# VERİ ANALİZ YÖNTEMLERİ

*Yrd. Doç. Dr. Hüseyin BAYRAKTAR [1], Dursun Yıldırım BAYAR [2], Ömer Faruk ERİŞ [3], Selami SUNGUN [4]*

[1] Coğrafi Bilgi Sistemleri Genel Müdürlüğü, 06530, Çankaya, Ankara, huseyin.bayraktar@csb.gov.tr
[2] Coğrafi Bilgi Sistemleri Genel Müdürlüğü, 06530, Çankaya, Ankara, dyildirim.bayar@csb.gov.tr
[3] Coğrafi Bilgi Sistemleri Genel Müdürlüğü, 06530, Çankaya, Ankara, omerfaruk.eris@csb.gov.tr
[4] Coğrafi Bilgi Sistemleri Genel Müdürlüğü, 06530, Çankaya, Ankara, selami.sungun@csb.gov.tr

## ÖZET

*Ülkemizde akıllı şehir politikalarına ulusal katmanda bütüncül bir bakış açısı getirerek ulusal politikalarla uyumlu şekilde yatırımları güvence altına almak amacıyla 2020-2023 Ulusal Akıllı Şehirler Strateji ve Eylem Planı hazırlanmıştır. 2020-2023 Ulusal Akıllı Şehirler Stratejisi ve Eylem Planı kapsamında tanımlanan eylemlerin, görev ve sorumlulukların gerçekleştirilmesine ulusal ölçekte katkı sağlanması ve başta yerel yönetimlerimiz olmak üzere tüm paydaşların kapasitesinin artırılması amacıyla "Akıllı Şehirler Kapasite Geliştirme ve Rehberlik Projesi" T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Coğrafi Bilgi Sistemleri Genel Müdürlüğü tarafından hayata geçirilmiştir. Proje kapsamında hazırlanan akıllı şehir külliyatında "veri analiz yöntemleri" konusu kapsamlı bir şekilde ele alınmış, bu konuda veri analiz yöntemleri eğitim kitabı, video ve sunumlar hazırlanmıştır.*

**Anahtar Sözcükler:** Akıllı Şehirler, Veri Analiz Yöntemleri, Stratejik Yönetim

## ABSTRACT

## DATA ANALYSIS METHODS

*The 2020-2023 National Smart Cities Strategy and Action Plan has been prepared in order to assure investments in line with national policies by bringing a holistic perspective to smart city policies at the national level in our country. Smart Cities Capacity Building and Guidance Project was implemented by the General Directorate of Geographic Information Systems of the Ministry of Environment, Urbanization and Climate Change, in order to contribute to the realization of the actions, duties and responsibilities that are defined within the scope of the 2020-2023 National Smart Cities Strategy and Action Plan, and to increase the capacity of all stakeholders, especially municipalities. In the smart city collection prepared within the scope of the project, the subject of "data analysis methods" have been comprehensively addressed, and a data analysis methods training book, videos and presentations have been prepared on this subject.*

**Keywords:** Smart Cities, Data Analysis Methods, Strategical Management

## 1. GİRİŞ

Akıllı şehir politikalarına ulusal katmanda bütüncül bir bakış getirerek birlikte çalışabilme yetisi kazanmak, belirlenen politikalarla uyumlu yatırımları önceliklendirerek yatırımların doğru proje ve faaliyetlerle uygulandığını güvence altına almak amacıyla ulusal ihtiyaçları ve öncelikleri bütüncül olarak göz önünde bulunduran, ekosistem paydaşlarının ortak aklı ile inşa edilen 2020-2023 Ulusal Akıllı Şehirler Stratejisi ve Eylem Planı hazırlanmıştır.

Hizmet alan kitlelerin yaşam kalitesinin artırılmasında gerek şehir genelinde bulunan algılayıcılardan gerekse diğer veri sağlayıcılardan toplanan verilerin analiz edilmesinin önemli bir yeri vardır. Bu çalışmada veri analiz yöntemlerine iki açıdan yaklaşılmıştır: Birincisi geleneksel veri analizlerinin yapıldığı "istatistiksel veri analizleri", ikincisi ise, makine öğrenmesi yöntemlerine dayalı olan "Veri Madenciliği" yöntemidir.

