdemir, S. and Ozdemir, Y. (2021). Evaluating Inclusive Campus Environment Design Criteria Using CFPR and FANP Methodologies, *WSEAS Transactions on Computer Research*, 9:92-102, (2021), DOI: 10.37394/232018.2021.9.11.
- [26] Tan, T., Mills, G., Papadonikolaki, E., & Liu, Z. (2021). Combining multi-criteria decision making (MCDM) methods with building information modelling (BIM): A review. *Automation in Construction*, 121, 103451.
- [27] S. Senturk, Y. Binici and N. Erginel, The theoretical structure of fuzzy analytic network process (FANP) with interval type-2 fuzzy sets, *IFAC-PapersOnLine*, 49(12)(2016): 1318-1322.
- [28] T. Wu and X. Liu, An interval type-2 fuzzy ANP approach to evaluate enterprise technological innovation ability, *Kybernetes* 45(9)(2016): 1486-1500.
- [29] S. Senturk, N. Erginel and Y. Binici, Interval Type-2 Fuzzy analytic network process for modelling a third-party logistics (3PL) company, *Journal of Multiple-Valued Logic & Soft Computing*, 28(2017).
- [30] Ozdemir, Y., Ozdemir, S., & Nalbant, K. G. (2021). A Hybrid Methodology for Prioritizing of Store Plan Alternatives Produced with Rule-Based Design. *International Journal of Information Technology & Decision Making*, 20(06), 1685-1709.
- [31] Ozdemir, Ş., Nalbant, K.G., Ozdemir, Y., "An Interval Type-2 Fuzzy Analytic Network Process For Prioritizing of Store Plan Alternatives Produced With Ruled Based Design", *CRPASE: Transactions of Applied Sciences*, 7(4):1-5, (2021).
- [32] Şimşek, B., İc, Y. T., & Şimşek, E. H. (2013). A TOPSIS-based Taguchi optimization to determine optimal mixture proportions of the high strength self-compacting concrete. *Chemometrics and Intelligent Laboratory Systems*, 125, 18-32.
- [33] Şimşek, B., & İc, Y. T. (2020). Fuzzy failure mode and effect analysis application to reduce risk level in a ready-mixed concrete plant: A fuzzy rule based system modelling approach. *Mathematics and Computers in Simulation*, 178, 549-587.
- [34] Abdelgawad, M., & Fayek, A. R. (2010). Risk management in the construction industry using combined fuzzy FMEA and fuzzy AHP. *Journal of Construction Engineering and management*, 136(9), 1028-1036.
- [35] Debnath, J., & Biswas, A. (2018). Assessment of occupational risks in construction sites using interval type-2 fuzzy analytic hierarchy process. In *Industry Interactive Innovations in Science, Engineering and Technology* (pp. 283-297). Springer, Singapore.
- [36] Kou, G., Olgu Akdeniz, Ö., Dinçer, H., & Yüksel, S. (2021). Fintech investments in European banks: a hybrid IT2 fuzzy multidimensional decision-making approach. *Financial Innovation*, 7(1), 1-28.
- [37] Bera, A. K., Jana, D. K., Banerjee, D., & Nandy, T. (2020). Supplier selection using extended IT2 fuzzy TOPSIS and IT2 fuzzy MOORA considering subjective and objective factors. *Soft Computing*, 24(12), 8899-8915.
- [38] Zhou, P., Zhou, P., Yüksel, S., Dinçer, H., & Uluer, G. S. (2019). Balanced scorecard-based evaluation of sustainable energy investment projects with it2 fuzzy hybrid decision making approach. *Energies*, 13(1), 82.
- [39] Liu, J., Lv, J., Dinçer, H., Yüksel, S., & Karakuş, H. (2021). Selection of renewable energy alternatives for green blockchain investments: A hybrid IT2-based fuzzy modelling. *Archives of Computational Methods in Engineering*, 28(5), 3687-3701.
- [40] Kahraman, C., Oztaysi, B., Sari I. U. & Turanoglu, E. (2014). Fuzzy analytic hierarchy process with interval type-2 fuzzy sets, *Knowledge-Based Systems*, 59, 48-57.

Research Article

Received: date:20.04.2022  
Accepted: date:26.06.2022  
Published: date:30.06.2022

# Medikal Sentetik Veri Üretimiyle Veri Dengelemesi

A.Fatih Deveci<sup>1</sup>, M. Fevzi Esen<sup>2,\*</sup>

<sup>1</sup> Tekhnelogos Araştırma ve Geliştirme Merkezi, İstanbul, Türkiye; fatih.deveci@tekhnelogos.com

<sup>2</sup> Sağlık Bilimleri Üniversitesi, Sağlık Bilişim Sistemleri Ana Bilim Dalı; fevzi.esen@sbu.edu.tr

Orcid: 0000-0002-3044-8397<sup>1</sup> Orcid: 0000-0001-7823-0883<sup>2</sup>

\* Correspondence Author, e-mail: fevzi.esen@sbu.edu.tr

**Öz:** Sağlık hizmetleri planlaması, klinik deneyler ve araştırma geliştirme çalışmaları gibi sağlık verisi kullanımını gerektiren alanlarda, kişisel sağlık verisinin elde edilmesi ve kullanımında etik, bürokratik ve operasyonel zorluklar yaşanmaktadır. Elektronik kişisel sağlık kayıtlarının güvenliği ve kişisel veri mahremiyeti konularındaki kısıtlamalar başta olmak üzere, klinik ve saha çalışmalarından veri elde edilmesinin maliyetli ve zaman alıcı olması, gerçek veriye en yakın şekilde yapay veri üretilmesini gerekli kılmaktadır. Bu çalışmada, son dönemde sağlık alanında artan veri kullanımı ihtiyacı doğrultusunda, sentetik veri kullanımının önemi ele alınarak, sentetik veri üretiminde kullanılan SMOTE, SMOTEENN, BorderlineSMOTE, SMOTETomek ve ADASYN yöntemlerinin performanslarının karşılaştırılması amaçlanmıştır. Çalışmada, gözlem ve sınıf sayısı birbirinden farklı ve ikisi de kamuya açık, 390 hastaya ait 15 değişkenden oluşan veri seti ile 19.212 COVID-19 hastasına ilişkin 16 değişkenden oluşan veri seti kullanılmıştır. Çalışma sonucunda SMOTE tekniğinin gözlem ve sınıf sayısının fazla olduğu veri setini dengelemede daha başarılı olduğu ve sentetik veri üretiminde hibrit tekniklere göre etkin olarak kullanılabileceği sonucuna ulaşılmıştır.

**Keywords:** Sentetik Veri, Smote, Smoteenn, Sağlık Bilişimi

## Data Balancing with Synthetic Medical Data Generation

**Abstract:** There are ethical, bureaucratic and operational difficulties in obtaining and using personal health data in the areas that require the use of sensitive health data such as health care planning, clinical trials and research and development studies. The cost and time consuming of obtaining data from clinical and field studies, especially the restrictions on the security of electronic personal health records and personal data privacy, necessitate the production of synthetic data as close to real data. In this study, it is aimed to compare the performances of SMOTE, SMOTEENN, BorderlineSMOTE, SMOTETomek and ADASYN methods that have been used in synthetic data production by considering the importance of synthetic data generation in line with the increasing need for data use in the health field. In the study, a dataset consisting of 15 variables belonging to 390 patients with different observation and class numbers and a dataset consisting of 16 variables related to 19,212 COVID-19 patients were used. It has been concluded that SMOTE is more successful in balancing the data sets with large number of observations and multiclass classification. This technique can be used effectively in synthetic data generation compared to hybrid techniques.

**Keywords:** Synthetic Data, Smote, Smoteenn, Health Informati

### 1. Giriş

ReportLinker [1] büyük veri ve analitiği raporuna göre son yıllarda büyük veri ve ilişkili piyasalar yıllık ortalama %4 oranında büyüyerek 76,1 milyar dolar düzeyine ulaşmış olup, bu rakamın 2025 yılında yaklaşık 117 milyar dolar seviyesine ulaşması öngörülmektedir. Söz konusu artışta, kurum ve kuruluşların uzaktan çalışma şartlarıyla birlikte araştırma, geliştirme ve pratik deneyimlerden elde edilen veriler ve diğer operasyonları sınırlayıcı önlemlerin sebep olduğu radikal değişikliklerin önemli rol oynadığı belirtilmektedir. Ayrıca, verinin elde edilmesinde kullanılan cihazların çeşitliliği ve miktarındaki artışla birlikte, cihazların nesnelere interneti ve bulut bilişimi teknolojileriyle entegrasyonu ve gerçek zamanlı veri kullanımının karar vericiler için önemli bir ihtiyaç haline gelmesinin, büyük veri analitiği piyasasına ivme kazandırdığı ifade edilmektedir [2]. Aynı raporda, 2024 yılına kadar yapay zeka

ve veri analitiği projelerinde kullanılan verilerin yaklaşık %60'ının sentetik olarak üretileceği belirtilmektedir.

Sağlık alanında kişisel demografik verilerin yanı sıra; klinik bulgular, laboratuvar testleri ve görüntüleme yöntemleri, reçete ve sosyal sigorta kayıtları, halk sağlığının izlenmesi, bireylerin hayat tarzı ve sosyo-ekonomik düzeyi, çevre ve işyeri koşulları, sağlık ihtiyaçları ve bakım hizmetleri gibi sağlığı ilgilendiren geniş bir alanda yüksek hacimli ve karmaşık türde büyük veri birikimi yaşanmaktadır. Elektronik hasta kaydı olarak nitelendirilen bu verilerin, hastalıkların zamanında ve doğru teşhis edilmesi, tedavi planlaması, hastalık seyri hakkında yeni bilgilerin keşfedilmesi, tıbbi araştırma ve geliştirme çalışmalarının gerçekleştirilmesi ve kişiselleştirilmiş (hassas) uygulamaların geliştirilmesindeki artan önemine dikkat çekilmektedir [3]. Buna karşın, birçok sağlık kuruluşu tarafından toplanan elektronik hasta kayıtları değer yaratan bir kaynak olsa da, hasta mahremiyeti endişeleri sebebiyle araştırmacıların çoğu tarafından erişilebilir nitelikte değildir. Araştırma ve geliştirme amacı dahil olmak üzere elektronik tıbbi kayıtlara erişim, gizlilik ve güvenlik kısıtlamaları, verinin üretilmesi – işlenmesi ile ilgili düzenlemeler ve verinin değeri nedeniyle sınırlandırılmaktadır. Ayrıca, araştırmacıların sağlık verisine erişimi her ne kadar mümkün olsa bile; verinin işlenmesi, korunması ve kullanımı konusundaki yasal gerekliliklerin yerine getirilmesi uzun bir süreç gerektirmekte olup bu durum, araştırma sonucundan ve bilginin paylaşılmasından sağlanacak faydayı da ciddi şekilde geciktirmektedir [4]. Bu sebeple, sağlık verilerinin araştırmacıların kullanımına sunulabilmesi amacıyla veriden kişilerin tanımlanabilir özelliklerinin kaldırılması veya gerçek verilere dayalı olarak sentetik veri üretilmesi, yenilikçi ve adaptif uygulamalar olarak göze çarpmaktadır [5].

Avrupa Birliği Genel Veri Koruma Tüzüğü'nün (GDPR) yürürlüğe girmesinden bu yana, bireysel sağlık verisinin araştırma ve geliştirme amaçlı kullanımı dahil olmak üzere, hasta mahremiyetinin sağlanması konularında birçok tartışma yürütülmekte olup, araştırmacıların sağlıkta büyük örneklemle çalışmaları sınırlandırılmaktadır. Verilerin korunması ve veri ihlallerinin bildirilmesine ilişkin faydalarının yanı sıra; düzenleme kapsamında kişisel düzeydeki tıbbi verilere erişimin engellenmesi veya sınırlandırılması, birçok yenilikçi araştırma, geliştirme ve eğitim fırsatına da engel olmaktadır [5]. Bu durum, bilgi teknolojileri araçlarının etkin şekilde kullanılarak, kişisel sağlık verilerinin yerini tutabilecek nitelikte alternatif veri kaynaklarının oluşturulmasını gerekli kılmaktadır.

Kişisel verilerin tanımlanmasını önlemeye yönelik olarak, veri setinden kişiyi tanımlayıcı bilgilerin kaldırılmasını amaçlayan veri maskeleyme, veri birleştirme, bölümlenme ve veri anonimleştirme gibi birçok teknikten yararlanılmaktadır. Veri setindeki değişkenlerin içeriği ve hassasiyetini dikkate alan söz konusu tekniklerin uygulanması sonucunda orijinal veri setinden çıkartılabilecek fayda azalmakta ve belirli ölçüde bilgi kaybı yaşanmaktadır. Uygulama adımlarında karşılaşılan prosedürel güçlükler, işlemlerin maliyetli ve zaman alıcı olması, hasta verilerinin yeniden tanımlanması ve geniş veri ağlarında paylaşılması riski ve kurumların hasta mahremiyeti – bilginin mülkiyeti endişeleri, gerçek veriye dayalı sentetik veri üretim tekniklerini popüler hale getirmektedir [6,26]. Bu sebeple, verinin gizliliği ve güvenliğini sağlamak amacıyla gerçek veri setindeki özelliklere, dağılımlara ve ilişkilere sahip, gerçek veriden herhangi bir ipucu içermeyen ve verinin tanımlanması riskinin en aza indirildiği sentetik veri üretimi iyi bir alternatif olarak kabul edilmektedir [7].

Sentetik veri, güvenlik - gizlilik riski olmayan, gerçek verilerle ilişkili yapay veri olup, verinin olmadığı veya elde edilmesinin zor ve maliyetli olduğu durumlarda hacim ve çeşitlilik sorunu olmadan belirli ihtiyaca veya koşula yönelik üretilebilmektedir. Sentetik veri, gerçek veriye ulaşımın olmadığı durumlarda modelleme ve simülasyon gibi bütünleşik yöntemlerle üretilebildiği gibi, makine öğrenmesi teknikleriyle eğitilmiş modeller kullanılarak da üretilmektedir [8]. Bu durum, araştırmacıların kendi amaçları doğrultusunda gerçek verilerin yerini alabilecek, keşfedici veya doğrulayıcı nitelikte ve etik endişeler içermeyen sentetik veri kullanımını veri anonimleştirmeye iyi bir alternatif kılmaktadır[9]. Sentetik verinin üretilmesi sırasında, silme, yok etme, gürültü ekleme, genelleştirme ve baskılama gibi anonimleştirme tekniklerinden de yararlanılmakta, veri içerisinde kişiyi tarif edici nitelikler değiştirilmekte veya kaldırılmaktadır. Sentetik veri, orijinal verilerle aynı istatistiksel özellikleri ve zamana bağlı özellikleri içermekte olup, gerçek kişilerle eşleştirilmemiş, tam ve geri dönüşü olmayan bir anonimleştirme sağlamaktadır[10].



Sağlık çalışmalarında kişisel veri setlerine ulaşımındaki zorluk ve verinin işlenmesi süreçlerinin çoğunlukla kısıtlanmış olması, bilimsel tekrarlanabilirlik sorununu ortaya çıkarmaktadır. Bunun yanı sıra; özellikle sağlık bilimlerinde olgu grubundaki gözlem sayısının toplam gözlem içerisindeki oranının düşük olması ve olgulara ait elde edilen verinin yetersiz kalması, kullanılacak istatistiksel modelin tahmin doğruluğu da düşürmektedir. Dengesiz veri olarak adlandırılan bu durum, model başarısının düşmesi, aşırı veya yetersiz uyum gibi problemlere de sebep olmaktadır. Örneğin, halk sağlığı planlamasına yönelik olarak, kişilerin beslenme ve sağlık durumuna ilişkin yeterli veriye ulaşılamadığında, verinin eksik/gürültülü olduğu durumda veya nadir görülen hastalık araştırmalarında bir sınıfa ait daha fazla gözlem elde edilirken diğer sınıfa ilişkin kısıtlı sayıda gözleme ulaşıldığında, sınıf dengesizliğinden bahsetmek mümkündür. Bu sebeple, dengesizliğin azaltılması veya ortadan kaldırılmasına yönelik olarak literatürde çeşitli sentetik veri üretimi teknikleri kullanımı önerilmektedir [7].

