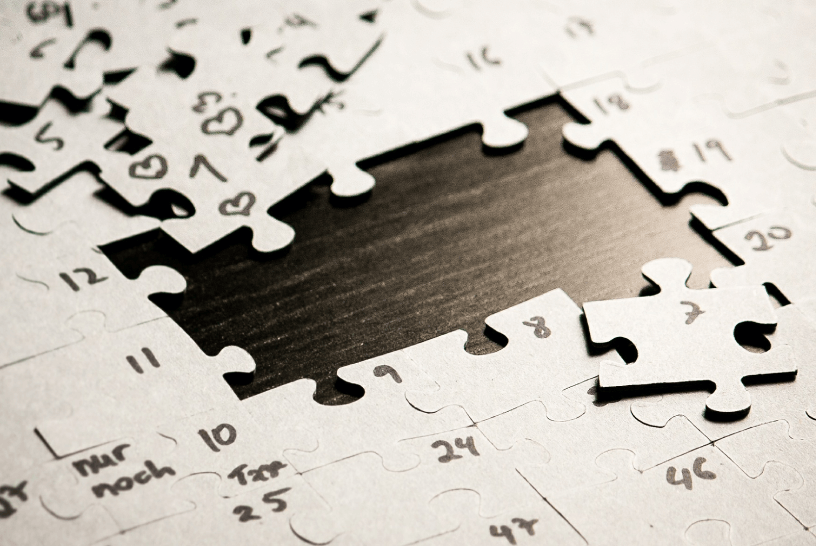
# Missing Data

## ****What is Missing Data?****

The reliability and consistency of visualizations and analyses are negatively affected when there is missing data in a dataset. This can lead to incorrect conclusions as the dataset fails to positively represent the target audience. Additionally, missing data can hinder the smooth operation of machine learning algorithms, preventing them from effectively solving classification problems.

Missing data refers to situations where at least one measurement for the variables in a study cannot be obtained. In other words, it indicates the absence of a value in any cell of the dataset.



**Causes of Missing Data:**

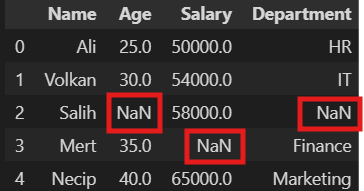
* A respondent typically not answering a specific question
* A respondent not completing the entire survey
* Data being accidentally deleted
* Incomplete measurements from data collection tools
* Loss of data during processing operations

These factors can lead to the occurrence of missing data.

In applications, missing data is represented by expressions such as “NULL,” “NaN,” or “NA.”

Missing data negatively affects the validity, reliability, and explainability of the study and also adversely impacts the functioning of machine learning algorithms. Therefore, missing data must be addressed before analyses in the research process. First, the amount and type of missing data are determined, and a missing data analysis is conducted to observe the missing data in the dataset. Subsequently, various methods and techniques have been developed for managing missing data, including data imputation, excluding missing observations from the analysis, and modern statistical methods (e.g., multiple imputation, expectation-maximization).

(Akbaş, U., & Koğar, H. (2020))



**Missing Data Analysis**

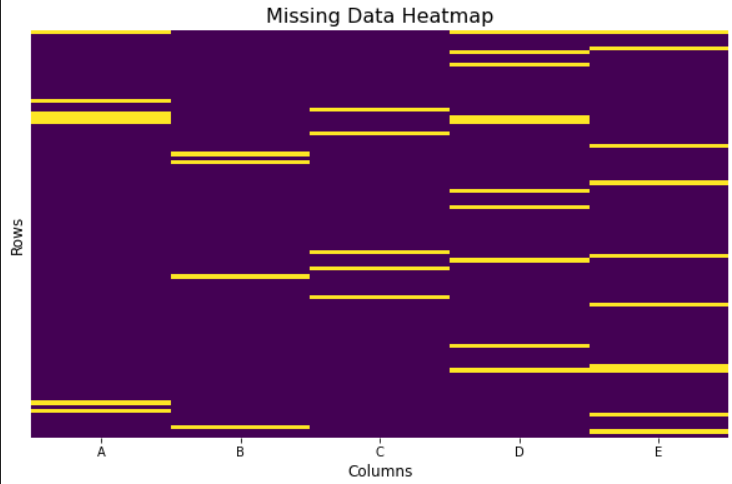
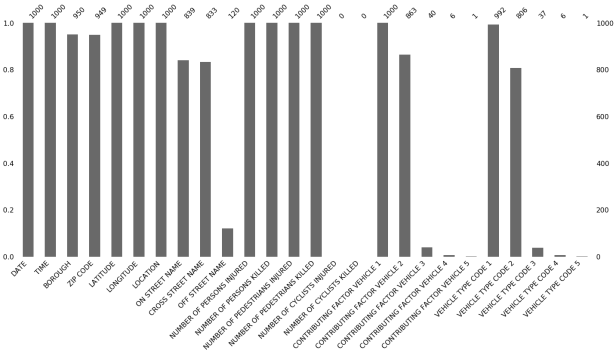
Missing data analysis is the process of examining missing observations or values in a dataset and evaluating the impact of these missing entries. It is conducted to understand the reasons for missing data, decide how to address them, and ensure the validity of the results.