Hizmet alan kitlelerin yaşam kalitesinin artırılmasında, kurumsal seviyede toplanan verilerin etkili olması beklenir. Bu kaliteyi iyileştirmek, ilgili kurumların gerektiğinde verileri kullanma nedenlerinin en önemlisidir. Diğer yandan, toplanan verilerden beklenen amaçlara uygun sonuçların elde edilmesi sağlanmalıdır. Veri analizleri bu noktada devreye girmektedir. Veri analizleri, toplanan veri ya da "büyük veriden" kurumun amaçlarına uygun çıktıları üretmek üzere ortaya atılan bilimsel esaslara dayalı faaliyetleri içerir.

## 2. VERİ VE BÜYÜK VERİ

IDC'nin araştırmalarına göre, günümüzde 5 milyardan fazla tüketici her gün verilerle etkileşim kurmaktadır. 2025'e kadar, her insanın her 18 saniyede en az bir etkileşimi olacaktır. Bu etkileşimlerin çoğu, 6 milyar veya dünya nüfusunun %75'i olabileceği öngörülmektedir. 2025'teki bu etkileşimler, internete bağlı dünyada ağa milyarlarca IoT cihazının bağlı olmasından kaynaklanacaktır. Bu cihazların, 2025'te 90 ZB'tan fazla veri oluşturması beklenmektedir (IDC, 2020). Veri hazırlığından veri analizi modellerini mükemmelleştirmeye kadar, önümüzdeki on yılın en önemli işleri veriler, verilerin yönetimi ve korunması, yönetişimi ve gelir elde edilmesi, analizi ve karar vermedeki rolü ile ilgili olacağı tahmin edilmektedir.

Analistlere göre veri ile ilgili işlerin kurulması ve yaygınlaşmasına ek olarak, "veri okuryazarlığı" birçok kuruluştaki tüm çalışanlar için iç eğitimin ana odak noktası haline gelecektir. Hemen hemen tüm ürün ve hizmetler ya veriye dayalı olacak ya da bir veri bileşenine sahip olacaktır ve geliştiricileri, yöneticileri ve satıcıları veri konusunda yetkin olmak zorunda bırakacaktır (Press, 2020). Veri miktarının gittikçe büyük rakamlarla artma eğiliminde olması veri analizlerinin de boyut değiştirmesine neden olmaktadır. Küçük veri kümeleriyle yapılan istatistiksel analizler, büyük veri karşısında amacına ulaşmada yetersiz kalmaktadır. Bu nedenle, söz konusu büyük veri analizlerinde yapay zekâ, makine öğrenmesi ve veri madenciliği gibi alanlara başvurulmaktadır.

Basit bir tanım yapmak gerekirse, genel olarak veri, toplanan ve bazı amaçlar için dönüştürülen, çoğunlukla analiz amaçlı kullanılan herhangi bir karakter kümesidir. Bunlar metin ve sayı, resim, ses veya video da dahil olmak üzere herhangi bir karakter olabilir. Veri türlerini farklı biçimlerde tanımlamak mümkündür. Birincisi; veri türleri arasında yer alan yapısal veridir. Özellikle tablo halinde gösterilebilen veriyi tanımlar. Örneğin Excel uygulamaları içinde yaratılan tablolar yapısal veri türlerine en uygun örneklerdir. İkincisi; yarı yapısal verilerdir. Özellikle bilgisayar uygulamalarında tercih edilen ve kod ile verinin birlikte oluşturulduğu veri türleridir. Bu veri türlerine en uygun XML verileri örnek olarak gösterilebilir. Üçüncüsü; yapısal olmayan veri türlerdir. Sosyal medya verisi, metinler ve bunlar gibi verileri temsil eder.

Bu veriler yer, zaman ve öznitelikler (yani açıklayıcı bilgiler) arasın - da bağlantı kurar. Bu tür veriler, şehir planlamasında ve çevresel etkilerin izlenmesinde analistlere yardımcı olur (Zuhair, 2016). Büyük veri, var olan geleneksel uygulamalarla etkili bir şekilde işlenemeyen büyük boyutlara ulaşmış veriyi belirtir. Büyük verilerin işlenmesi, toplanmamış ham verilerle başlar ve genellikle tek bir bilgisayarda depolanması olanaksızdır.