Bu çalışmada, literatürde belirtilen [11,12] metodolojiden hareketle, veri çoğaltım ve hibrit yaklaşımlara dayalı olan SMOTE, SMOTEENN, BorderlineSMOTE, SMOTETomek ve ADASYN tekniklerinin, iki ve çok sınıflı veri setlerinde sınıf dengesizliklerinin giderilmesindeki performansları ve dolayısıyla sentetik veri üretiminin kıyaslanması amaçlanmıştır. Bu kapsamda, gözlem ve sınıf sayısı birbirinden farklı iki sağlık veri seti kullanılarak orijinal veri seti sınıf dağılımları ile söz konusu yöntemler uygulandıktan sonra oluşan sınıf dağılımları karşılaştırılmıştır.

## 2. Sentetik Veri Üretim Süreci

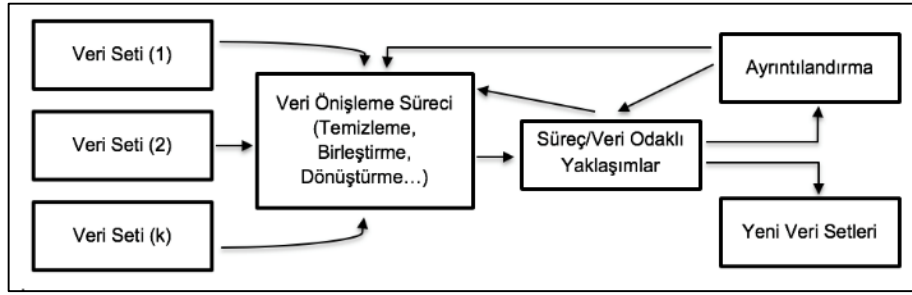
Sentetik veri üretimi, kişiyi tanımlayan ve hassas nitelikte olan verilerin prototipini sağlamaktadır. Böylelikle araştırmacıların veri gizliliği ile ilgili düzenlemeleri ihlali riskine maruz kalmasının engellenmesi ve araştırmaların zengin, geniş ve kontrol edilebilir bir veri alanı ile test edilmesine de imkan tanınmaktadır. Ayrıca sentetik veriyle, araştırmacıların gerçek tıbbi kayıt kullanımına gerek kalmadan, problemlerin çözümünde eğitilebilir modeller kullanabilmesine, çalışma sonuçlarında hassasiyetin artırılması, istenilen tür/boyutta veri üretilmesi ve daha fazla senaryo ile çalışılması sağlanmaktadır.

Veri bilimi ve makine öğrenmesinde yeni çözümler ve hype döngüsü raporuna göre sentetik veri aşağıda belirtilen durumlarda kullanılmaktadır [14]:

- Kişisel ve hassas nitelikteki verilerin mahremiyeti yani, gizlilik gereksinimleri veri erişilebilirliğini veya nasıl kullanılabileceği sınırlandırıldığında, veri merkezli karar destek sistemlerinin tasarlanması sürecinde,
- Mevcut veri setinin homojen olduğu ve heterojen veri setlerine ihtiyaç duyulduğunda,
- İhtiyaç duyulan verinin olmaması ya da erişilebilir veri seti bulunmadığında,
- Geçerli veri setleri olsa da, düzenlenmemiş farklı veri setlerinin birleştirilmesi veya bir araya getirilmesine ihtiyaç olduğu durumlarda,
- Gerçek verinin test edilmesi ve doğrulanmasında,
- Makine öğrenimi algoritmaları için eğitim veri setinin üretiminin çok pahalı olduğu durumlarda,
- Veri gizliliği korunurken, veriye erişimin en üst düzeye çıkarılması durumunda,
- Yazılımların test edilmesi ve doğrulanmasında,
- Akademik araştırma ve eğitim amaçlı olarak.

Sentetik verilerin üretiminde birçok farklı yaklaşım kullanılmaktadır [15]. Şekil 1'de gösterildiği üzere söz konusu yaklaşımlardaki ortak noktalar, farklı kaynaklardan elde edilen yapılandırılmış/yapılandırılmamış verilerin ön işleme sürecinden geçirilerek kullanılacak sayısal yaklaşıma göre esnek ve genelleştirilebilir adımların oluşturulabilmesidir.

Eksik verilerin tamamlanmasında kullanılan klasik istatistik yaklaşımların (gürültü ekleme, döndürme, kırpma vb.) yeni kohortların ve hastalık gruplarının oluşturulmasında etkin olarak kullanılamayacağı belirtilmektedir [4]. Ayrıca, görüntü işleme, metin madenciliği ve büyük veri analitiği gibi karmaşık ve zor süreçler içeren konularda klasik istatistik yaklaşımlar yetersiz kaldığı için, benzetim teknikleri ve makine öğrenmesi yaklaşımlarından yararlanılmaktadır.



**Figür 1.** Sentetik Veri Üretim Süreci

Literatürde sentetik veri üretiminde kullanılan yaklaşımlar kabaca ikiye ayrılmaktadır. Süreç odaklı yaklaşımlarda, bir olayın temelini oluşturan süreçlere ilişkin hesaplamalı ve matematiksel modeller kullanılmaktadır. Kesikli olay benzetimi, etmen tabanlı modelleme ve Monte Carlo benzetimi süreç odaklı yaklaşımlara örnek olarak verilebilir. Veri odaklı yaklaşımlarda ise, makine öğrenimi kullanılmakta olup, mevcut veriler öğrenilerek çıkarımlarda bulunmaktadır. Destek vektör makinaları, Bayes ağları ve derin öğrenme bu yöntemlere örnek olarak verilebilir.

Derin öğrenmeye dayalı yöntemler ise, bir varyasyonel otomatik kodlayıcıdan (VAE - Variational Autoencoder) ya da üretici bir rakip ağdan (GAN - Generative Adversarial Network) yararlanmaktadır. VAE'ler, kodlayıcılardan ve kod çözücülerden oluşan denetimsiz makine öğrenimi modelleridir. VAE'nin kodlayıcı kısmı, veriyi orijinal veri setini basit ve kompakt bir şekle sıkıştırmakla sorumlu olup, kod çözücü analiz işlemi ve temel verilerin temsiliyi oluşturmak için kullanılmaktadır. Bir VAE, hem giriş verilerinin hem de çıkış verilerinin benzer olduğu, giriş verileri ve çıkış arasında optimal bir ilişkiye sahip olmak amacıyla eğitilmektedir. GAN modellerinde, jeneratör sentetik verilerin üretilmesinden sorumlu olup, ayırıcı ise üretilen verileri gerçek bir veri setiyle karşılaştırmakta ve hangi verilerin sahte olduğunu belirlemeye çalışmaktadır. Her seferinde oluşturulan ve sahte olanların ayırt edildiği ağlar birbirlerine karşı eğitilerek gerçekçi hale getirilmektedir [17].

**Tablo 1.** Veri Dengeleme Yaklaşımları(Data Balancing Approaches)

Veri Azaltım Yaklaşımları	Veri Çoğaltım Yaklaşımları (Sentetik Veri Üretimi)	Hibrit Yaklaşımlar (Sentetik Veri Üretimi & Veri Azaltım)
Random Undersampling	Random Oversampling	SMOTETomek
Condensed Nearest Neighbor Rule (CNN)	Synthetic Minority Oversampling Technique (SMOTE)	SMOTEENN
Near Miss Undersampling	Borderline-SMOTE	-
Tomek Links Undersampling	Borderline Oversampling with SVM	-
Edited Nearest Neighbors Rule (ENN)	Adaptive Synthetic Sampling (ADASYN)	-
One-Sided Selection (OSS)	-	-
Neighborhood Cleaning Rule (NCR)	-	-

Sentetik veri üretimi yaklaşımlarının temelinde, örneklem azaltma (under sampling), örneklem artırma (over sampling) ve hibrit örnekleme (hybrid sampling) gibi üç temel veri dengeleme yöntemi bulunmaktadır (bkz. Tablo 1) [18]. Çoğunluk sınıfına ait veri sayısının azınlık sınıfı veri sayısına çekildiği örneklem azaltma yönteminde veri seti dengeli hale getirilmeye çalışılmaktadır. Örneklem arttırmada azınlık sınıfına ait veri sayısı çoğunluk sınıfına ait veri sayısına yaklaştırılmakta olup, yeni sentetik veriler üretilerek azınlık sınıfına eklenmektedir. Hibrit yaklaşımda ise, hem çoğunluk sınıfından örneklem azaltma yapılmakta, hem de azınlık sınıfına ait örneklerin dağılıma uygun olarak örneklem artırılması yapılarak her iki yöntemden yararlanılmaktadır.

### 3. Literatür Araştırması

Sağlıkta sentetik veri üretimi konusundaki çalışmalar geniş bir uygulama alanına sahip olup, tıbbi görüntüleme başta olmak üzere kanser araştırmaları, gen dizileme, biyogözetim, ilaç geliştirme ve farmakovijilans alanında yoğunlaşmaktadır.

Gerçekleştirilen çalışmalarda sentetik veri üretiminin yanı sıra; araştırmacıların klinik kararlarını destekleyici nitelikte ve kişisel verilerin korunması ilkeleri çerçevesinde bilgi güvenliğiyle uyumlu iş akışı sağlayan sistem modelleri üstünde durulmaktadır. Örneğin, gerçek elektronik tıbbi kayıtlardan sentetik kayıtlar oluşturması amacıyla gerçekleştirilen Sentetik Elektronik Tıbbi Kayıt Oluşturucu (EMERGE) çalışmasında, salgın ve acil durumlarda kayıt kontrolü ve tarihsel veri doğrulama işlemlerinin gerçekleştirilmesi amaçlanmaktadır. Standartlaştırılmış bir test veri setinin oluşturulması farklı algoritmaların ve prosedürlerin karşılaştırılmasına izin vermenin yanı sıra, yenilikçi ve adaptif algoritmaların geliştirilmesi için de önemli bir veri kaynağı niteliği taşımaktadır [10]. Gözlemsel Tıbbi Veri Kümesi Simülatörü (OSIM) çalışmasında ise zaman, cinsiyet ve yaş gibi demografik değişkenler ele alınarak, gerçek verilerin olasılık dağılımlarına uygun hastalık sınıflarının oluşturulması ve reçete edilebilecek ilaçlar konusunda sentetik veri sentezi sunulmaktadır [19]. Bir diğer çalışmada, 823 kardiyovasküler hastadan oluşan veri seti üzerinde düşük ve yüksek risk sınıflaması vakaları ele alınmış, iki sınıf arasındaki dengesizliğin farklı tekniklerle giderilmesi amacıyla kullanılacak teknikler üzerinde değerlendirmelerde bulunulmuştur [20]. Çalışma sonucunda SMOTE uyarlanmış veri azaltma yaklaşımının sentetik veri üretiminde etkin olarak kullanılacağı gösterilmiştir. 13-24 yaş aralığında 132 hastaya ait sol el MR görüntüsü ve 20 farklı hastaya ait bilgisayarlı tomografi görüntülerinin kullanıldığı medikal görüntüleme çalışmasında, Evrişimli Sinir Ağı tabanlı algoritmalar ve bilgisayarlı görü teknikleri kullanılmıştır. Çalışmada, üretilen sentetik görüntülerle vücut pozunu tahminleme gerçekleştirilmiştir [21].

Farklı hastalık sınıfları ve örneklem sayılarına sahip medikal veri setleri üzerine yapılan bir diğer çalışmada, Destek Vektör Makinaları, k-En Yakın Komşu yöntemi ve Çok Katmanlı Algılayıcılar algoritmaları kullanılarak toplamda 2471 hasta kaydını içeren veri setinin dengelenmesi amaçlanmıştır. Çalışmada, sınıflama algoritmalarının performansları karşılaştırılmıştır [22]. Optimizasyon, istatistik ve makine öğrenme tekniklerinin sentetik zaman serisi oluşturmadaki etkinliği üzerine gerçekleştirilen çalışmada ise 42.240 hasta kaydı içeren AreM (Activity Recognition system based on Multisensor data) veri seti ve 122 kayıttan oluşan EEG verileri ele alınmıştır. Çalışma sonucunda sentetik veri setinin orijinal veri setindeki öznelikleri koruduğu tespit edilmiştir [23]. Açık kaynak kodlu Synthea (Sentetik Sağlık) üreticisinde ise, algoritma testi, araştırma sonuçları doğrulama, güvenlik ve gizlilik kontrolü, fizibilite analizi ve diğer akademik araştırmacılar için sentetik boylamsal veri oluşturulması mümkün kılınmakta olup, klinik keşif ve bilimsel çıkarım yönünden araştırmacılara karar desteği sağlanmaktadır. Karaciğer lezyonu bilgisayarlı tomografi görüntüleri kullanılarak Çekişmeli Üretici Ağlarıyla (GANs) sentetik medikal görüntü üretilmesini amaçlayan bir diğer çalışmada, Evrişimli Sinir Ağı (CNN) ile sınıflama modeli tahmin başarısının artırılması sağlanmıştır. Buna göre klasik sınıflandırma yaklaşımlarıyla %78,6 duyarlılık ve %88,4 özgüllük sağlanırken, sentetik veri artırma ile %85,7 duyarlılık ve %92,4 özgüllük sağlanmıştır. Çalışma sonucunda, sentetik veri üretmeye yönelik bu yaklaşımın diğer tıbbi sınıflandırma uygulamalarına genelleştirilebileceği ve böylelikle radyologların tanıyı iyileştirme çabalarının desteklenebileceği gösterilmiştir [24]. Otomatik kodlayıcılar ve üretken çekişmeli ağların bir kombinasyonuyla yüksek boyutlu ayrışık örneklem oluşturmayı amaçlayan medGAN sisteminde ise, girdi olarak gerçek hasta kayıtları kullanılmakta olup, epidemiyoloji alanında yüksek kaliteli, çok boyutlu ve gerçeğe yakın sentetik veri üretimi sağlanmaktadır [25].

Bir başka çalışmada, 2010 ve 2015 yılları arasında teşhis edilen meme, solunum ve katı olmayan kanser vakalarından oluşan 360.000 hastaya ilişkin veri setini kullanarak olasılık modelleri, sınıflandırmaya dayalı isnat modelleri ve çekişmeli üretken ağ olmak üzere üç sınıf sentetik veri üretme yaklaşımı değerlendirilmiştir [4]. Çalışmada, üretilen sentetik veri kümelerinin kalitesini değerlendirmeye yönelik ölçütler sunulmakta olup, kategorik gizli Gauss sürecinin gerçek veri üretiminde daha başarılı olduğu sonucuna ulaşılmıştır. 1042 farklı muayenehaneden elde edilen 27,5 milyon birinci basamak hastasına ilişkin veriden hareketle yapılan bir diğer çalışmada, hastaların sentetik veriyle yeniden tanımlanması

riskini ölçerek, makine öğrenmesi sınıflandırıcılarından elde edilen duyarlılık analizi sonuçlarını ele almıştır [7]. Ayrıca, grafik modelleme ile adapte edilen aykırı değer analizi de gerçekleştirilmiş olup, orijinal veriden istatistiksel olarak anlamlı derecede farklı olmayan sentetik veri kümelerinin tespit edilmesi amaçlanmıştır. Çalışma sonucunda, hastaları tanımlayıcı nitelikte veri üretme riskinin çok düşük olduğu belirtilmiştir. Acil servis hastaları üzerine gerçekleştirilen bir başka çalışmada, 27.963 acil servis taburcu verisi ele alınmış olup, değişkenlere ilişkin gizli olasılık dağılımlarının tespiti Kullback-Liebler uzaklığı ile gerçekleştirilmiştir. Bireysel gizliliği koruyan Elektronik Sağlık Kayıtları (EHR) oluşturmak için Gretel'in açık kaynaklı, sentetik veri kütüphanesinden faydalanılmış olup, gerçek veri setiyle uyumlu sentetik veri setlerinin oluşturulabileceği sonucuna ulaşılmıştır [8].