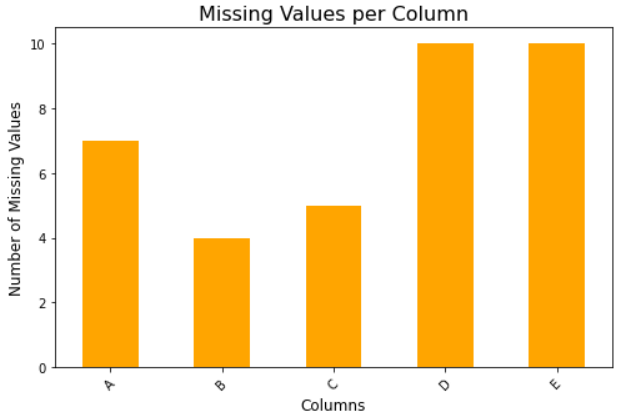
**1.Identifying Missing Data**  
The dataset is examined to identify missing data and ensure that they are marked as NaN, NULL, or NA. The amount of missing data is determined. You can represent the extent of missing data both numerically and visually.If a variable contains a large amount of missing data, it is necessary to evaluate whether it is appropriate to include this variable in the analysis.

Visualizing missing data helps to interpret the analysis more easily.

(Dayanıklı, 2021)

**Examples of Visualizing Missing Data**





**2.Identifying Types of Missing Data**

Missing data can be classified into three categories: MCAR, MAR, or MNAR. This classification influences which methods are to be used.

**2.1 MCAR (Missing Completely at Random):**

The loss of data occurs completely randomly, with no systematic relationship between the missing data and any observed data. For example, a respondent might skip certain questions in a survey by accident.

**2.2 MAR (Missing at Random):**

The missing data may be related to other observed variables but is not directly related to the missing values themselves. For instance, people who do not answer a particular question in a survey may have a specific characteristic (e.g., age) that correlates with their missing data.

**2.3 MNAR (Missing Not at Random):**

There is a relationship between the missing data and both observed and unobserved variables. In other words, the missing data is directly related to the observed data. For example, people with a high income might not respond to income-related questions, which can be an example of MNAR.  
(Jason S. Haukoos, Craig D. Newgard, 2007)

Various methods are applied to determine the type of missing data. The most useful method involves correlation analysis of the missing data, where the relationship between missing data is examined. Based on this, decisions are made regarding whether the variable should be included in the analysis and how to handle the missing data.

**3.Solutions for Missing Data**

The famous statistician Ronald Fisher once said, “The best way to handle missing data is to not have missing data at all.” In order to make accurate analyses, missing data must be addressed. To determine the best method for dealing with missing data, the amount, type, and importance of the missing data in the variable should be considered. Based on this, the missing data can be handled by either deleting the rows/columns or imputing values using different techniques.

**3.1 Deletion Methods**

Deleting missing data is a method of excluding incomplete observations from analysis and is generally classified into two main approaches: listwise deletion and pairwise deletion.

**3.1.1 Listwise Deletion:**

In this method, all observations with missing data in one or more variables are excluded from the analysis. If any data is missing in an observation, that entire observation is excluded from the analysis.