Büyük veri, daha isabetli kararlar vermek ve stratejik iş eylemlerine yol açabilecek bilgileri analiz etmek için kullanılır. Açık veri, araştırmacıların veriye kolayca ulaşabilmeleri ve ilgili alanda analizlerini yapabilmeleri için erişimin sağlanmasıdır. Böyle bir amaca ulaşmak için de açık veri platformlarının desteklenmesi gerekir. Bu tür yenilikçi yaklaşımın ülkedeki araştırma kültürüne katkıda bulunması beklenir. Böylece açık hale getirilen verinin analiz edilmesi için gerekli farkındalığın oluşturulması ve kolaylaştırıcı bir ortamın sağlanması mümkün olacaktır (T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, 2020, s:591).

## 3. VERİ ANALİZİ

Veri analizi, faydalı sonuçlar çıkarmak için verileri toplama ve düzenleme sürecidir. Veri analizi süreci, verilerden bilgi elde etmek için analitik ve mantıksal akıl yürütmenin kullanılmasıdır. Veri analizinin temel amacı, elde edilen bilginin bilinçli kararlar vermede kullanılabilmesi için verilerde anlam aramaktır. Özellikle günümüzde açık veriye dayalı açık veri platformlarının kurulması veri analizi çalışmalarının gerçekleştirilmesi açısından gerekli görülmektedir.

Bu açıdan bakıldığında özellikle ülkemizdeki Ulusal ve Yerel Akıllı Şehir Açık Veri Platformları üzerinde büyük veri ve veri analizi çalışmalarının yapılmasına imkân sağlayan araçların yer almasının planlanması önemli bir yaklaşımdır (T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, 2020, s:583). Veri analizleri ülkemiz üniversitelerinin birçoğunda müfredatlara girmiştir. Ancak bu konuda özellikle açık veri, büyük veri, yapay zekâ, makine öğrenmesi gibi alanlarındaki eğitimlerin yaygınlaştırılması gerekli görülmektedir. Eğitim sistemindeki kurum ve kuruluşların veri analiz kabiliyetinin güçlendirilmesi, okul bazında veriye dayalı planlama ve yönetim sisteminin hayata geçirilmesi hedeflenmelidir (T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, 2020, s:374).

Veri analiz yöntemlerine iki farklı açıdan yaklaşılmaktadır. Bunlardan birincisi geleneksel veri analizlerinin yapıldığı "istatistiksel veri analizleri" alanıdır. Bir diğeri ise makine öğrenmesi yöntemlerine dayalı olan "Veri Madenciliği" adıyla son yıllarda rağbet gören bir alandır. Kitapçıkta istatistiğin bazı veri analiz yöntemlerine değinilmektedir. Ardından veri madenciliğinde hangi süreçlerin izlenerek ne gibi analizlerin yapıldığı üzerinde durulmaktadır.

## 3.1 İstatistiksel Veri Analizleri

İstatistiksel veri analizlerinin gerçekleştirilebilmesi için öncelikle bu analizlerin gerektirdiği veri kümelerinin oluşturulması gerekmektedir. Örneğin şehrin ekonomisini oluşturan mevcut varlık ve faaliyetlerin yönetimi kapsamında yapılacak veri analizine dayanak teşkil etmesi açısından istatistiki bilgilerin temin edilmesi önem taşımaktadır (T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, 2020, s:344) İstatistiksel veri analizi, çeşitli istatistiksel işlemleri gerçekleştirme sürecidir. Verileri anlamlandırmak ve bazı çıkarımlar yapmak üzere istatistiksel analizlerden yararlanılır. Bu analizlerin yapılabilmesi için aşağıdaki bölümlerde istatistiğin merkezi eğilim ölçüleri, yayılım ölçüleri, korelasyon analizleri, regresyon analizleri ve zaman serileri analizleri ele alınarak incelenmektedir.