#### 4. Veri Seti ve Yöntem

Çalışmada, sentetik veri üretiminde kullanılan SMOTE, SMOTEENN, BorderlineSMOTE, SMOTETomek ve ADASYN yöntemlerinin performanslarının karşılaştırılması amacıyla, gözlem ve sınıf sayısı birbirinden farklı iki veri seti kullanılmıştır. İlk veri setinde 390 hastaya ait 15 değişkenden oluşan diyabet hastalığı teşhisleri, ikinci veri setinde ise 19.212 COVID-19 hastasına ilişkin 16 kesikli ve sürekli değişkenden oluşan hastane yatış ve sosyodemografik veriler bulunmaktadır. Literatürde önerilen, azınlık sınıfı üzerinden yeni gözlemler üreterek azınlık sınıfı gözlem sayısını arttırmaya yönelik metodolojiden hareketle, sentetik veri üretimi için veri çoğaltım ve hibrit yaklaşımlar kullanılmıştır[38]. Söz konusu yaklaşımların ortak özelliği, çoğunluk sınıfının aşırı öğrenilmesini önlemek için azınlık sınıfından sentetik numuneler oluşturulmasıdır. Ayrıca, aşırı örneklemenin Rastgele Veri Çoğaltım veya diğer adıyla Rastgele Aşırı Örnekleme (random oversampling) metoduyla yapılmasının en basit yaklaşım olduğu belirtilen çalışmadan yola çıkarak [20], bu çalışmada ilk olarak, veri çoğaltım (oversampling) ve sonrasında örnekleme azaltımı (undersampling) da içeren hibrit teknikler kullanılmıştır. Gürültülü ve eksik gözlemler veri setinden çıkartılarak ve herhangi bir normalizasyon uygulanmadan veri analize hazır hale getirilmiş, sınıflar arası heterojen yapı giderilmiştir.

#### 5. Bulgular

Çalışmada kullanılan diyabet teşhislerine dağılım frekansları ve sınıflar arası dengesizlik oranı Tablo 3'de verilmiştir. Diyabet teşhisi iki sınıflı bir değişken olarak (0=diyabet tanısı yok, 1= diyabet tanısı var) tanımlanmıştır. Orijinal veri setinde tüm vakaların %84,61'i diyabet tanısı olmayanlardan oluşmaktadır. Sınıflar arası dengesizliği oranı ise diyabetli olmayan hasta sayısının diyabetli hasta sayısına oranlanmasıyla elde edilmiştir. Buna göre uygulama öncesi orijinal veri setinde sınıf dengesizlik oranı %81,81 olarak bulunmuştur.

**Tablo 3.** İki sınıflı diyabet Veri Seti Analiz Sonuçları

Yöntem	Orijinal Veri Seti-1 Sınıf Dağılımı ve Yüzde		Dengesizlik Oranı	Yöntem Uygulandıktan Sonra Sınıf Dağılımı ve Yüzde		Yöntem Uygulandıktan Sonra Dengesizlik Oranı
	Diyabet Tanısı: Yok	Diyabet Tanısı: Var		Diyabet Tanısı: Yok	Diyabet Tanısı: Var	
SMOTE	330	60	% 81,81	330	330	%0,00
SMOTEENN	330	60	% 81,81	262	312	%16,02
ADASYN	330	60	% 81,81	330	325	%1,01
BorderlineSMOTE	330	60	% 81,81	330	330	%0,00
SMOTETomek	330	60	% 81,81	325	325	%0,00

Sentetik veri üretme ve hibrit yaklaşımlar uygulandıktan sonra elde edilen sınıf dengesizliği oranına bakıldığında, SMOTE, Borderline-SMOTE ve SMOTETomek uygulamaları sonrasında dengesizliğin giderildiği sonucuna ulaşılmıştır. SMOTETomek yaklaşımında, çoğunluk sınıfı gözlem sayısından veri azaltılarak, azınlık sınıfına sentetik veri eklenerek dengeleme gerçekleştirilirken, borderlineSMOTE ve SMOTE yöntemlerinde ise azınlık sınıfı için sentetik veri üretilmiştir.

Tablo 4’de 19.212 COVID-19 vakasından oluşan dengesiz sınıf dağılımına sahip veri setinde ise hastaların çalışma statüleri 10 farklı sınıf şeklinde ifade edilmiştir. Örneğin COVID-19 hastaları içerisinde tam zamanlı çalışan sayısı 4595 kişi, engelli 4380 kişi ve asker/ordu görevi yürüten toplamda iki kişi görülmektedir. Uygulama sonrası sonuçlara bakıldığında, SMOTE veri çoğaltımı yönteminde toplamda 54,640 yeni vaka üretildiği ve her bir sınıf yüzdesinin eşitlendiği görülmektedir. Aynı şekilde, Borderline-SMOTE uygulaması sonrasında her bir sınıf orijinal veri setindeki çoğunluk sınıfındaki vaka sayısına eşitlenmiş olup, sadece son sınıfta (asker/ordu görevi olanlar) orijinal veri setindeki vaka sayısı sabit kalmıştır. SMOTETomek yönteminde ise sınıf frekanslarının birbirine yaklaşmış olduğu izlenirse de, eşitliğin sağlanamadığı görülmektedir.

Tablo 5’de, iki sınıflı veri setinde uygulanan yöntemlere ilişkin olarak değişkenlerin ortalama değerleri ve çarpıklık katsayıları verilmiştir. Buna göre, orijinal veri seti değerlerine en yakın ortalama ve çarpıklık katsayısı değerlerinin her bir değişken için yöntemlere göre farklılık gösterdiği tespit edilmiştir. Örneğin, orijinal veri setinde 107,33 olarak gözlenen ortalama kan şekeri değerinin ADASYN yöntemiyle üretilen sentetik veri değerine yakın iken, orijinal veri setinde 46,77 ortalama ile gözlenen yaş değişkenine en yakın değerlerin SMOTEENN yöntemiyle oluşturulduğu gözle çarpılmaktadır.

**Tablo 4.** Çok Sınıflı Veri Seti Analiz Sonuçları

Sınıf*	Orijinal Veri Seti Frekans	SMOTE Uyg. Sonrası Frekans	SMOTEENN Uyg. Sonrası Frekans	ADASYN Uyg. Sonrası Frekans	Borderline-SMOTE Uyg. Sonrası Frekans	SMOTETomek Yöntemi Uyg. Sonrası Frekans
1	5464(%28,44)	5464(%10)	345(%1,06)	5464(%22,1)	5464(%11,11)	4848(%10,29)
2	4595(%23,92)	5464(%10)	3683(%11,29)	4380(%17,7)	5464(%11,11)	5303(%11,25)
3	4380(%22,80)	5464(%10)	936(%2,87)	2078(%8,43)	5464(%11,11)	4883(%10,36)
4	2078(%10,82)	5464(%10)	3634(%11,14)	1522(%6,17)	5464(%11,11)	5247(%11,14)
5	1522(%7,92)	5464(%10)	3240(%9,93)	4595(%18,6)	5464(%11,11)	5171(%10,98)
6	610(%3,18)	5464(%10)	5004(%15,34)	610(%2,47)	5464(%11,11)	5393(%11,44)
7	465(%2,42)	5464(%10)	4932(%15,11)	57(%0,23)	5464(%11,11)	5368(%11,39)
8	57(%0,30)	5464(%10)	5412(%16,59)	465(%1,89)	5464(%11,11)	5451(%11,57)
9	39(%0,20)	5464(%10)	5438(%16,67)	5481(%22,2)	5464(%11,11)	5458(%11,58)
10	2(%0,01)	5464(%10)	-	-	2(%0,01)	-

\*Hastaların çalışma statülerini ifade etmektedir. Kodlanan sınıflar: 1- İşsiz, 2- Tam zamanlı, 3- Engelli, 4- Yarı zamanlı, 5- Emekli, 6- Tam zamanlı öğrenci, 7- Bilinmiyor, 8- Serbest meslek, 9- Yarı zamanlı öğrenci, 10- Asker/ordu görevi

**Tablo 5.** İki Sınıflı Veri Seti Temel İstatistikleri

Yöntem	Ort. Yaş (Çarpıklık katsayısı)	Ort. HDL Kolesterol (Çarpıklık katsayısı)	Ort. Glukoz (Çarpıklık katsayısı)	BMI ort. (Çarpıklık katsayısı)	Cinsiyet (mod)	Ort. Boy (Çarpıklık katsayısı)	Ort. Ağırlık (Çarpıklık katsayısı)
Orijinal Veri Seti	46,77 (0,41)	50,26 (0,74)	107,33 (0,96)	28,77 (0,44)	Kadın	65,95 (-0,03)	177,40 (0,32)
SMOTE	51,69 (-0,24)	47,96 (0,55)	140,72 (1,20)	29,65 (0,36)	Kadın	65,80 (-0,16)	182,86 (0,23)

SMOTEENN	50,99 (-0,28)	47,63 (0,57)	148,30 (1,28)	29,38 (0,25)	Kadın	65,92 (-0,05)	182,42 (0,19)
ADASYN	51,22 (-0,34)	48,25 (0,62)	117,71 (0,79)	29,68 (0,70)	Kadın	65,48 (0,40)	181,25 (0,41)
Borderline SMOTE	51,21 (-0,35)	48,88 (0,56)	118,56 (0,54)	29,75 (0,69)	Kadın	65,16 (0,13)	179,83 (0,55)
SMOTE Tomek	51,68 (-0,35)	47,77 (0,51)	143,39 (1,31)	29,43 (0,35)	Kadın	65,88 (-0,09)	182,30 (0,18)

Tablo 6’da çok sınıflı veri setine ilişkin tanımsal istatistiklere bakıldığında, yaş değişkeninin orijinal veri setinde simetriğe yakın hafif sağa çarpık dağılım gösterdiği ve bu duruma en yakın SMOTE tekniğiyle veri üretildiği göze çarpmaktadır. Aynı şekilde, cinsiyet, etnisite, din, medeni durum, sağlık sigortası, hastaların polikliniğe başvuru tarihi (gün ve ay) olarak orijinal veri seti istatistiklerinin SMOTE ile üretilen veri ile uyumlu olduğu tespit edilmiştir.

**Tablo 6.** Çok Sınıflı Veri Seti Temel İstatistikleri

Yöntem	Ort. Yaş (Çarpıklık katsayısı)	Cinsiyet (mod)	Etnisite (mod)	Din (mod)	Medeni durum (mod)	Sağlık sigortası (mod)	Başvuru ay (mod)	Başvuru gün (mod)
Orijinal Veri Seti	50,14 (0,17)	Kadın	Beyaz	Katolik	Bekar	Medicaid	Haziran	Pazartesi
SMOTE	47,99 (0,19)	Kadın	Beyaz	Katolik	Bekar	Medicaid	Haziran	Pazartesi
SMOTEENN	41,91 (0,70)	Kadın	Beyaz	Yok	Bekar	Özel	Mayıs	Cumartesi
ADASYN	44,80 (0,69)	Kadın	Beyaz	Katolik	Bekar	Medicaid	Haziran	Pazartesi
Borderline SMOTE	41.05 (0,32)	Erkek	Beyaz	Katolik	Bekar	Medicaid	Nisan	Cumartesi
SMOTE Tomek	42,67 (0,56)	Kadın	Beyaz	Katolik	Bekar	Özel	Mayıs	Cumartesi

## 6. Sonuç ve Tartışma

Sağlık bilişim alt yapısı ve hasta kayıtlarının saklanma şekli, makine öğrenmesi teknikleri ve yapay zekâ tabanlı sistemlerin ihtiyaç duyduğu nicelikte veriye ulaşımı zorlaştırmaktadır. Veri paylaşımındaki teknik ve bürokratik zorluklarla birlikte, kişisel verilerin gizliliği ve güvenliği sebebiyle gerçek veriye erişimin kısıtlanması da sağlık alanında gerçekleştirilecek bilimsel araştırmaları sınırlandırmaktadır. Gerçek hasta verisine bu denli kısıtlı erişim, araştırmacıların sentetik veri ve ilişkili yaklaşımlara olan ilgisini arttırmaktadır.

Literatürde ikili sınıfları içeren veri setlerinin dengelenmesinde bir çok veri çoğaltımı, azaltımı ve hibrit yöntem önerilmiştir. Sağlık alanında incelenen problemlerin genellikle çok sınıflı dengesiz veri setlerini içermesi, çok sınıflı veri setlerinin dengelenmesinin ikili sınıflara göre güç oluşu ve bu konuda gerçekleştirilmiş çalışmaların az oluşu, çok sınıflı değişkenler içeren veri setlerinin dengelenmesi problemini literatürde sık tartışılan konulardan biri haline getirmektedir [39]. Çalışma sonucunda, iki sınıflı veri setinin dengelenmesinde sentetik veri üretilmesine dayanan SMOTE ve Borderline-SMOTE yöntemlerinin başarılı olduğu, çok sınıflı veri setinin dengelenmesinde SMOTE’un daha iyi bir performansa sahip olduğu tespit edilmiştir. Buna karşın, SMOTE yönteminde her bir sınıfta eşit frekans dağılımı (%10) görülmekteyken; borderline-SMOTE yönteminde orijinal veri setinden sentetik veri üretimi için son sınıfta (asker/ordu görevi olanlar) bulunan vaka sayısı yetersiz kalmış olup, bu sınıf için sentetik veri üretilmemiştir. Bu durum borderline-SMOTE yönteminde azınlık sınıfının parçalara bölünerek veri çoğaltımının sağlanması durumuyla ilgili olup, azınlık sınıfı sayısının çok düşük olduğu durumlarda veri hacminin veri üretimi için yeterli derecede bilgi sağlayamadığına işaret etmektedir [40].

Hibrit yaklaşımlardan olan SMOTETomek algoritmasının ikili sınıf içeren veri setinin dengelenmesindeki etkinliğine karşın, çok sınıflı veri setinin dengelenmesinde aynı durumun gerçekleşmediği; veri çoğaltım yaklaşımlarının sınıf dengesizliğini gidermede daha etkin olduğu görülmektedir. Buna göre, iki ve çok sınıflı veri setleri için sentetik veri üretiminde veri çoğaltım yöntemlerinin hibrit yöntemlere göre sınıflandırma performanslarının belirgin şekilde farklılık gösterdiği tespit edilmiştir.

Hibrit yaklaşımların çok sınıflı veri setlerinin dengelenmesinde kullanımına ilişkin olarak söz konusu yaklaşımların sınıflandırma performanslarının ölçümünde doğruluk, duyarlılık, kesinlik ve seçicilik gibi metriklerin de göz önünde bulundurulmasının faydalı olacağı düşünülmektedir. Ayrıca, çalışmada elde edilen sonuçlar doğrultusunda, düşük sınıflandırma performansı ve veri setinden gözlem çıkartarak bilgi kaybına yol açması nedeniyle, hibrit yöntemlerin sağlık araştırmalarında kullanımının dezavantajlarının da göz önünde bulundurulması gerekmektedir. Ayrıca sağlık gibi çok değişkenli ve çok sınıflı veri setlerinin bulunduğu bir alanda, değişken seçimi yöntemleriyle gürültülü değişkenlerin veri setinden çıkartılması veya çarpık dağılıma sahip değişkenlerin sentetik veri üretimindeki etkilerinin tespit edilmesi ve düzeltilmesine yönelik yaklaşımların kullanılması önerilebilir.