**Advantages:** Simple and straightforward method. It can reduce bias caused by missing data.

**Disadvantages:** When the missing data rate is high, it can significantly reduce the sample size, which in turn may lower statistical power. Also, if the missing data is not random, it may lead to biased results.

**3.1.2 Pairwise Deletion:**

In this method, only the observations with missing data in a specific variable are excluded from the analysis for that variable.

**Advantages:** Allows more data to be used in analysis, preserving the sample size.

**Disadvantages:** Since different observations are used in different analyses, the consistency of the results may be affected. Additionally, this method can also create bias if the missing data is not random.

Deletion methods are typically selected based on the size of the dataset, the missing data rate, and the reasons for the missing data. However, the impact of these methods on bias and sample representativeness must be considered. Researchers should be careful when deleting missing data and consider alternative methods as well.

**3.2 Imputation Methods**

These methods involve assigning predicted values to replace missing data. They are preferred to reduce the impact of missing data and improve the validity of the analysis.

**3.2.1 Mean Imputation:**

The mean of a variable is calculated for all observations, and the missing values are replaced by this mean value.

**Advantages:** Easy to apply and provides a quick solution.

**Disadvantages:** It can distort the data distribution and reduce variability, leading to biased results.

**3.2.2 Nearest Neighbor Mean Imputation:**

The mean of the observations closest to the missing data is used to replace the missing value.

**Advantages:** Provides more realistic estimates by considering similar observations.

**Disadvantages:** It can be computationally more complex, and selecting the neighbors can introduce bias.

**3.2.3 Nearest Neighbor Median Imputation:**

The median of the observations closest to the missing data is used to replace the missing value.

**Advantages:** Less influenced by outliers and provides more robust estimates.

**Disadvantages:** Neighbor selection issues can still arise.

**3.2.4 Linear Interpolation:**

Missing data is predicted based on the linear relationship between the observations before and after the missing value.

**Advantages:** Effective for time series data.

**Disadvantages:** Only valid for linear relationships and can be misleading in complex data structures.

**3.2.5 Linear Trend Extrapolation:**

A value is assigned to the missing data based on the trend of existing data.

**Advantages:** Can provide more accurate predictions by considering trends in the data.

**Disadvantages:** It may lead to incorrect results if trends are misinterpreted.

**3.2.6 Expectation-Maximization (EM):**

The probability distribution of the missing data is estimated, and the best predictions are made for the missing values. This method works iteratively.

**Advantages:** Provides more accurate predictions by considering the structure of missing data.

**Disadvantages:** Computationally complex and time-consuming.

**3.2.7 Regression Imputation:**

Regression imputation is one of the methods that provides the closest values after multiple imputations. It is effective when the missing data rate is low.

**Advantages:** Takes into account the relationships between variables in the dataset. The assigned values do not affect the standard deviation or distribution of the variable.

**Disadvantages:** High correlation may occur between the variable with missing data and the predictors, reducing uncertainty in the assigned values.

**3.2.8 Multiple Imputation:**

Multiple predictions are made for missing data, and these predictions are combined to provide final results. This method also takes uncertainty into account.

**Advantages:** Reduces bias and provides more reliable results.

**Disadvantages:** Complex and time-consuming to implement.

**4. Performance Evaluation**

Each dataset has unique characteristics, so the effectiveness of missing data handling methods can vary based on the dataset. To determine which method is more effective, several evaluation techniques are used.

### 4.1 Model Fit

The suitability of applied methods can be evaluated by model fit criteria such as AIC, BIC, and R-squared. Lower AIC or BIC values indicate a better fit, while higher R-squared values indicate greater accuracy in the model's predictions.

### 4.2 Simulation Studies

Simulation studies can be conducted to assess the performance of different methods. In these studies, methods are compared under various scenarios, such as different missing data rates and data structures. This helps identify which methods perform best under various conditions.

### 4.3 Validity and Reliability Analyses

It is essential to compare the validity (the accuracy of the scale's factor structure) and reliability (the consistency of the scale) of the methods. High reliability coefficients (e.g., Cronbach's alpha) and valid factor structures indicate the effectiveness of the method.

### 4.4 Literature Review

Previous research findings and recommendations can provide guidance on which methods are more effective. The results from similar studies in the literature can assist in the decision-making process.

These criteria and methods help researchers identify which missing data treatment method is more effective. In every case, the specific research context and the characteristics of the dataset should also be considered.

#### Example:

In a study by Kürşad M. and Nartgün Z. (2015), different missing data percentages (5%, 10%, and 20%) were simulated. The dataset was used to evaluate the performance of missing data methods, applying both deletion and imputation techniques. Performance evaluation methods like AIC, BIC, and R-squared were used for comparison.

The results indicated that multiple imputation was the most effective method for addressing missing data. Expectation-Maximization and regression imputation also provided effective alternatives, but multiple imputation yielded better results, especially with higher missing data rates.

**Referances**

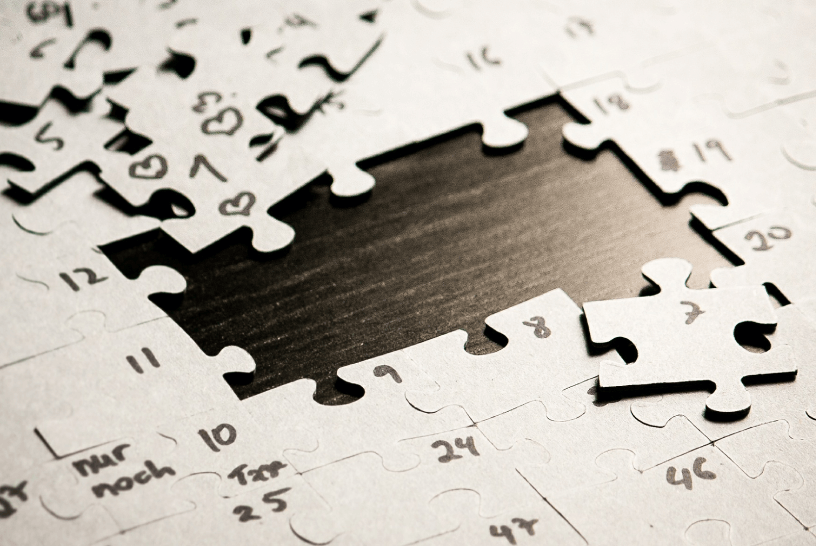
* Akbaş, U., & Koğar, H. (2020). Nicel Araştırmalarda Kayıp Veriler ve Uç Değerler. Ankara: Pegem Akademi Yayıncılık.
* Kürşad,M.,& Nartgün,Z.(2015).Kayıp Veri Sorununun Çözümünde Kullanılan Farklı Yöntemlerin Ölçeklerin Geçerlik ve Güvenirliği Bağlamında Karşılaştırılması.Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,2015; 6(2); 254-267
* Jason S. Haukoos, Craig D. Newgard, (2007),Advanced Statistics: Missing Data in Clinical Research—Part 1: An Introduction and Conceptual Framework, Acad Emerg Med.
* Dayanıklı A.S., (2021), Kayıp Veri Nedir? Kayıp Veri Analizi Nasıl Yapılır?,Ravenfo.

# Kayıp/Eksik Veri

## Kayıp veri nedir?

Veri setinde kayıp değer olması durumunda yapılan görselleştirmelerin ve analizlerin güvenirliği, tutarlılığı olumsuz etkilenecek, hedef kitleyi pozitif bir şekilde temsil edemeyerek yanlış çıkarımlar yapmamıza neden olacaktır. Aynı zamanda eksik veriler, makine öğrenmesi algoritmalarının sorunsuz çalışmasına engel olacak, sınıflandırma problemlerini yapamayacaktır.

Kayıp veri, bir araştırmada yer alan değişkenlere ait ölçümlerden en az birinin elde edilemediği durumları ifade eder. Yani veri setinde herhangi bir hücrede değer olmaması durumudur.



**Kayıp veri sebepleri:**

* Genellikle bir bireyin belirli bir maddeye yanıt vermemesi
* Bireyin tüm anketi tamamlamaması
* Verilerin yanlışlıkla silinmesi
* Veri toplanan aletlerin eksik ölçüm yapması
* Veri üzerinde yapılan işlemler sırasında yaşanan kayıplar

Gibi nedenlerden dolayı kayıp veriler oluşmaktadır.

Kayıp veriler uygulamalarda “NULL”, ”NaN”, “NA” şeklinde ifadelerle gösterilmektedir.