Aynı anda hem azınlık sınıfı için sentetik veri üretilmesi ve çoğunluk sınıfından örnek azaltılması prensibine dayanan hibrit yöntemlerle medikal veri setlerinde daha tutarlı makine öğrenimi modellerinin test edilmesi sağlanabilir. Böylelikle, farklı veri çoğaltım teknikleri ile azınlık sınıfına sentetik veri üretilmesi sağlanıp sonrasında ise, çeşitli veri azaltım teknikleri ile sadeleştirme yapılarak her iki yöntemin avantajlı taraflarından faydalanılabilir. Büyük hacme, çeşitliliğe ve yüksek sınıf dengesizliğine sahip veri setleri için sentetik veri üretiminde kullanılan algoritmaların etkinliği konusu, gelecekte gerçekleştirilebilecek çalışmalar için önemli bir alan olarak görülmektedir.

**Yazar Katkıları:** Bu çalışmada A.D, literatür araştırması ve uygulama kısmının destek verilmesi, M.F.E, çalışmanın yönetilmesi verilerin hazırlanması ve analizlerin yapılması ve kontrol edilmesi; konularında katkı sağlamışlardır.

**Finansman:** Bu araştırma her hangi bir fon desteği almadı.

**Çıkar çatışmaları:** Yazarlar çıkar çatışması olmadığını beyan etmemektedir.

**Not:** Bu çalışma 30.09.2021-01.10.2021 tarihleri arasında düzenlenen II. Tıpta Bilişim Kongresinde özet bildiri olarak sunulan bildirinin genişletilmiş halidir.

## Kaynakça

- [1] ReportLinker (2021). Big Data Industry. <https://www.reportlinker.com/market-report/Advanced-IT/513221/Big-Data,20.07.2021>
- [2] Gartner (2021). Top Strategic Technology Trends for 2021, <https://www.gartner.com/en/publications/top-tech-trends-2021,13.07.2021>
- [3] Jacob, P.D. (2020). Management of patient healthcare information: Healthcare-related information flow, access, and availability, In *Fundamentals of Telemedicine and Telehealth* (ss. 35-57) (Eds. Shashi Gogia), Academic Press.
- [4] Goncalves, A., Ray, P., Soper, B., Stevens, J., Coyle, L., & Sales, A. P. (2020). Generation and evaluation of synthetic patient data. *BMC Medical Research Methodology*, 20(1), 1–40. <https://doi.org/10.1186/s12874-020-00977-1>
- [5] Yale, A., Dash, S., Dutta, R., Guyon, I., Pavao, A., & Bennett, K. P. (2020). Generation and evaluation of privacy preserving synthetic health data. *Neurocomputing*, 416: 244–255. <https://doi.org/10.1016/j.neucom.2019.12.136>
- [6] Rocher, L., Hendrickx, J.M. & de Montjoye, YA. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun*, 10: 3069.
- [7] Tucker, A., Wang, Z., Rotalinti, Y., & Myles, P. (2020). Generating high-fidelity synthetic patient data for assessing machine learning healthcare software. *Npj Digital Medicine*, 3(1). <https://doi.org/10.1038/s41746-020-00353-9>
- [8] Walonoski, J., Klaus, S., Granger, E., Hall, D., Gregorowicz, A., Neyarapally, G., Watson, A., & Eastman, J. (2020). SyntheaTM Novel coronavirus (COVID-19) model and synthetic data set. *Intelligence-Based Medicine*, 1–2: 100007. <https://doi.org/10.1016/j.ibmed.2020.100007>
- [9] Dube, K., Gallagher, T. (2014). Approach and Method for Generating Realistic Synthetic Electronic Healthcare Records for Secondary Use. In: Gibbons J., MacCaull W. eds. *Foundations of Health Information Engineering and Systems*. FHIES 2013. *Lecture Notes in Computer Science*, vol 8315. Berlin, Heidelberg: Springer.
- [10] Buczak, A. L., Babin, S., & Moniz, L. (2010). Data-driven approach for creating synthetic electronic medical records. *BMC medical informatics and decision making*, 10, 59. <https://doi.org/10.1186/1472-6947-10-59>
- [11] Zeng, M., Zou, B., Wei, F., Liu, X., & Wang, L. (2016). Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data. *Proceedings of 2016 IEEE International Conference of Online Analysis and Computing Science, ICOACS 2016, 2016, 225–228*. <https://doi.org/10.1109/ICOACS.2016.7563084>
- [12] Liu, N., Li, X., Qi, E., Xu, M., Li, L., & Gao, B. (2020). A novel ensemble learning paradigm for medical diagnosis with imbalanced data. *IEEE Access*, 8, 171263–171280. <https://doi.org/10.1109/ACCESS.2020.3014362>

- [13] Liu, Y., Li, X., Chen, X., Wang, X., & Li, H. (2020). High-Performance Machine Learning for Large-Scale Data Classification considering Class Imbalance. *Scientific Programming*, 2020. <https://doi.org/10.1155/2020/1953461>
- [14] Gartner (2020). Hype Cycle for Data Science and Machine Learning-2020. <https://www.gartner.com/en/documents/3988118/hype-cycle-for-data-science-and-machine-learning-2020>, 19.07.2021
- [15] Ayala-Rivera, V., Portillo-Dominguez, A. O., Murphy, L., & Thorpe, C. (2016). COCOA: A synthetic data generator for testing anonymization techniques. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9867 LNCS, 163–177. [https://doi.org/10.1007/978-3-319-45381-1\\_13](https://doi.org/10.1007/978-3-319-45381-1_13)
- [16] Marathe, M. V. (2006). Synthetic Data for Data Mining to Support Epidemiological Modeling. *Network Dynamics and Simulation Science Laboratory, Virginia Tech*, 1 Agustus 2021 tarihinde <https://www.cs.dartmouth.edu/~cbk/sdm06/marathe-data.sdm.pdf> adresinden alındı.
- [17] Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models With Big Data. *Frontiers in artificial intelligence*, 3, 4. <https://doi.org/10.3389/frai.2020.00004>
- [18] Bekkar, M., & Alitouche, T. A. (2013). Imbalanced Data Learning Approaches Review. *International Journal of Data Mining & Knowledge Management Process*, 3(4). <https://doi.org/10.5121/ijdkp.2013.3402>
- [19] Murray, R. E., Ryan, P. B., & Reisinger, S. J. (2011). Design and validation of a data simulation model for longitudinal healthcare data. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2011, 1176–1185.
- [20] Rahman, M. M., & Davis, D. N. (2013). Addressing the Class Imbalance Problem in Medical Datasets. *International Journal of Machine Learning and Computing*, May 2014, 224–228. <https://doi.org/10.7763/ijmlc.2013.v3.307>
- [21] Riegler, G., Urschler, M., Ruther, M., Bischof, H., & Stern, D. (2015). Anatomical Landmark Detection in Medical Applications Driven by Synthetic Data. *Proceedings of the IEEE International Conference on Computer Vision, 2015-February*, 85–89. <https://doi.org/10.1109/ICCVW.2015.21>
- [22] Belarouci, S., & Chikh, M. A. (2017). Medical imbalanced data classification. *Advances in Science, Technology and Engineering Systems*, 2(3), 116–124. <https://doi.org/10.25046/aj020316>
- [23] Shamsuddin, R., Maweu, B. M., Li, M., & Prabhakaran, B. (2018). Virtual patient model: An approach for generating synthetic healthcare time series data. *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, February 2019, 208–218*. <https://doi.org/10.1109/ICHI.2018.00031>
- [24] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321, 321–331. <https://doi.org/10.1016/j.neucom.2018.09.013>
- [25] Zhang, Z., Yan, C., Mesa, D. A., Sun, J., & Malin, B. A. (2020). Ensuring electronic medical record simulation through better training, modeling, and evaluation. *Journal of the American Medical Informatics Association*, 27(1). <https://doi.org/10.1093/jamia/ocz161>
- [26] Benaim, A. R., Almog, R., Gorelik, Y., Hochberg, I., Nassar, L., Mashiach, T., Khamaisi, M., Lurie, Y., Azzam, Z. S., Khoury, J., Kurnik, D., & Beyar, R. (2020). Analyzing medical research results based on synthetic data and their relation to real data results: Systematic comparison from five observational studies. *JMIR Medical Informatics*, 8(2), 1–14. <https://doi.org/10.2196/16492>
- [27] Gherardini, M., Mazomenos, E., Mencias, A., & Stoyanov, D. (2020). Catheter segmentation in X-ray fluoroscopy using synthetic data and transfer learning with light U-nets. *Computer Methods and Programs in Biomedicine*, 192, 105420. <https://doi.org/10.1016/j.cmpb.2020.105420>
- [28] Hernandez-Matamoros, A., Fujita, H., & Perez-Meana, H. (2020). A novel approach to create synthetic biomedical signals using BiRNN. *Information Sciences*, 541, 218–241. <https://doi.org/10.1016/j.ins.2020.06.019>
- [29] Shi, G., Wang, J., Qiang, Y., Yang, X., Zhao, J., Hao, R., Yang, W., Du, Q., & Kazihise, N. G. F. (2020). Knowledge-guided synthetic medical image adversarial augmentation for ultrasonography thyroid nodule classification. *Computer Methods and Programs in Biomedicine*, 196, 105611. <https://doi.org/10.1016/j.cmpb.2020.105611>
- [30] Stolfi, P., Valentini, I., Palumbo, M. C., Tieri, P., Grignolio, A., & Castiglione, F. (2020). Potential predictors of type-2 diabetes risk: machine learning, synthetic data and wearable health devices. *BMC Bioinformatics*, 21(17), 1–20. <https://doi.org/10.1186/s12859-020-03763-4>
- [31] Vaden, K. I., Gebregziabher, M., Dyslexia Data Consortium, & Eckert, M. A. (2020). Fully synthetic neuroimaging data for replication and exploration. *NeuroImage*, 223. <https://doi.org/10.1016/j.neuroimage.2020.117284>
- [32] Vilardell, M., Buxó, M., Clèries, R., Martínez, J. M., Garcia, G., Ameijide, A., Font, R., & Civit, S. (2020). Missing data imputation and synthetic data simulation through modeling graphical probabilistic dependencies between variables (ModGraProDep): An application to breast cancer survival. *Artificial Intelligence in Medicine*, 107: 101875. <https://doi.org/10.1016/j.artmed.2020.101875>
- [33] Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., & Pinheiro, P. R. (2020). CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection. *IEEE Access*, 8: 91916–91923. <https://doi.org/10.1109/ACCESS.2020.2994762>
- [34] Dai, F., Song, Y., Si, W., Yang, G., Hu, J., & Wang, X. (2021). Improved CBSO: A distributed fuzzy-based adaptive synthetic oversampling algorithm for imbalanced judicial data. *Information Sciences*, 569, 70–89. <https://doi.org/10.1016/j.ins.2021.04.017>
- [35] Karbhari, Y., Basu, A., Geem, Z. W., Han, G. T., & Sarkar, R. (2021). Generation of synthetic chest X-ray images and detection of COVID-19: A deep learning based approach. *Diagnostics*, 11(5), 1–19. <https://doi.org/10.3390/diagnostics11050895>
- [36] Palmér, E., Karlsson, A., Nordström, F., Petruson, K., Siversson, C., Ljungberg, M., & Sohlén, M. (2021). Synthetic computed tomography data allows for accurate absorbed dose calculations in a magnetic resonance imaging only workflow for head and neck radiotherapy. *Physics and Imaging in Radiation Oncology*, 17(December 2020), 36–42. <https://doi.org/10.1016/j.phro.2020.12.007>



- 
- [37] Vepa, A., Saleem, A., Rakhshan, K., Daneshkhah, A., Sedighi, T., Shohaimi, S., Omar, A., Salari, N., Chatrabgoun, O., Dharmaraj, D., Sami, J., Parekh, S., Ibrahim, M., Raza, M., Kapila, P., & Chakrabarti, P. (2021). Using machine learning algorithms to develop a clinical decision-making tool for covid-19 inpatients. *International Journal of Environmental Research and Public Health*, 18(12), 1–22. <https://doi.org/10.3390/ijerph18126228>
- [38] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). snopes.com: Two-Striped Telamonia Spider. *Journal of Artificial Intelligence Research*, 16(Sept. 28), 321–357. <https://arxiv.org/pdf/1106.1813.pdf%0Ahttp://www.snopes.com/horrors/insects/telamonia.asp>,
- [39] Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N. & Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: an experimental review. *J Big Data*, 7: 70.
- [40] Susan, S. & Kumar, A. (2021). The balancing trick: Optimized sampling of imbalanced datasets—A brief survey of the recent State of the Art. *Engineering Reports*, 3: e12298. <https://doi.org/10.1002/eng2.12298>

Research Article

Received: date: 04.06.2022

Accepted: date: 29.06.2021

Published: date: 30.06.2022

# Yeşil Hastanelerin Gelişimi İçin Stratejik Unsurların Belirlenmesi: Enerji Bazlı Faktörlere Yönelik DEMATEL Yöntemi ile Bir Analiz

Serhat Yüksel<sup>1\*</sup>, Hasan Dinçer<sup>2</sup>, Büşra Çelebi<sup>3</sup>

<sup>1</sup> Istanbul Medipol University, Faculty of Business and Management Sciences, Istanbul, Turkey; serhatyüksel@medipol.edu.tr

<sup>2</sup> Istanbul Medipol University, Faculty of Business and Management Sciences, Istanbul, Turkey; hdincer@medipol.edu.tr

<sup>3</sup> Istanbul Medipol University, Health Management Graduate Student, Istanbul, Turkey; busra.celebi@std.medipol.edu.tr

Orcid: 0000-0002-9858-1266 Orcid: 0000-0002-8072-031X Orcid: 0000-0002-7412-5418

\*Corresponding Author, e-mail: serhatyüksel@medipol.edu.tr

**Öz:** Bu çalışmanın amacı yeşil hastanelerin gelişimi için önem arz eden stratejik unsurların belirlenmesidir. Bu kapsamda, ilk olarak detaylı bir literatür taraması gerçekleştirilmiştir. Yapılan bu incelemeler sayesinde, enerji bazlı dört farklı faktör tespit edilmiştir. Bu kriterler karbon emisyonunun azaltılması, temiz enerji kullanımı, enerji verimliliğinin sağlanması ve düşük enerji tüketen ekipmanların seçilmesi şeklindedir. Bu kriterler arasından en önemlilerinin tespit edilebilmesi için DEMATEL yöntemiyle bir analiz gerçekleştirilmiştir. Sonuç olarak, yeşil hastanelerin gelişebilmesi için en fazla önem arz eden kriter temiz enerji kullanımıdır. Ayrıca, karbon emisyonunun azaltılması da önem arz eden diğer önemli bir faktördür. Elde edilen bu sonuçlardan anlaşılacağı üzere, hastanelerin temiz enerji kullanımına acilen geçmeleri gerekmektedir. Bu süreçte, küçük ölçekli güneş panelleri dikkate alınabilmektedir. Bu paneller sayesinde, hastaneler hem kendi enerjilerini üretebilmekte hem de kullanılan enerji çevre dostu olabilmektedir. Bu bağlamda, yapılacak kapsamlı araştırma ve geliştirme çalışmaları sayesinde, bu enerji türlerinin maliyetlerinin düşürülebilmesi mümkün olabilecektir. Belirtilen bu hususa ek olarak, vergi indirimi ve düşük faizli kredi kullanımı gibi devlet teşvikleri sayesinde, küçük ölçekli güneş panellerinin maliyet avantajı elde edebilmesi mümkün olabilecektir.

**Anahtar Kelimeler:** Yeşil Hastane, Temiz Enerji, DEMATEL

## Identifying Strategic Elements for the Development of Green Hospitals: An Analysis of Energy-Based Factors with the DEMATEL Method

**Abstract:** The aim of this study is to determine the strategic elements that are important for the development of green hospitals. In this context, firstly, a detailed literature review was carried out. Thanks to these investigations, four different energy-based factors were determined. These criteria are reducing carbon emissions, using clean energy, ensuring energy efficiency and choosing low energy-consuming equipment. In order to determine the most important of these criteria, an analysis was carried out with the DEMATEL method. As a result, the most important criterion for the development of green hospitals is the use of clean energy. In addition, reducing carbon emissions is another important factor. As can be understood from these results, hospitals urgently need to switch to the use of clean energy. In this process, small-scale solar panels can be considered. Thanks to these panels, hospitals can both produce their own energy and the energy used can be environmentally friendly. In this context, it will be possible to reduce the costs of these types of energy, thanks to comprehensive research and development studies. In addition to this mentioned issue, it will be possible for small-scale solar panels to gain cost advantages thanks to government incentives such as tax reductions and low-interest loans.

**Keywords:** Green Hospital, Clean Energy, DEMATEL

## 1. Giriş

Sağlık kurumları yoğun bir şekilde kullanılan cihazlar ve sürekli hizmet vermesi gerektiğinden çok miktarda enerji ve su tüketmektedir. Bu yüzden, sürdürülebilir şekilde enerji ve su elde edilmemesi giderek artan çevre sorunlarını oluşturmaktadır. Sağlık kurumlarında hem zararlı hem de zararlı olmayan atıklar oluşmaktadır [1]. Bu atıkların planlı yönetilmemesinden kaynaklı ciddi şekilde hasta ve çalışan odaklı sağlık sorunları ve çevre sorunları ardı ardına gelmektedir. Ayrıca, sağlık hizmetlerinin sunulduğu hastanelerde yatan hastaların ortam kalitesinden etkilenerek iyileşme süreleri etkilenmektedir [2]. Ortam kalitesi görünüm, ses ve hava kalitesi genel olarak etki eden unsurlar olarak belirlenmiştir. Önemli bir sorun olan gürültü kirliliğine hastaların tolerans göstermesi zordur. Çünkü stres seviyesini giderek arttırarak hastayı daha da hastalığını tetiklemektedir. Ek olarak insan sağlığı ve kâr amacı birlikte değerlendirildiğinde sağlık kurumlarında maliyet sorununu da ortaya çıkartmaktadır.

Yüksek enerji tüketimi artması ile ekonomik zararlardan söz edilmektedir. Cihazları çok kullanan hastanelere büyük bir rol düşmektedir. Bu yüzden performansı yüksek cihazlar ve otomasyon kullanımları enerji tasarrufu sağlayarak çevrenin ekonomi üzerinde kurduğu olumsuz baskıyı azaltacaktır [3]. Oluşan atıkların yüzdeleri olarak üretimini sağlık sektörü sağlayan ve enerjiyi en çok yiyen sektör olarak araştırmalar sonucu belirlenmiştir. Tıbbi cihaz ve donanımların yaydığı kirlilik ve radyolojik atıklar insan ve hava ilişkili negatif etkilemektedir [4]. Hastanelerde sağlık personeli diğer çalışan sektördeki çalışanlarla kıyaslandığında hasta bina sendromu, binayla ilişkili hastalık ve çoklu kimyasal duyarlılık fazladır.

Bu çalışmada yeşil hastanelerin gelişimi için önem arz eden stratejik unsurların belirlenmesi hedeflenmiştir. Bu çerçevede, ilk olarak kapsamlı bir literatür taraması gerçekleştirilmiştir. Bu analiz neticesinde enerji bazlı dört farklı faktör tespit edilmiştir. Bu kriterler karbon emisyonunun azaltılması, temiz enerji kullanımı, enerji verimliliğinin sağlanması ve düşük enerji tüketen ekipmanların seçilmesi şeklindedir. Bu kriterler arasından en önemlilerinin belirlenebilmesi için DEMATEL yöntemiyle bir analiz gerçekleştirilmiştir. Bu sayede, sağlık kuruluşlarının “yeşil hastane” olabilmesi için gerekli olan etkin stratejilerin belirlenebilmesi mümkün olabilecektir.

Bu çalışmada beş farklı bölüm bulunmaktadır. Bu bölümün ardından, ikinci bölümde yeşil hastanelerin tanımı, önemi ve avantajları hakkında genel bilgiler verilecektir. Üçüncü bölümde ise yeşil hastane kriterleri açıklanacaktır. Çalışmanın dördüncü bölümü ise enerji bazlı faktörlere yönelik DEMATEL yöntemiyle yapılan analizi içermektedir. Son bölümde ise bulgular ve öneriler tartışılacaktır.

## 2. Yeşil Hastane Kavramı

Yeşil iş insanları belirli bir kurum için çalışırken kaliteli iş standartlarına sahip olması, enerji ve hammadde kullanımlarının minimize edilmesi, sera gazı limitleri olan, atıkların ve oluşan kirliliğin etkili yönetim ile en aza indirgenmesi ve ekosistemi sürdüren ve koruyan olarak tanımlanabilmektedir. Yeşil iş kavramı sağlık sektöründe hastanelere entegre edilmiştir. Çünkü sağlık sisteminin oluşturduğu çevre sorunları çözümsüz hale gelmeden yeşil hastane kavramı ile sorunlar minimize edilmeye çalışılmıştır [5]. Bir hastanenin yeşil hastane unvanına sahip olabilmesi inşaatında ve hizmetinde çevreye verilen zararları en aza indirecek şekilde süreçlerin yönetilmesi ile olabilmektedir. Hastanelerin kendine has politikaları bulunmaktadır. Bu politikalar dâhil edilerek yeşilimsi yöntemler vurgulanarak yeşil hastane kavramı oluşturulmuştur. Mesela hastaneler de toksik maddeleri yok etmek ve sağlıklı çevre imkânı ile sağlık hizmetlerini iyileştirme kavramın ortaya çıkışının gereklilik amacıdır [6].

İnsan merkezi düşünce sisteminden doğa odaklı düşünce sistemi geçiş artan çevre sorunlarıyla artış göstermiştir. Bu yüzden yapı sektöründe sürdürülebilir olması gerekliliği düşüncesi ile yeşil yapılar ortaya çıkmıştır. İnsan ekonomik kazanç elde ederken ve diğer faaliyetlerinde bina yapılarında çok vakit geçirmektedir. Bundan dolayı yapı içinde bulunan insanların sağlığı önem arz etmektedir. Ek olarak insanların yaşamlarını devam ettirebilmeleri için ihtiyaç duyduğu doğal gereksinimleri kaynakların verimli tüketilmesi ile sağlamaktadır. İnsan gününün büyük bir bölümünü iş ve okul gibi alanlarda geçirmektedir [7]. Bu alanlarda verimli olabilmesi amacıyla yeşil yapılar önem kazanmıştır. Yeşil yapı hareketleri bir bakıma yapılan işe engel oluşturmadan çevre dostu bir tasarım oluşturulmasıdır. Çevre dostu tasarım atık ve kirlilik tüketimini azaltması açısından üzerinde durulması gerekmektedir [8]. Yeşil

yapılar yapımı, işletilmesi ve sonlandırılması süreçlerinde hep çevre problemlerini çözmesi bakımından önem arz etmektedir.

Bir yeşil yapı ve normal yapı araştırmalara göre değerlendirildiğinde enerji tüketimi ve kullanım alışkanlığı, karbondioksit emisyonları yayma miktarı, su tüketim oranında, atık üretim miktarı ve bakım maliyetleri konularında farklılıklar göstermektedir. Bu bağlamda, sürdürülebilirlik önem arz ettiği için yeşil yapılar bu tüketim ve farklılıklarda hep çevre dostu yaklaşımlar gösterdiği sonucuna ulaşmıştır. Hastanelerde yüksek performans sağlanması dahilinde enfeksiyon riski önlenebilecektir [9]. Hastanelerde yeşil hastane kavramının önemi değerlendirildiğinde kullanılan kaynakların artması ile maliyet sorunu oluşması, standartların yükselmesi ile sağlık kurumlarının artışı, yeni yapılan hastanelere teşvik oluşturması, hastanelerin enerji ve su gibi kaynakları çok yoğun kullanması ve sağlık örgütlerinin insan haklarından doğan sağlık hakkının öneminin artması gibi sebepler gösterilmektedir [10]. Temel olarak yeşil hastane kavramı çevresel olumsuz etkileri azaltmak, ekonomik açıdan tasarruf sağlanması ve halk sağlığının iyileştirilmesi olarak önemi vurgulanmaktadır. Yeşil hastanelerin çok tükettikleri enerjiyi aza indirmek, su tasarrufu elde etmek, etkili bir atık yönetimi sağlayarak gereksiz atıkları azaltmak, geri dönüşüm prosedürlerini yerine getirmek, satın alma faaliyetinde zararlı madde almamak, yeşilimsi uygulamaları desteklemek, insanlarla olumlu ilişkileri arttırmak, yürüme gibi sağlığa yararlı aktiviteleri arttırmak ve yerel tüketimi desteklemek gibi politikaları olması önemini ortaya çıkarmaktadır [11].

Yeşil hastaneler etkin bir atık yönetimi sağlayarak alternatifler geliştirmektedir. Böylelikle sağlık sisteminden kaynaklı aşırı atık üretimi azaltılması sağlanmaktadır. Yeşil hastaneler çevre dostu faaliyetleri ile hasta ve çalışanların sürdürülebilir şekilde besin ihtiyacını karşılamaktadır. Ayrıca, güvenilir bir çalışma ortamı oluşturmak amacıyla yenilenebilir enerji kaynakları üretim ve tüketimini gerçekleştirmektedir. Aslında enerji ve kaynak taleplerini azaltılmasıyla verimlilik elde edilerek maliyet tasarrufu sağlanmaktadır [12]. Ek olarak, rahat bir ortam oluşturarak hastaların iyileşme sürelerini azaltmakta ve sağlık hizmetlerinin sürdürülebilirliği için sağlık profesyonellerinin performansını arttırmaktadır. Böylelikle hastanelerde hasta çıktılarının olumlu olması sağlanmaktadır. Hastanelerde ciddi şekilde bir kimyasal madde kullanımı vardır.

Bu kimyasal maddelerin zararları minimize edilerek ortamın güvenliği arttırılmaktadır. Bir diğer avantajı etrafındaki kurumlarda ve kişilerde olumlu bir bakış açısı oluşturarak tercih edilir olabilmesidir. Hava kalitesinin iyileştirilmesi ile hasta bina sendromundan kaynaklı sağlık sorunları önlenmektedir [13]. Çevre dostu malzeme kullanımı ve uygulamaları sera gazı salınımı ve küresel ısınmayı azaltmaktadır. Böylece ekolojik sistemin sürdürülebilirliğine katkı sağlayarak bio-çeşitlilik korunmaktadır. Yeşil hastanelerin verdiği avantajları temel olarak beş kategoride toplanabilmektedir. İlk kategori araştırma sonuçlarına göre daha iyi hasta çıktısı ve güvenliği sağlanmasıdır [14,15]. Çünkü yeşil hastaneler gün ışığı alarak hastaların hastanede kalma süresini kısaltmaktadır.

Ek olarak, kaliteli bir havalandırma sistemi sağlandığından dolayı enfeksiyon riskini azaltmaktadır. İkinci kategori personelin güvenliği iyileştirmesidir. Bu bağlamda, kimyasallara maruz kalınma süreleri azaltılırsa araştırma sonuçlarının gösterdiği çıktılara göre sağlık personellerinde görülen hastalıklar azalacaktır [16]. Yeşil hastane kriterlerine sahip hastanelerde hasta ve çalışan memnuniyeti yüksektir. Bunun altında yatan sebep iç ortam kalitesinin doğa ile ilişkili bir görünüme sahip olduğu bulunmuştur. Ayrıca, hastane statüsü ve personelin aidiyet duygusu artarak verimliliği artar [17]. Bir yeşil işletme uzun vadede incelendiğinde işletme maliyeti tasarrufu sağlayacak kriterleri oluşturmaktadır. Son kategori ise yeşil hastanelerde verimlilik artışı olmasıdır. Böylelikle personelin hasta olması veya devamsızlık yapması önlenerek hastanede sunulan hizmet kalitesi artar ve süreklilik sağlanır [18].

## 2. Yeşil Hastane Kriterleri

Atık çevreye bırakılan her türlü zararlı malzemeye denmektedir. Atıklar türlerine göre evsel atık, tıbbi atık, tehlikeli atık ve inşaat atıkları olarak sınıflandırılmaktadır. Evsel atıklar yiyecek atıkları, plastik şişeler gibi malzemeleri kapsamaktadır. Tıbbi atık genellikle kullanılmış ilaç ve tıbbi malzemeleri ve tehlikeli atıklar pil, boya veya çeşitli kimyasallardır [19]. Çevreye büyük zararı dokunan inşaat atıkları ise yıkılan veya ev tadilatında oluşan atıklardır. Atık malzemelerin sınıflandırılmasının yapılması üretilen atıkların etkin yönetimini sağlamaktadır. Etkin bir şekilde atık azaltmak amacıyla yönetilen

süreçlerde dikkat edilmesi gerekenler bir binanın tasarımı, üretim veya hizmetin sunumu ve malzeme alımıdır. Hastanelerin üretilen atıkları mümkün oldukça yok etmek veya azaltmak amacıyla planlar oluşturması gerekmektedir. Bu planlar üretilen atıklardan ısı ve enerji elde etmek veya daha az kaynak kullanmak gibi tasarruf sağlama amacındadır.

Hastaneler birçok atık üretmek yanında tehlikeli atıkları da barındıran bir kurumdur. Bu tehlikeli atıklar patolojik atıklar, bulaşıcı atıklar, metal atıklar, radyoaktif atıklar, nükleer atıklar, ilaç atıkları ve kimyasal atıklardır [20]. Bu atıkların miktarı ülkeden ülkeye, bölgeden bölgeye veya kurumdan kuruma değişmektedir. Bu değişimlerin sebepleri yapılan atık yönetim uygulamaları, hastanenin sunduğu hizmet, hastane çeşidi, mümkün olan geri dönüştürülmüş atıklar, hasta sayısı ve bulunulan yerin ekonomik durumudur. Bu hususlar dikkate alınarak hastanelerin atık yönetim planlaması oluşturması gerekmektedir.

Dünya Sağlık Örgütü gündeminde karbon emisyonu insan sağlığını tetiklediğinden dolayı sağlık kuruluşlarına yönelik planlar geliştirilmesi yer almıştır. Hastaneler kullanılan cihazlar ve sürekli hizmet vermesi gerektiği gibi sebeplerden ötürü enerjiyi çok yoğun kullanan kurumlardır. Bu yüzden temiz ve uygun enerji tüketerek yayılan karbon emisyonunu en aza indirmesi gerekmektedir. Bu bağlamda, hastanelerin kullanacağı enerji ve ısı miktarını azaltmasına yönelik planlar uygulanmalı ve tasarım yapılmalıdır. Ek olarak, sera gazı emisyonu yayan enerji kaynakları yerine sürdürülebilir olan yenilenebilir enerji kaynaklarına yönelim gösterilmesi söylenmektedir. Tıbbi cihazların sızıntı kontrolleri sağlanmalı, anestezi gazlarının ölçülü kullanımı planlanmalı ve oksit gazı yerine hidrojen ve peroksit gazları kullanılması şeklinde alternatifler geliştirilmelidir [21]. Atıkların yok edilmesi için ayrılan alanların kontrolü düzenli olarak sağlanmalıdır.