Kayıp veriler çalışmanın geçerliliğini, güvenirliğini, açıklanabilirliğini olumsuz etkileyip makine öğrenmesi algoritmalarının çalışmasını da olumsuz etkilediği için araştırma sürecinde analizler öncesinde kayıp verilerin giderilmesi gerekir. İlk olarak kayıp verilerin miktarını, türünü belirleyip kayıp veri analizi yapılır veri setindeki kayıp veriler görülür. Sonrasında kayıp verilerin giderilmesi yönetimi için çeşitli yöntemler ve teknikler geliştirilmiştir, bu yöntemler arasında veri atama, eksik gözlemleri analiz dışı bırakma ve modern istatistiksel yöntemler (örneğin, çoklu atama, beklenti maksimizasyonu) bulunmaktadır.

(Akbaş, U., & Koğar, H. (2020))

metin, ekran görüntüsü, yazı tipi, sayı tahtası, skorbord içeren bir resim

Açıklama otomatik olarak oluşturuldu

## Kayıp Veri Analizi

Kayıp veri analizi, bir veri setinde eksik olan gözlemlerin veya değerlerin incelenmesi ve bu eksikliklerin etkilerinin değerlendirilmesi sürecidir. Kayıp veri analizi, kayıp verilerin nedenlerini anlamak, bu verilerin nasıl ele alınacağına karar vermek ve sonuçların geçerliliğini sağlamak amacıyla gerçekleştirilir.

1. **Kayıp verilerin belirlenmesi**

Veri setindeki veriler incelenerek eksik olan veriler belirlenir NaN, NULL veya NA olarak işaretlendiğinden emin olunur. Bu kayıp verinin miktarı belirlenir. Kayıp verilerin miktarını sayısal ve görsel olarak gösterebilirsiniz.

Bir değişken çok sayıda kayıp veriye sahipse bu değişkenin analizde kullanılmasının ne kadar uygun olup olmadığının kararını vermek gerekir.

Kayıp veriler görsel hale getirilerek analizi daha rahat yorumlayabilmemize fayda sağlar.

(Dayanıklı, 2021)

## Kayıp Verilerin Görselleştirilmesine Örnekler IMG_256 ekran görüntüsü, dikdörtgen, çizgi, metin içeren bir resim Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

1. **Kayıp veri Türlerinin Belirlenmesi**

Kayıp veriler MCAR, MAR veya MNAR kategorilerine ayrılabilir. Bu sınıflandırma, hangi yöntemlerin kullanılacağını etkiler.

* 1. **MCAR**:

Verilerin kaybı, gözlemler arasında herhangi bir sistematik ilişki olmaksızın tamamen rastgele gerçekleşir. Örneğin, bir anketin bazı sorularının yanıtlanmaması, katılımcının o soruyu atlaması nedeniyle olabilir.

**2.2 MAR :**

Kayıp veriler, gözlemlenen diğer değişkenlerle ilişkili olabilir, ancak kayıp olan verilerle doğrudan ilişkili değildir. Örneğin, bir anketin belirli bir sorusunu yanıtlamayan kişilerin, o soruyla ilgili belirli bir özellikleri (örneğin, yaş) olabilir.

**2.3 MNAR:**

Kayıp veriler, gözlemlenen ve gözlemlenmeyen değişkenler arasında bir ilişki vardır. Yani, kayıp olan verilerin durumu, gözlemlenen verilerle doğrudan ilişkilidir. Örneğin, gelir düzeyi yüksek olan kişilerin, gelirle ilgili soruları yanıtlamaması MNAR durumuna örnek teşkil edebilir.

(Jason S. Haukoos, Craig D. Newgard, (2007))

Kayıp veri türlerinin belirlenmesinde çeşitli yöntemler uygulanır. Bunlardan en kullanışlı olanı kayıp verilere ilişkin korelasyon analizi kayıp veriler arasındaki ilişkiye bakılarak türüne (Rastgele ilişkisiz mi yoksa ilişkili bir kayıp veri mi olduğuna ) dair bir yorum yapılır ve buna göre veri analizine o değişkenin dahil edilip edilmeyeceği, kayıp verilerin doldurulup doldurulmayacağına, çözüm yöntemine yönelik bir çıkış yoluna başvurulur.