Yapılan araştırmalara göre diğer işletmeler ile hastaneler kıyaslandığında hastanelerin 2,5 kat daha fazla enerji tükettiği belirlenmiştir. Ek olarak hastaneler diğer ticari işletmelerine göre yedi gün yirmi dört saat hizmet vermesi gerekmektedir. Bu hizmeti yüksek teknolojik tıbbi cihazlar ile gerçekleştirilmektedir. Bu cihazlar sürekli açık kalarak çok enerji tüketmektedir. Öncelikle etkin bir enerji tasarrufu için ısı sızmaları ve kayıplarını tespit etmek, hava filtreleme sistemini verimli kurmak ve ısı enerjisi tek sistemde sağlama planları yapılması gerekmektedir. Enerji kullanılması kontrol edilerek enerjinin en çok harcadığı üniteler tespit edilmelidir. Ek olarak, enerjiyi yoğun tüketen cihaz ve sistemler de belirlenmelidir. Akabinde bu alan ve sistemlerde enerji tasarrufuna yönelik çalışmalar yapılması gerekmektedir [22]. Hastanelerin enerji verimliliği sağlamsındaki en önemli karar mekanizması satın alma birimidir.

Bu yüzden satın alma yöneticileri enerji tasarrufu sağlayan malzeme ve cihaz alma konusunda çalışmalar yapılmalıdır. Hastanelerin karar mekanizmalarının enerji alanında uzmanlardan yardım alması, tasarruf amacıyla uygulamalar yapmak ve sürdürülebilir önlemler gibi görevleri bulunmaktadır. Düzenli kontrollerin sağlanması ve doğru kararlar alabilmesi adına enerji tasarrufunu sağlamaya yönelik sadece bu işlemlerle ilgilenen personel çalıştırılması önerilmektedir [23]. Işıklandırma ihtiyacı ledli lamba gibi malzemelerle sağlanarak uzun süre enerji verimliliği elde edilmesi amaçlanmalıdır. Ayrıca ışığa duyarlı sensör sistemleri ve açık renkli mobilya kullanımı gibi yöntemler ile ortam kalitesinin aydınlığı sağlanmalıdır. Mümkün oldukça daha küçük kurumların kullanılması ve güneş enerjisi sistemlerin kurulabilmesine ortam hazırlanması enerji verimliliği sağlamak için faydalı yöntemlerdir.

Hastaneler için su çok önemli bir kaynaktır. Çünkü hastaların hijyenini sağlması, tıbbi gereçlerin bakımının yapılması ve kesintisiz su ihtiyacının sağlanması su tüketiminde dikkatli davranılması gerektiğini göstermektedir. Bu yüzden hastanelerin bölümleri tek tek değerlendirerek yıllık, aylık ve hatta günlük kontrolleri ve planlaması yapması gerekmektedir. Hastaneler temelde çamaşırhane, sterilizasyon, havalandırma, tedavi süreci, tıbbi cihazlar, teknik birimler alanlarında su kullanmaktadır [24]. Bu alanlarda su tasarrufu sağlamak amacıyla yağmur sularını değerlendirmek için sarnıç sistemi oluşturmak, kaçak varsa yapılması gerekenleri yaparak tamir etmek, hastane bahçelerinde teknolojik su ve gri su sistemleri kullanmak faydalı uygulamalardır. Hastanelerde çok sık kullanılan tuvalet için düşük akımlı rezervuar ve uzun süreli yatışlarda kullanılan duşlar için düşük debili musluk başlığı kullanmak su verimliliği için gerekli uygulamalardır.

Ayrıca, bulaşık ve çamaşır makinelerinde su tasarrufu sağlayan teknolojiler kullanmak da gerekmektedir. Mesela bu alanlara su saatleri takması örnek bir uygulama olarak verilebilir. Ayrıca, hastaneler için büyük bir sorun olan atık suların yönetimi su tasarrufu sağlamak için değerlendirilmesi gereken önemli bir konudur. Arıtma sistemine kirleticilerin, patojen bakterilerin ve cıva maddesinin karışmaması yönünde dikkat edilmesi ve personelin gerekli bilinç seviyesine ulaştırılması gerekmektedir. Ek olarak ısıl dezenfektan kullanılması su maliyetini düşürecektir. Hastanelerde olabildiğince su ve enerji dengesi sağlanmalıdır. Çünkü birbirlerini olumlu etkilerse su ve enerji tasarrufu birlikte sağlanacaktır. Yağmur suları ve atık suları arıtılarak kullanılarak hem su tasarrufu hem de su kalitesi artacaktır. Bu yüzden atık suları veya lağım suları geri kazandırma yöntemleri geliştirilmelidir. Enerji tasarrufu konusunda uzman bulundurulması gerektiği gibi su tasarrufu konusunda da eğitilmiş uzmanların bulunması gerekmektedir.

Hastanelerde yapım aşamalarında veya yıkılması söz konusu olduğunda çevreye duyarlı yapı oluşturmak gerekmektedir. Özellikle insan sağlığını iyileştirmek temel amacı olan hastanelerin bu kritere dikkat etmesi yeşil hastane olması yolunda iyi bir adım olacaktır. Hastane dış ve iç mekânını oluştururken geri dönüştürülebilir malzemeler kullanmak, iç ortam refahını artırıcı bir alan oluşturmak, tedavi hizmetlerinde temiz teknolojilere yer verilmesi için elverişli tasarım kurulması ve yeterli depolama alanları oluşturmak temel amaç olmalıdır. Ek olarak, hasta ve personel odaklı çevre dostu tasarım uygulamaları üzerinde de durulması gerekmektedir. Bu bağlamda, sağlık personellerin odalarını tıbbi malzemelere yakın olarak yapmak, bisiklet kullanımı teşviki için alan oluşturmak ve stresten uzaklaştırmak adına spor salonu oluşturma uygulamaları yararlı olacaktır.

Hasta ve hasta yakını refahı amacıyla pencerelerin uzaktan kontrolü sağlayacak uygulama, ulaşılabilir şekilde sağlık bilgisi edinme alanları, daha geniş hasta odaları ve ziyaretçilerin dinlenme alanlarına uygun tasarımlar yapmak yeşil hastane kriterlerine katkı sağlayacaktır. İlave olarak hastane kullanıcıları için yer yön tabelalarına yer vermek ve iyi bir ses yalıtımı yapmak şeffaf ve sağlıklı bir tasarım oluşturacaktır. Temelde oluşan tasarım sürdürülebilir yapı malzemeleri kullanılması, alternatif enerji kullanım fırsatları, doğal havalandırma ve aydınlatma ve tüm işleri çevre dostu malzemeler ile gerçekleştirerek çevreye verilen olumsuzlukları minimize edecektir [25]. Ayrıca dış cephesinde uzun ömürlü malzeme kullanılması ilerleyen zamanlarda sürdürülebilirlik amaçları için avantajlı olacaktır. Bir hastanenin çevre dostu tasarlanması uzun süreç de birçok avantaj sağlayacaktır. Bu avantajlar enfeksiyonlardan hastane kullanıcılarını korumak, tedavi maliyetlerini düşürmek, personel iş yükünün azaltılarak sunulan hizmetin kalitesini ve piyasa statüsünü artırmaktır.

Hastaneler sağlık personeli, hastalar, hasta yakınları ve idari personelleri barındırır. Bu hastane kullanıcılarını düşünerek ulaşımın kolay ve kısa sürede sağlanacağı alanlara kurmak gerekmektedir. Mümkün oldukça dijital görüşme sistemleri ve evde bakım hizmet stratejileri üzerinde odaklanılması yeşil hastane uygulamalarına katkı sağlayacaktır. Hasta, hasta yakını veya sağlık personelinin ulaşım faaliyetini kişisel araçlardan toplu taşıma, bisiklet ve yürüyüşe yöneltme stratejileri geliştirilmesi çevreye verilen zararı azaltacaktır. Enerji verimliliğini maksimize etmek amacıyla elektrikli veya hibrit araçlar sağlanmalıdır. Hastanenin tedarikini yakın veya yakıt kullanımında enerji tasarrufu sağlayan işletmelerden alınması önerilmektedir. Mekân tasarımında hastane etrafına veya içine bisiklet kullanım alanları yapmak ulaşımın çevreye vereceği zararı indirecektir. Ayrıca, hastanelerin gıda tedariki sağlanırken yerel ve organik ürünler seçmek gerekmektedir. Çünkü ithal edilip uzaktan gelen yiyecekler ulaşım mesafesini artırarak sera gazı emisyonu havaya yayılacaktır.

Yeşil yapı uygulamaları proje maliyetlerini düşürme amacı da gütmüştür. Bu yüzden, inşaatın çevreye verdiği hasarı minimize etmek amacıyla yerel ve geri kazanım sağlanacak malzeme seçimi büyük önem arz etmektedir. Hatta inşaat aşaması ve sonrası ve ham maddeden ürün veya hizmet çıktısına olan tüm süreçlerin değerlendirilerek malzeme kullanımı çevreye duyarlı seçilmesi gerekliliği yeşil hastane kriterlerin de önemlidir. Genellikle inşaat süreçlerinde taş, kum ve petrol gibi çevreye zararlı malzemeler kullanılarak ekosistemin dengesini bozmaktadır. Bu bağlamda, yeşil yapı gereklilikleri ortaya çıkmıştır. İhtiyaç olunan maddelerin tedariki çevre dostu malzeme satın alma konusunda eğitilmiş kişilerin yapması gerekmektedir. Tedariki sağlanan ürünlerin miktarı ve tavsiye talimatı bilgilerinin kontrollerini sağlamak gerekmektedir. Mesela satın alınan malzeme aldıktan işlemi bitene

kadar olan tüm süreçleri değerlendirilmelidir. Özellikle alternatifi sağlanıyorsa toksik ve biyobirikim maddelerin alınmaması önerilmektedir. Malzeme içeriğinde dikkat edilmesi gerek iki malzeme türü vardır. Bunlar cıva pilli ve PVC kasa teknolojileri tercih edilmemelidir.

Sürekli kapalı alanlarda faaliyetler gerçekleşmektedir. Bu yüzden, insan sağlığı etkilenmekte ve iş verimliliği etkilenmektedir. Sıcaklıkların artışı ile klima gibi serinletici sistemler hava ortamını kirletmektedir. Bu bağlamda, gün ışığından yararlanma arttırılmalı ve açılması mümkün olan pencereler kullanılarak ortam kalitesi yükseltilmelidir. Ortama zararı dokunacak halı, boya kullanımlarına dikkat edilmeli ve gürültü azaltılmalıdır. Eğer iç ortam kalitesi artarsa iş yerlerinde üretkenlik artacak, okullarda başarı sağlanacak ve hastanelerde yatış süreleri azalacaktır [26]. Doğa görünümü alan stres seviyesini azaltarak hastalığın çabuk iyileşmesini sağlamaktadır. Örnek olarak ağaç, doğal malzemeler ve bitki bulunan hastanelerde hastaların ağrılarının dindiği söylenmektedir.

Tehlikeli maddeleri belirli bir kalıpta tanımlamak ve türlerine ayırmak zordur. Çünkü her ülkeye göre tanımı ve sınıflandırma değişmektedir. Hastanelerin içerisinde yer alan bazı zararlı maddelere örnek olarak radyoloji alanından gelen zararlı atıklar ve sterilizasyon gazları verilebilmektedir. Bu tehlikeli maddeler yönetimi için teslim alma, kullanma, depolama ve bertaraf konularında yazılı olarak mevzuatlar oluşturulmalı ve onlara göre hareket edilmelidir. Ayrıca satın alma birimi de talimatlara uygun satış faaliyetini gerçekleştirmelidir. Her tehlikeli maddenin Malzeme güvenlik veri formları (MSDS) olup olmadığı kontrol edilmelidir. Bu beklenen yazılı talimat içeriği kullanım alanı, maddenin tanımı, çevre ve insan sağlığı üzerindeki riskleri, koruyucu önlemler ve kullanma talimatları, acil durum ve ilk yardım talimatları, uygun bertaraf yöntemleri konularını içermeli ve kullanıcılara bunları dikkatlice okumaları için sürekli eğitimsel uyarılar yapılmalıdır. Hastaneye giren tehlikeli madde adım adım bileşiğin ismi ve atık sınıfı belirtilmeli, miktarı gibi kullanım kılavuzu, yeterli güvenlik önlemleri ve bertaraf nasıl edileceğini şeffaf bir şekilde kullanıcılara ifade edilmelidir. Tehlikeli maddelere ilişkin yapılacak diğer uygulamalar ise kullanılması gereken tehlikeli maddelerin çevreye daha az zarar veren eş değerleri araştırılarak kullanılması etkili bir yönetim stratejisidir. Tüm bu aşamaların eksiksiz gerçekleşmesi adına personel eğitimi önem arz etmektedir.

### 3. Enerji Bazlı Faktörlere Yönelik DEMATEL Yöntemi İle Bir Analiz

#### 4.1. Kriter Seti

Bu çalışmada hastanelerin “yeşil hastane” statüsüne geçebilmesi için gerekli olan kriterlerin belirlenmesi hedeflenmektedir. Bu bağlamda, enerji bazlı kriterler inceleme kapsamına alınmıştır. Yapılan geniş çaplı literatür analizi sonucunda dört farklı enerji bazlı kriter belirlenmiş ve detayları Tablo 1’de gösterilmiştir.

**Tablo 1.** Enerji Bazlı Kriter Listesi

Değişkenler	Literatür
Karbon Emisyonunun Azaltılması (K1)	[21]
Temiz Enerji Kullanımı (K2)	[20]
Enerji Verimliliğinin Sağlanması (K3)	[15]
Düşük Enerji Tüketen Ekipmanların Seçilmesi (K4)	[8]

#### 4.2. DEMATEL Yöntemi

DEMATEL yöntemi “Decision making trial and evaluation laboratory” ifadesinin ilk harflerinden elde edilmektedir [27]. Bir hususu etkileyen çok fazla değişken var iken, DEMATEL yardımıyla en önemlisinin bulunması hedeflenmektedir [28,29]. Özetle, DEMATEL çok kriterli karar verme tekniklerinden biridir. Bu çerçevede, elde edilen kriter setlerine ilişkin sorular hazırlanmaktadır [29]. Bu sorulara yönelik uzmanlardan görüş alınmaktadır [30]. Uzman görüşlerinin (değerlendirmelerinin) ortalaması alınarak direk ilişki matrisi elde edilmektedir [31,32]. Bunun ardından, söz konusu matris normalize edilmektedir [33,34]. Daha sonraki aşamada ise toplam ilişki matrisi hesaplanmaktadır [35,36]. Son olarak da toplam matrisindeki rakamlar dikkate alınarak kriter ağırlıkları hesaplanmaktadır [37,38].

### 4.3 Analiz Sonuçları

Analiz sürecinde ilk olarak, değişkenler ile ilgili sorular hazırlanmış ve uzmanlardan bu hususlara yönelik değerlendirmeler talep edilmiştir. Bu süreçte “0-hiç”, “1-az”, “2-orta”, “3-fazla” ve “4-çok fazla” şeklinde 5 farklı skala dikkate alınmıştır. Elde edilen uzman görüşlerine Tablo 2’de yer verilmiştir.

**Tablo 2. Uzman Görüşleri**

Uzman 1				
Kriterler	K1	K2	K3	K4
K1	0	4	2	1
K2	2	0	1	1
K3	2	1	0	1
K4	1	2	3	0
Uzman 2				
Kriterler	K1	K2	K3	K4
K1	0	4	2	2
K2	1	0	2	1
K3	1	2	0	1
K4	1	1	3	0
Uzman 3				
Kriterler	K1	K2	K3	K4
K1	0	4	1	2
K2	1	0	2	2
K3	1	2	0	1
K4	1	1	3	0

Elde edilen uzman görüşlerinin ortalamaları alınarak direk ilişki matrisi elde edilmektedir. Tablo 3 bu matris hakkında detaylı bilgi vermektedir.