1. **Kayıp verilere çözüm yöntemleri**

Ünlü istatistikçi Ronald Fisher “kayıp verilerle baş etmenin en iyi yolu hiç kayıp veriye sahip olmamaktır” demiştir.

Verilerimizde doğru analiz yapabilmek için kayıp verilere sahip olmamak eğer varsa kurtulmak gerekir. Kurtulma yöntemlerinin hangisinin doğru olduğuna karar vermek için öncelikle kayıp verilerin sayısına ve türlerine bakılıp kayıp verinin bulunduğu değişkenin önemine bakılmalıdır. Bu doğrultuda kayıp verinin olduğu sütun veya satır silinerek veya kayıp veriler çeşitli yöntemlerle tahmine dayalı doldurularak kurtulma arayışlarına girilmelidir.

* 1. **Silme Yöntemleri**

Kayıp verileri silme, eksik gözlemleri analiz dışı bırakma yöntemidir ve genellikle iki ana yaklaşım altında sınıflandırılır: **liste bazında silme** ve çi**ftler bazında silme.**

**3.1.1 Liste Bazında Silme (Listwise Deletion)**:

Bu yöntemde, bir veya daha fazla kayıp veri içeren tüm gözlemler analizden çıkarılır. Yani, bir gözlemde herhangi bir eksik veri varsa, o gözlem tamamen analiz dışı bırakılır.

**Avantajları**: Basit ve anlaşılır bir yöntemdir. Kayıp verilerin neden olduğu yanlılıkları azaltabilir.

**Dezavantajları**: Kayıp veri oranı yüksek olduğunda, örneklem büyüklüğünü önemli ölçüde azaltabilir ve bu da istatistiksel gücü düşürebilir. Ayrıca, kayıp verilerin rastgele olmaması durumunda yanlı sonuçlara yol açabilir , .

**3.1.2 Çiftler Bazında Silme (Pairwise Deletion)**:

Bu yöntemde, her bir analiz için gerekli olan verilerin bulunduğu gözlemler kullanılır. Yani, bir değişkenin kayıp olduğu durumlarda, yalnızca o değişkenin eksik olduğu gözlemler analiz dışı bırakılır.

**Avantajları**: Daha fazla veri kullanarak analiz yapma imkanı sağlar, bu da örneklem büyüklüğünü korur.

**Dezavantajları**: Farklı analizlerde farklı gözlemler kullanıldığı için sonuçların tutarlılığı etkilenebilir. Ayrıca, bu yöntem de kayıp verilerin rastgele olmaması durumunda yanlılık yaratabilir .

Kayıp verileri silme yöntemleri, genellikle veri setinin büyüklüğüne, kayıp veri oranına ve kayıp verilerin nedenine bağlı olarak seçilir. Ancak, bu yöntemlerin yanlılık ve örneklem temsili üzerindeki etkileri göz önünde bulundurulmalıdır. Araştırmacılar, kayıp verileri silerken dikkatli olmalı ve alternatif yöntemleri de değerlendirmelidir.

* 1. **Değer Atama Yöntemleri**

Kayıp verilerin yerine tahmin edilen değerlerin atanması amacıyla kullanılan çeşitli tekniklerdir. Bu yöntemler, kayıp verilerin etkisini azaltmak ve analizlerin geçerliliğini artırmak için tercih edilir.

**3.2.1 Seri Ortalaması (Mean Imputation)**:

Tüm gözlemler için belirli bir değişkenin ortalaması hesaplanır ve kayıp verilerin yerine bu ortalama değer atanır.

**Avantajları**: Uygulaması kolaydır ve hızlı bir çözüm sunar.

**Dezavantajları**: Verinin dağılımını bozabilir ve varyansı azaltabilir, bu da yanlı sonuçlara yol açabilir .

**3.2.2 Yakın Noktaların Ortalaması (Nearest Neighbor Mean Imputation)**:

Kayıp verinin yakınındaki gözlemlerin ortalaması alınarak eksik değer yerine atanır.

**Avantajları**: Daha gerçekçi tahminler sağlar çünkü benzer gözlemler dikkate alınır.

**Dezavantajları**: Hesaplama açısından daha karmaşık olabilir ve yakın komşuların seçimi yanlılık yaratabilir .

**3.2.3 Yakın Noktaların Medyanı (Nearest Neighbor Median Imputation)**:

Kayıp verinin yakınındaki gözlemlerin medyanı kullanılarak eksik değer doldurulur.

**Avantajları**: Aykırı değerlerden daha az etkilenir ve daha sağlam bir tahmin sağlar.

**Dezavantajları**: Yine, komşu seçimi sorunları yaşanabilir .

**3.2.4 Doğrusal Değer Kestirimi (Linear Interpolation)**:

Kayıp veri, eksik veriden önceki ve sonraki tam gözlemler arasındaki doğrusal ilişkiye dayanarak tahmin edilir.

**Avantajları**: Zaman serisi verileri için etkili bir yöntemdir.

**Dezavantajları**: Sadece doğrusal ilişkilerde geçerlidir ve karmaşık veri yapılarında yanıltıcı olabilir .

**3.2.5 Noktanın Doğrusal Eğimi (Linear Trend Extrapolation)**:

Kayıp veri yerine, mevcut verilerin eğilimine uygun bir değer atanır.

**Avantajları**: Verinin eğilimlerini dikkate alarak daha doğru tahminler yapabilir.

**Dezavantajları**: Eğilimlerin yanlış yorumlanması durumunda hatalı sonuçlar verebilir .

**3.2.6 Beklenti Maksimizasyonu (Expectation-Maximization, EM)**:

Kayıp verilerin olasılık dağılımı tahmin edilerek, eksik veriler için en iyi tahminler yapılır. Bu yöntem, iteratif bir süreçle çalışır.

**Avantajları**: Kayıp verilerin yapısını dikkate alarak daha doğru tahminler sağlar.

**Dezavantajları**: Hesaplama açısından karmaşık ve zaman alıcı olabilir.

**3.2.7 Regresyon Atama:**

Regresyon ataması, çoklu atama yönteminden sonra en yakın değerleri veren yöntemlerden biri olarak belirtilmiştir. Bu yöntem de kayıp veri oranı düşük olduğunda etkili sonuçlar vermektedir.

**Avantajları:** veri seti içerisindeki değişkenlerin birbiriyle olan ilişkisini de göz önünde bulundurur.Atanan değerler birbirinden farklı olacağı için değişkenin standart sapma değerini etkilemez ve dağılımını bozmaz.

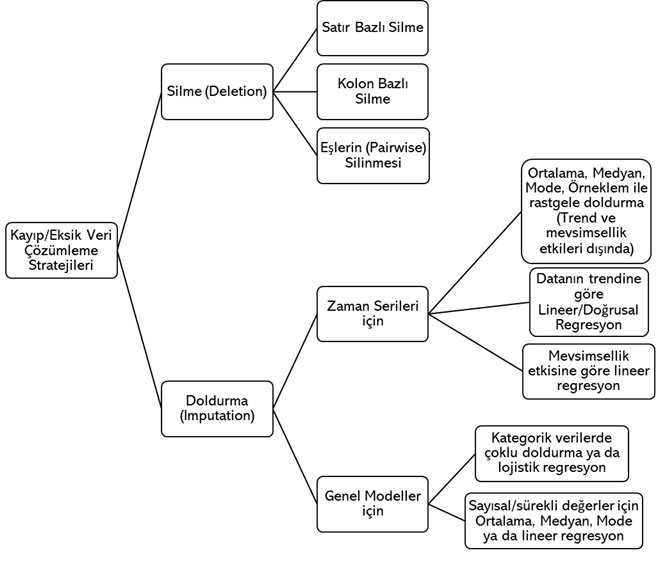
**Dezavantajları:** Atanan değerler diğer değişkenlerdeki gözlemlerden tahmin edildiği için modelde kullanılan değişkenlerle kayıp verilere sahip değişken arasında yüksek korelasyon oluşur. Bu durum değerlerdeki belirsizliği düşürür.

**3.2.8 Çoklu Atama (Multiple Imputation)**:

Kayıp veriler için birden fazla tahmin yapılır ve bu tahminler birleştirilerek nihai sonuçlar elde edilir. Bu yöntem, belirsizliği de dikkate alır.