**Tablo 3. Direk İlişki Matrisi**

Kriterler	K1	K2	K3	K4
K1	0	4	1,666667	1,666667
K2	1,333333	0	1,666667	1,333333
K3	1,333333	1,666667	0	1
K4	1	1,333333	3	0

Bir sonraki aşamada, direk ilişki matrisindeki satırların toplamı hesaplanmaktadır. Bu süreçte, direk ilişki matrisindeki tüm değerler en büyük satır toplamına bölünmektedir. Bu sayede normalize matris elde edilmektedir. Bu matrisin detaylarına Tablo 4’te yer verilmiştir.

**Tablo 4. Normalize Matrisi**

Kriterler	K1	K2	K3	K4
K1	0	0,545455	0,227273	0,227273
K2	0,181818	0	0,227273	0,181818
K3	0,181818	0,227273	0	0,136364
K4	0,136364	0,181818	0,409091	0

Daha sonra, “ $X*(I-X)^{-1}$ ” formülü dikkate alınarak toplam ilişki matrisi oluşturulmaktadır. Bu formülde, “X” normalize matrisi, “I” ise birim matrisi ifade etmektedir. Toplam ilişki matrisine ait değerler Tablo 5’te paylaşılmıştır.



**Tablo 5.** Toplam İlişki Matrisi

Kriterler	K1	K2	K3	K4
K1	0,43724	1,090025	0,835702	0,638791
K2	0,438816	0,460947	0,615568	0,449299
K3	0,422205	0,621374	0,400793	0,39995
K4	0,448492	0,668465	0,798933	0,332415

Toplam ilişki matrisindeki satır ve sütun toplamları (D ve R) hesaplanmaktadır. Daha sonra bu değerlerin de toplamı (D+R) belirlenmektedir. D+R değeri kriterlerin ağırlıklarının hesaplanmasında dikkate alınmaktadır. Kriter ağırlıkların detaylarına Tablo 6'da yer verilmiştir.

**Tablo 6.** Kriter Ağırlıkları

Kriterler	Kriter Ağırlıkları
Karbon Emisyonunun Azaltılması (K1)	0,262087677
Temiz Enerji Kullanımı (K2)	0,265229761
Enerji Verimliliğinin Sağlanması (K3)	0,248112969
Düşük Enerji Tüketen Ekipmanların Seçilmesi (K4)	0,224569592

Tablo 6'dan anlaşılacağı üzere yeşil hastanelerin gelişebilmesi için en fazla önem arz eden kriter temiz enerji kullanımınıdır. Belirtilen bu hususa paralel olarak, karbon emisyonunun azaltılması da önem arz eden diğer kriterlerdendir. Buna karşın, enerji verimliliğinin sağlanması ve düşük enerji tüketen ekipmanların seçilmesi ise nispeten daha düşük kriter ağırlıklarına sahip olmuşlardır.

## 5. Sonuç ve Tartışma

Bu çalışmada yeşil hastanelerin gelişimi için önem arz eden stratejik unsurların belirlenmesi amaçlanmıştır. Bu bağlamda, ilk olarak kapsamlı bir literatür taraması gerçekleştirilmiştir. Yapılan bu incelemeler sayesinde, enerji bazlı dört farklı faktör tespit edilmiştir. Bu kriterler karbon emisyonunun azaltılması, temiz enerji kullanımı, enerji verimliliğinin sağlanması ve düşük enerji tüketen ekipmanların seçilmesi şeklindedir. Bu kriterler arasından en önemlilerinin tespit edilebilmesi için DEMATEL yöntemiyle bir analiz gerçekleştirilmiştir. Bu sayede, sağlık kuruluşlarının “yeşil hastane” olabilmesi için gerekli olan etkin stratejilerin belirlenebilmesi mümkün olabilecektir. Sonuç olarak, yeşil hastanelerin gelişebilmesi için en fazla önem arz eden kriter temiz enerji kullanımınıdır. Ayrıca, karbon emisyonunun azaltılması da önem arz eden diğer önemli bir faktördür. Öte yandan, enerji verimliliğinin sağlanması ve düşük enerji tüketen ekipmanların seçilmesi ise nispeten daha düşük kriter ağırlıklarına sahip olmuşlardır.

Elde edilen bu sonuçlardan anlaşılacağı üzere, hastanelerin temiz enerji kullanımına acilen geçmeleri gerekmektedir. Bu süreçte, küçük ölçekli güneş panelleri dikkate alınabilmektedir. Bu paneller sayesinde, hastaneler hem kendi enerjilerini üretebilmekte hem de kullanılan enerji çevre dostu olabilmektedir. Buna karşın, bu enerji türlerinin en büyük dezavantajı başlangıç maliyetinin çok yüksek olmasıdır. Bu durum da temiz enerji kullanımının sürekliliğini tehlikeye sokmaktadır. Bu problemin engellenebilmesi için küçük ölçekli güneş panellerinin maliyet avantajı elde etmesi sağlanmalıdır. Bu bağlamda, yapılacak kapsamlı araştırma ve geliştirme çalışmaları sayesinde, bu enerji türlerinin maliyetlerinin düşürülebilmesi mümkün olabilecektir. Belirtilen bu hususa ek olarak, vergi indirimi ve düşük faizli kredi kullanımı gibi devlet teşvikleri sayesinde, küçük ölçekli güneş panellerinin maliyet avantajı elde edebilmesi mümkün olabilecektir.

**Yazar Katkıları:** Bu çalışmada giriş literatür taraması analizler ve sonuç bölümü S.Y., H.D. ve B.Ç. tarafından yapılmıştır.

**Finansman:** Bu çalışma için herhangi bir finansal kaynak sağlanmamıştır.

**Çıkar Çatışması:** Yazarlar herhangi bir çıkar çatışması olmadığını beyan etmemektedir.

**Not:** Bu çalışma Büşra Çelebi'nin İstanbul Medipol Üniversitesi Sağlık Yönetimi yüksek lisans programındaki tezinden türetilmiştir

## Kaynakça

- [1] Dong, W., Zhao, G., Yüksel, S., Dinçer, H., & Ubay, G. G. "A novel hybrid decision making approach for the strategic selection of wind energy projects". *Renewable Energy*, vol. 185, 2022, pp. 321-337.
- [2] Li, X., Zhao, L., Yu, J., Liu, X., Zhang, X., Liu, H., & Zhou, W. "Water splitting: from electrode to green energy system" *Nano-Micro Letters*, vol. 12, no. 1, 2020, pp. 1-29.
- [3] Kou, G., Yüksel, S., & Dinçer, H. "Inventive problem-solving map of innovative carbon emission strategies for solar energy-based transportation investment projects" *Applied Energy*, vol. 311, 118680, 2022.
- [4] Tsang, C. H. A., Huang, H., Xuan, J., Wang, H., & Leung, D. Y. C. "Graphene materials in green energy applications: recent development and future perspective" *Renewable and Sustainable Energy Reviews*, vol. 120, 109656, 2020.
- [5] Mensah-Darkwa, K., Zequine, C., Kahol, P. K., & Gupta, R. K. "Supercapacitor energy storage device using biowastes: A sustainable approach to green energy" *Sustainability*, vol. 11, no. 2, 414, 2019.
- [6] Zhang, Y., Zhang, Y., Gong, C., Dinçer, H., & Yüksel, S. "An integrated hesitant 2-tuple Pythagorean fuzzy analysis of QFD-based innovation cost and duration for renewable energy projects" *Energy*, vol. 248, 123561, 2022.
- [7] Adom, P. K., Amuakwa-Mensah, F., Agradi, M. P., & Nsabimana, A. "Energy poverty, development outcomes, and transition to green energy" *Renewable Energy*, vol. 178, 2021, pp. 1337-1352.
- [8] Li, J., Yüksel, S., Dinçer, H., Mikhaylov, A., & Barykin, S. E. "Bipolar q-ROF hybrid decision making model with golden cut for analyzing the levelized cost of renewable energy alternatives" *IEEE Access*, vol. 10, 2022, pp. 42507-42517.
- [9] Liobikienė, G., & Dagiliūtė, R. "Do positive aspects of renewable energy contribute to the willingness to pay more for green energy?" *Energy*, vol. 231, 120817, 2021.
- [10] Kostis, P., Dinçer, H., & Yüksel, S. "Knowledge-Based Energy Investments of European Economies and Policy Recommendations for Sustainable Development" *Journal of the Knowledge Economy*, 2022, pp. 1-33.
- [11] Tan, H., Li, J., He, M., Li, J., Zhi, D., Qin, F., & Zhang, C. "Global evolution of research on green energy and environmental technologies: A bibliometric study" *Journal of Environmental Management*, vol. 297, 113382, 2021.
- [12] Wu, X., Dinçer, H., & Yüksel, S. "Analysis of crowdfunding platforms for microgrid project investors via a q-rung orthopair fuzzy hybrid decision-making approach" *Financial Innovation*, vol. 8, no. 1, 2022, pp. 1-22.
- [13] Stucki, T. "Which firms benefit from investments in green energy technologies?—The effect of energy costs" *Research Policy*, vol. 48, no. 3, 2019, pp. 546-555.
- [14] Dinçer, H., Yüksel, S., Aksoy, T., & Hacıoğlu, Ü. "Application of M-SWARA and TOPSIS methods in the evaluation of investment alternatives of microgeneration energy technologies" *Sustainability*, vol. 14, no. 10, 6271, 2022.
- [15] Dinçer, H., Yüksel, S., Uluç, G. S., & Çağlayan, Ç. "Green Nuclear Energy: A Solution of Environmental Sustainability for Emerging Economies" In *Environmental Sustainability, Growth Trajectory and Gender: Contemporary Issues of Developing Economies* (pp. 63-73). Emerald Publishing Limited, 2022.
- [16] Mukhtarov, S., Yüksel, S., & Dinçer, H. "The impact of financial development on renewable energy consumption: Evidence from Turkey" *Renewable Energy*, 2022.
- [17] Byun, J., Kwon, O., Park, H., & Han, J. "Food waste valorization to green energy vehicles: sustainability assessment" *Energy & Environmental Science*, vol. 14, no. 7, 2021, pp. 3651-3663.
- [18] Yüksel, S., Dinçer, H., Eti, S., & Adalı, Z. "Strategy improvements to minimize the drawbacks of geothermal investments by using spherical fuzzy modelling" *International Journal of Energy Research*, 2022.
- [19] Bhuiyan, M. A., Dinçer, H., Yüksel, S., Mikhaylov, A., Danish, M. S. S., Pinter, G., ... & Stepanova, D. "Economic indicators and bioenergy supply in developed economies: QROF-DEMATEL and random forest models" *Energy Reports*, vol. 8, 2022, pp. 561-570.
- [20] Yüksel, S., & Dinçer, H. "Identifying the strategic priorities of nuclear energy investments using hesitant 2-tuple interval-valued Pythagorean fuzzy DEMATEL" *Progress in Nuclear Energy*, vol. 145, 104103, 2022.
- [21] Wan, Q., Zhao, X., Liu, H., Dinçer, H., & Yüksel, S. "Assessing the New Product Development Process for the Industrial Decarbonization of Sustainable Economies" *SAGE Open*, vol. 12, no. 1, 21582440211067231, 2022.
- [22] Zhao, Y., Korsakienė, R., Dinçer, H., & Yüksel, S. "Identifying Significant Points of Energy Culture for Developing Sustainable Energy Investments" *SAGE Open*, vol. 12, no. 1, 21582440221087262, 2022.
- [23] Jabeen, G., Ahmad, M., & Zhang, Q. "Factors influencing consumers' willingness to buy green energy technologies in a green perceived value framework" *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 16, no. 7, 2021, pp. 669-685.
- [24] Chakraborty, S. K., & Mazzanti, M. "Energy intensity and green energy innovation: Checking heterogeneous country effects in the OECD" *Structural Change and Economic Dynamics*, vol. 52, 2020, pp. 328-343.
- [25] Krawinkler, A., Breitenacker, R. J., & Maresch, D. "Heuristic decision-making in the green energy context: Bringing together simple rules and data-driven mathematical optimization" *Technological Forecasting and Social Change*, vol. 180, 121695, 2022.
- [26] Teodor, A. H., & Bruce, B. D. "Putting photosystem I to work: truly green energy" *Trends in Biotechnology*, vol. 38, no. 12, 2020, pp. 1329-1342.
- [27] Haiyun, C., Zhixiong, H., Yüksel, S., & Dinçer, H. "Analysis of the innovation strategies for green supply chain management in the energy industry using the QFD-based hybrid interval valued intuitionistic fuzzy decision approach" *Renewable and Sustainable Energy Reviews*, vol. 143, 110844, 2021.
- [28] Kou, G., Olgu Akdeniz, Ö., Dinçer, H., & Yüksel, S. "Fintech investments in European banks: a hybrid IT2 fuzzy multidimensional decision-making approach" *Financial Innovation*, vol. 7, no. 1, 2021, pp. 1-28.

- 
- [29] B. Leblebicioğlu and A. Keskin , "Evaluation of Supplier Selection Criteria with Fuzzy DEMATEL Method: An Application on the Pharmacy Industry", *Avrupa Bilim ve Teknoloji Dergisi*, no. 23, pp. 236-242, Apr. 2021, doi:10.31590/ejosat.864116
- [30] Zhao, Y., Xu, Y., Yüksel, S., Dinçer, H., & Ubay, G. G. "Hybrid IT2 fuzzy modelling with alpha cuts for hydrogen energy investments" *International Journal of Hydrogen Energy*, vol. 46, no. 13, 2021, pp. 8835-8851.
- [31] Yuan, G., Xie, F., Dinçer, H., & Yüksel, S. "The theory of inventive problem solving (TRIZ)-based strategic mapping of green nuclear energy investments with spherical fuzzy group decision-making approach" *International Journal of Energy Research*, vol. 45, no. 8, 2021, pp. 12284-12300.
- [32] Xie, Y., Zhou, Y., Peng, Y., Dinçer, H., Yüksel, S., & an Xiang, P. "An extended pythagorean fuzzy approach to group decision-making with incomplete preferences for analyzing balanced scorecard-based renewable energy investments" *IEEE Access*, vol. 9, 2021, pp. 43020-43035.
- [33] Ding, Z., Yüksel, S., & Dinçer, H. "An Integrated Pythagorean fuzzy soft computing approach to environmental management systems for sustainable energy pricing" *Energy Reports*, vol. 7, 2021, pp. 5575-5588.
- [34] Liu, J., Lv, J., Dinçer, H., Yüksel, S., & Karakuş, H. "Selection of renewable energy alternatives for green blockchain investments: A hybrid IT2-based fuzzy modelling" *Archives of Computational Methods in Engineering*, vol. 28, no. 5, 2021, pp. 3687-3701.
- [35] Meng, Y., Wu, H., Zhao, W., Chen, W., Dinçer, H., & Yüksel, S. "A hybrid heterogeneous Pythagorean fuzzy group decision modelling for crowdfunding development process pathways of fintech-based clean energy investment projects" *Financial Innovation*, vol. 7, no. 1, 2021, pp. 1-34.
- [36] Fang, S., Zhou, P., Dinçer, H., & Yüksel, S. "Assessment of safety management system on energy investment risk using house of quality based on hybrid stochastic interval-valued intuitionistic fuzzy decision-making approach" *Safety science*, vol. 141, 105333, 2021.
- [37] Jun, Q., Dinçer, H., & Yüksel, S. "Stochastic hybrid decision-making based on interval type 2 fuzzy sets for measuring the innovation capacities of financial institutions" *International Journal of Finance & Economics*, vol. 26, no. 1, pp. 573-593.
- [38] Eti, S. "Ulakbim İndeksinde Taranan Sosyal Bilimler Alanındaki Dergilerde Öne Çıkan Konu ve Yöntemlerin Metin Madenciliği Yaklaşımı ile Belirlenmesi" *Uluslararası Hukuk ve Sosyal Bilim Araştırmaları Dergisi*, vol. 1, no. 1, 2019, pp. 61-66.
- [39] Liu, W., Dinçer, H., Eti, S., & Yüksel, S. "Gaussian-based hybrid approach to Entropy for analyzing energy efficiency of emerging economies" *Energy Reports*, vol. 7, 2021, pp. 2501-2511.