**Avantajları**: Yanlılığı azaltır ve daha güvenilir sonuçlar sağlar.

**Dezavantajları**: Uygulaması karmaşık ve zaman alıcıdır



(Kürşad,M.,& Nartgün,Z.(2015))

1. **Performans Değerlendirmesi**

Her veri seti farklı özelliklere sahiptir bu nedenle veri setlerine göre uygulanan kayıp veriden kurtulma yöntemlerinin başarısı değişebilmektedir.Hangi kayıp veri yönteminin daha etkili olduğuna karar vermek için yöntemler bulunmaktadır.

* 1. **Model Uygunluğu**

Uygulanan yöntemlerin model uygunluğu (örneğin, AIC, BIC, R-Kare gibi kriterler) değerlendirilerek, hangi modelin veriye daha iyi uyduğuna karar verilebilir. Daha düşük AIC veya BIC değerleri, daha iyi bir model uyumunu gösterir. R-Kare değeri ise modelin doğruluğunu açıklar ne kadar yüksekse modelin uyumunun o kadar iyi olduğunu gösterir.

* 1. **Simülasyon Çalışmaları**

Farklı yöntemlerin performansını değerlendirmek için simülasyon çalışmaları yapılabilir. Bu çalışmalarda, çeşitli senaryolar altında (farklı kayıp oranları ve veri yapıları) yöntemlerin sonuçları karşılaştırılır.

* 1. **Geçerlik ve Güvenirlik Analizleri**

Yöntemlerin geçerlik (ölçeğin faktör yapısının doğruluğu) ve güvenirlik (ölçeğin tutarlılığı) değerleri karşılaştırılmalıdır. Yüksek güvenirlik katsayıları (örneğin, Cronbach alfa) ve geçerli faktör yapıları, yöntemin etkinliğini gösterir.

* 1. **Literatür Taraması**

Önceki araştırmaların bulguları ve önerileri, hangi yöntemlerin daha etkili olduğu konusunda bilgi sağlayabilir. Literatürdeki benzer çalışmaların sonuçları, karar verme sürecinde rehberlik edebilir .

Bu kriterler ve yöntemler, araştırmacıların hangi kayıp veri yönteminin daha etkili olduğunu belirlemelerine yardımcı olur. Her durumda, spesifik araştırma bağlamı ve veri setinin özellikleri de dikkate alınmalıdır.

(Dayanıklı A.S., (2021), Kürşad,M.,& Nartgün,Z.(2015))

**Örneğin;**

Kürşad M. ve Nartgün Z. 2015 yılında yaptığı çalışmada farklı büyüklüklerde (%5, %10 ve %20) kayıp veri oranları altında oluşturulan veri setleri kullanmıştır. Bu veri setleri, kayıp veri yöntemlerinin performansını değerlendirmek amaçlanmıştır.

Bu amaç doğrultusunda silme ve atama yöntemleri uygulanmış ve bahsettiğimiz performans değerlendirme yöntemleri kullanılarak bir sonuca varılmıştır.

Sonuç olarak, çoklu atama yöntemi, kayıp veri sorununu çözmek için en etkili yöntem olarak öne çıkmaktadır. Beklenti maksimizasyonu ve regresyon ataması yöntemleri de etkili alternatiflerdir, ancak çoklu atama yöntemi, özellikle yüksek kayıp veri oranlarında daha iyi sonuçlar sunmaktadır.

## Kaynakça

* Akbaş, U., & Koğar, H. (2020). Nicel Araştırmalarda Kayıp Veriler ve Uç Değerler. Ankara: Pegem Akademi Yayıncılık.
* Kürşad,M.,& Nartgün,Z.(2015).Kayıp Veri Sorununun Çözümünde Kullanılan Farklı Yöntemlerin Ölçeklerin Geçerlik ve Güvenirliği Bağlamında Karşılaştırılması.Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,2015; 6(2); 254-267
* Jason S. Haukoos, Craig D. Newgard, (2007),Advanced Statistics: Missing Data in Clinical Research—Part 1: An Introduction and Conceptual Framework, Acad Emerg Med.
* Dayanıklı A.S., (2021), Kayıp Veri Nedir? Kayıp Veri Analizi Nasıl Yapılır?,Ravenfo.