Book Review

Received: date:27.03.2022

Accepted: date:15.06.2022

Published: date:30.06.2022

# SPSS ile İstatistik Rehberi

## Doç. Dr. İbrahim Demir

İstanbul: Efe Akademi Yayınları, 2020, 1. Baskı, 538 sayfa, ISBN: 978-605-60173-2-2

Hasan Aykut Karaboğa<sup>1</sup>

<sup>1</sup>Amasya Department of Measurement and Evaluation in Education, Amasya, Turkey; h.aykut.karaboga@amasya.edu.tr  
Orcid: 0000-0001-8877-3267<sup>1</sup>

### Değerlendiren: Hasan Aykut Karaboğa

Bu çalışmada Doç. Dr. İbrahim Demir tarafından “SPSS ile İstatistik Rehberi” isimli eser incelenmiştir. Eserin ilk baskısı Eylül 2020 tarihinde Efe Akademi Yayınları tarafından yayınlanmıştır. Toplam 538 sayfalık bu eserin genç araştırmacılar ve istatistiğe meraklı tüm okuyucular için alanda önemli bir boşluğu doldurması beklenmektedir. Bilindiği gibi günümüzde istatistiksel analizlerde paket programlar uygulama açısından büyük kolaylıklar sağlamaktadır. Sosyal bilimler ve eğitim bilimleri alanında SPSS en yaygın kullanılan paket programlardan biridir. Ancak araştırma verilerinin analizinde ve doğru yorumlanmasında rehber niteliğinde yol gösterici kaynak sıkıntısı yaşanmaktadır. Ele alınan bu eser sosyal bilimler ve eğitim bilimleri alanında araştırmalar yapan, istatistik ve SPSS paket programı konusunda yeterli bilgiye sahip olmayan kişilere analizleri adım adım uygulayarak öğrenmelerini ve sonuçları raporlayabilmelerini sağlamaya yönelik olarak hazırlanmıştır.

Kitabın önemli özelliklerinden biri bölümlerin birbirinden bağımsız olarak hazırlanmış olmasıdır. Böylece okuyucuların yalnızca ilgilendiği bölümü okuyup zaman kaybetmelerinin önüne geçilmiştir. Kitabı oluşturan 17 bölümde araştırma sorularının oluşturulması, örnekleme, ölçekler, temel SPSS bilgileri, betimsel istatistik tabloları, normallik testleri, hipotez testleri, parametrik ve parametrik olmayan testler, varyans analizi, korelasyon analizi, regresyon analizi, faktör analizi ile geçerlilik ve güvenilirlik analizi konuları ele alınmıştır.

Kitabın ilk bölümünde bilimsel araştırmanın ilk aşaması olan bilimsel araştırma süreci anlatılmıştır. İlk olarak anket hazırlama sürecinin anlatıldığı bu bölümde amaca uygunluk, hedef kitle, pilot testler, soru içerikleri ve türleri, hazırlık sürecinde yapılan hatalar ve davranış araştırmaları için örnek desenler ele alınmıştır. Araştırmanın en önemli ve ilk aşaması olan hipotezlerin oluşturulması da bu kısımda vurgulanmıştır.

İkinci bölümde yapılacak analizleri belirleyebilmek ve analizlere yön vermek amacıyla değişkenlerin isimlendirilmesinden bahsedilmiştir. Ölçek türleri, değişken türleri, ölçme kavramı ve değişken kavramı ve değişken kavramı bilgileri verilmiştir.

Üçüncü bölümde ikinci bölüme destek olarak SPSS programında bahsedilmiştir. Programın başlatılması, değişkenlerin programda tanımlanması, örnek anket tanımlanması ve veri girişi editörü açıklanmıştır. Ayrıca araştırmacıların SPSS programına verileri kolaylıkla aktarmasını sağlayan SPSS'e Excel'den veri aktarılması ele alınmıştır. Bu bölümde kayıp verilerin sisteme girişi konusunda ek bilgiler de verilmiştir. Alıştırma kısmında kullanıcıların öğrenmesini kolaylaştıracak örnek bir anket verilerek anketin SPSS'e girişi istenmiştir. Girişi yapılan anket için “ilkders.sav” adlı dosya ile karşılaştırma yapılarak kontrolü teşvik edilmiştir.

SPSS programının açılır menüleri dördüncü bölümde açıklanmıştır. Dosyaların açılması, verilerin SPSS'e aktarılması, yeni dosyaların oluşturulması, dosyaya veri ekleme veya veri düzenleme, tercihlerle programın kişiselleştirilmesi, eldeki veri setinden belirli koşullar ile örnek seçimi, yeni bir değişken tanımlanması ve veri dönüştürme bu bölümde ele alınmıştır. Ayrıca sosyal bilimlerde ve eğitim bilimlerinde sonuçların anlaşılır olmasını sağlamak için yaygın olarak kullanılan grafik menüsü ve grafiklerin özelleştirilmesi konularına da bu bölümde değinilmiştir.

Verilerin doğru şekilde anlaşılmasını sağlayan betimsel istatistikler, mod, medyan, aritmetik ortalama gibi ölçüler ile frekans tabloları, çapraz tablolar beşinci bölümde anlatılmıştır. Bir sonraki bölümde ise istatistiksel analizlerde en önemli varsayımlardan olan normallik varsayımı incelenmiştir. Parametrik istatistiksel yöntemlerin kullanılabilmesini sağlayan normallik yaklaşımının test edilmesinde kullanılan görsel ve sayısal yöntemler yine bu bölümde açıklanmıştır. Uygulama kısmında ise değişkenin örnek sayısına göre Kolmogorov-Smirnov ya da Shapiro-Wilk testlerinin kullanımı gösterilmiştir. Normallik testlerinin hipotezleri de örnek olarak verilen durumlar için okuyucuya sunulmuştur.

İstatistiğin önemli başlıklarından biri olan grup karşılaştırmaları ve ortalamadan farklılıkların tespiti konuları hipotez testleri başlığı altında altıncı bölümde verilmiştir. İstatistik derslerinde bahsedildiği gibi çalışmalarda çoğunlukla ana kütleyle ulaşmak mümkün değildir. Bu nedenle örneklem alınarak ana kütle parametreleri hakkında çıkarımlarda bulunulmaktadır. Hipotez testleri ile örneklemde elde edilen değerlerin belirli bir güven düzeyinde ana kütle parametrelerinden farklılıkları araştırılmaktadır. Bu amaçla hipotez testlerinde yer alan önemli kavramları, hipotez yapısına göre kullanılacak testleri ve test sonucunda elde edilen değerlerin yorumlanmasını yedinci bölümde bulmak mümkündür. Ek olarak etki büyüklüğü de uygulamalar üzerinden anlatılmıştır. Tek örneklem t-testi, bağımsız örneklem t-testi ve bağımlı örneklem t-testi raporlama ilkeleri dikkate alınarak anlaşılır şekilde tanıtılmıştır.

Bir sonraki bölümde çalışmalarda sıklıkla karşılaşılan parametrik yöntemlerden olan varyans analizi ele alınmıştır. Bu bölümde tek yönlü varyans analizi, iki yönlü varyans analizi, tekrarlı ölçümlerde varyans analizi ile karışık ölçümler için tekrarlı varyans analizi uygulamalar üzerinden işlenmiştir. Kovaryans analizi ise farklı bir kısımda diğer varyans analizi yöntemleri olarak açıklanmıştır. Her iki bölümde de durumlara uygun olarak verinin nasıl test edileceği, hangi testin kullanılması gerektiği gibi incelikler ek bilgi ve pratik bilgi şeklinde yönlendirmelerle vurgulanmıştır. Bu bölümlerin teoriye girilmeden pratik örnekler üzerinden anlatılması okuyucuların kafa karışıklığına uğramadan ve bilgi bombardımanı ile istatistik tabloları arasında boğulmadan konuyu özümsemesini sağlayabilir.

Bilindiği gibi analiz edilen veriler her zaman normal dağılmamaktadır. Böyle durumlarda araştırmacılar parametrik testleri kullanamayacağından alternatif yöntemlere yönelmektedirler. Dokuzuncu bölümde t-testlerinin ve varyans analizi yöntemlerinin parametrik olmayan alternatifleri verilmiştir. Bölümde öncelikle parametrik olmayan yöntemlerin neler olduğu, bu yöntemlerin ne zaman tercih edilmesi gerektiği anlatılmış, ardından örnekler üzerinde yöntemler açıklanmıştır.

Onuncu bölümde yine parametrik olmayan yöntemlerden olan Ki-kare testleri verilmiştir. Özellikle isimsel değişkenlerin analizinde uygunluk, homojenlik ve bağımsızlık amacıyla kullanılan ki-kare testlerinin ayrıca sıralı ve aralıklı ölçek türleri için de kullanılabileceği gösterilmiştir. Çapraz tablolar için McNemar, Odds oranı ve Kappa analizi de birer örnekle anlatılmıştır.

Hipotez testlerinin önemli bir diğer başlığı olan ilişki analizleri on birinci bölümden itibaren ele alınmıştır. On birinci bölümde iki değişken arasındaki ilişkinin en önemli göstergelerinden olan korelasyon analizi sunulmuştur. Sıralı, aralıklı ve oransal değişken türleri için parametrik ve parametrik olmayan korelasyon analizi yöntemleri örneklerle verilmiştir.

Aralarında ilişki bulunan değişkenlerin bağımlılık yapısının incelendiği modeller regresyon analizi olarak belirtilmiştir. Buna göre bir bağımlı ve bir veya birden fazla bağımsız değişkenin ilişkisi doğrusal regresyon analizi ile on ikinci bölümde incelenmiştir. Bu kısım sosyal bilimlerde ve eğitim bilimlerinde önemli bir yer tutmaktadır. Regresyon analizi varsayımları, yöntem seçimi, basit doğrusal regresyon analizi, kukla değişkenli regresyon analizi, doğrusal hale getirilebilen modeller ve eğri tahmini örnek üzerinden ve programda uygulama adımları verilerek anlatılmıştır. Katsayıların yorumlanması,

modelin açıklayıcılık değerinin yorumlanması, çoklu doğrusal bağlantı problemi ve otokorelasyon problemi uygulama kısmında anlaşılır şekilde detaylandırılmıştır. Modelin belirlenmesinde stepwise, backward, forward ve remove yöntemleri kısaca açıklanmıştır.

Bağımlı değişkenin isimsel olduğu durumlarda klasik regresyon analizini kullanmak mümkün olmamaktadır. Klasik regresyon analizi yerine kullanılacak regresyon yöntemleri on dördüncü bölümde verilmiştir. Bağımlı değişkeni isimsel, sıralı ya da aralıklı değişken olan bu yöntemlerin regresyon analizi gibi varsayımları yoktur. Sırasıyla lojistik, multinominal, ordinal ve probit regresyon analizi yöntemleri hakkında uygulamalı olarak bilgiler verilmiş ve analiz çıktıları yorumlanmıştır.

Sosyal bilimlerde ve eğitim bilimlerde son yıllarda yaygın olarak kullanılan yöntemlerden bir de aracılık ilişkilerinin test edilmesidir. Moderatör ve mediatör olarak da bilinen regresyon analizinin bu özel hali on beşinci bölümde kısaca anlatılmıştır. Aracılık modellerinde bağımsız değişkenin/değişkenlerin bağımlı değişken üzerindeki doğrudan ve dolaylı etkisi incelenmektedir. Bu bölümde çoğunlukla mediatör ilişki üzerinde durulmuştur. Literatürde birçok mediatör ve moderatör ilişki modeli bulunmaktadır. Bu konunun anlaşılması için farklı kaynaklardan destek alınması önerilmektedir.

Ölçek geliştirme çalışmalarının bel kemiğini oluşturan faktör analizi on altıncı bölümde verilmiştir. Çok sayıda değişkenin daha az sayıda, anlamlı ve birbirinden bağımsız değişkenlere dönüştürülmesi için faktör analizi kullanılmaktadır. Açıklayıcı ve doğrulayıcı faktör analizi olarak ikiye ayrılan bu yöntemlerden SPSS programında yalnızca açıklayıcı faktör analizi uygulanabilmektedir. Oldukça önemli olan bu yöntemin varsayımları, aşamaları ve faktör analizi sonuçlarının yorumlanması bu bölümde anlatılmıştır. Ayrıca literatürde sıklıkla birbirine karıştırılan Temel Bileşenler Analizi ile faktör analizi farklı ek bilgi olarak verilmiştir. Son bölümde ise ölçek geliştirme çalışmalarının aşamaları olan geçerlilik ve güvenilirlik analizi okuyucuya genel fikir vermek için kısaca açıklanmıştır. Kitabın en önemli avantajı uzun zaman içerisinde derlenen ders ve kurs notlarından oluşmasıdır. Sade ve anlaşılır bir dilin kullanıldığı bu kitapta özellikle sosyal bilimlerin çeşitli alanlarında ve eğitim bilimlerinde çalışmalar yapan, çalışmalarını bilimsel bir ürüne dönüştürmek isteyen araştırmacılara yol gösterici olabilecek niteliktedir. Araştırmacıları ayrıntıya boğmadan, kullanılan analizleri amaçları ve kısıtları ile veren yönlendirici bir eserdir. Aşağıda verilen özellikler çalışmayı kıymetli hale getirmektedir:

- ❖ Konuların paket program üzerinden anlatılması,
- ❖ Analizler için uyarı amacıyla verilen çeşitli ek bilgilerin renkli kareler içerisinde dikkate çekici olarak sunulması,
- ❖ Bir önceki bölümde neler anlatıldığı ve ilgili bölümde neler anlatılacağı kısaca verilmesi,
- ❖ Her bölümün giriş sayfasında konu başlıklarının sunulması,
- ❖ Analiz sonuçlarının anlaşılır şekilde açıklanması,
- ❖ Analizlerin hipotezlerinin analize başlarken verilmesi,
- ❖ Analiz sonuçlarının APA formatında makale yapısına uygun olarak verilmesi,
- ❖ Her anlatılan konunun muhakkak örnek üzerinden verilmesi,
- ❖ Bölüm sonlarında öğrenmeyi pekiştirecek alıştırmaların sunulması

Diğer taraftan kitabın çeşitli eksiklikleri de göze çarpmaktadır. Öncelikle kitapta küçük yazım hataları bulunmaktadır. Bu hataların dikkatle kontrol edilerek düzeltilmesi önerilmektedir. Ayrıca çalışmaya dizin eklenmesi okuyucu açısından faydalı olacaktır. Ek olarak konuyu derinlemesine irdelemek ve konunun teorisini öğrenmek isteyen okuyucular düşünülerek bölüm sonlarına kaynak önerileri eklenmesi faydalı olacaktır. Kitabın yeni baskılarında ayırma analizi ve kümeleme analizi gibi çok değişkenli istatistik konularının da kitaba bölüm olarak eklenmesinin kitabı bir başucu kitabına dönüştürebileceği öngörülmektedir.

## Kaynakça

- [1] Demir, I. (2020). SPSS ile İstatistik Rehberi (1. Baskı). İstanbul: Efe Akademi Yayınları.