## 3.2 Merkezi Eğilim Ölçüleri

Merkezi eğilim ölçüsü, bir veri kümesinin merkez noktasını veya ortasını temsil eden bir özet istatistiktir. İstatistikte yaygın olarak merkezi eğilim ölçülerinden söz edilebilir. Bunlar arasında aritmetik ortalama, medyan, mod gibi kavramlar en ön sırada yer alır.

## 3.3 Aritmetik Ortalama

Aritmetik ortalama, gözlem değerlerinin toplamı hesaplandıktan sonra bu toplam değerin gözlem sayısına bölünmesi suretiyle hesaplanır. Gözlemlerin her biri X1 ,X2 ,…,XN biçiminde ifade edilirse, X ortalama formülü şu şekilde karşımıza çıkar:

## 3.4 Medyan

Merkezi eğilim ölçüleri içinde yer alan bir diğer kavram medyan olarak isimlendirilir. Bir dizinin meydanını bulmak için önce dizi küçükten büyüğe doğru sıralanır ve elde edilen yeni dizinin ortadaki değeri medyan olarak belirlenir. Eğer ortada iki değer varsa onların aritmetik ortalaması medyan olarak hesaplanır.

## 3.5 Mod

Mod, bir dizinin elemanları arasında hangilerinin en çok tekrarlı olduğunu ortaya koyan bir ölçüttür. Her dizide tekrarlı değer yok ise mod değeri elde edilemez.

## 3.6 Yayılım Ölçütleri

Bir sayısal değerin bir ortalama etrafındaki yayılma eğilimi istatistikte önemli bir kavrama işaret eder. Sözü edilen bu yayılma eğilimlerine o verinin yayılımı adı verilir. Yayılım ölçüleri arasında en 23 24 Akıllı Şehirler Veri Analiz Yöntemleri önemlileri ortalama sapma, standart sapma ve varyans kavramları sayılabilir. 4.1.2.1. Ortalama Sapma Ortalama sapma, dizi elemanlarının her birinin ortalamadan sapmalarının ortalamasıdır. Bu tanıma dayalı olarak, X dizinin ortalaması olmak üzere, ortalama sapma aşağıdaki formüle göre hesaplanır.

## 3.7 Standart Sapma

Bir veri dizisi içindeki gözlemlerin her birinin aritmetik ortalamadan sapmalarının kareli ortalaması standart sapma olarak isimlendirilir. Bu ölçüt ss standart sapmayı ve X dizi değerlerinin ortalamasını belirtmek üzere, aşağıdaki gibi formüle edilir.

## 3.8 Varyans

İstatistiksel analizlerde yaygın biçimde bir yayılım ölçütü olarak kullanılan bir diğer kavram varyanstır. Varyans bir veri kümesinin standart sapmasının karesi olarak tanımlanır.

## 3.9 Korelasyon Analizi

İstatistik analizlerde kullanılan önemli kavramlar arasında yer alan korelasyon analizleri iki değişken arasındaki ilişkinin derecesini belirlemede kullanılır. Korelasyon analizlerini yapabilmek için öncelikle korelasyon katsayısı adı verilen bir ölçütün hesaplanması gerekir. Bu katsayının birçok türü olmasına karşılık en çok tercih edileni Pearson korelasyon katsayısıdır.

## 3.10 Regresyon Analizi

Korelasyon araştırması yardımıyla iki değişken arasında bir ilişkinin olup olmadığı yukarıda izah edildiği biçimde ortaya konmaktadır. Bağımsız değişkenlerdeki değişmelerden yararlanılarak bağımlı değişkenin ne ölçüde değişeceği de tahmin edilebilmektedir. Bir değişkenden edinilen bilgi yardımıyla diğer değişken için tahmin yapılmak istenirse buna regresyon analizi adı verilir. Regresyon analizleri, geçmiş dönemlere ilişkin verilerin kullanılarak bu verilerin gelecekte ne gibi sonuçlara neden olabileceğini tahmin etmek üzere kullanılır.

## 3.11 Regresyon Doğrusu

Regresyon analizlerini yapabilmek için x bağımsız değişken ile y bağımlı değişken arasındaki matematiksel model bir doğru denklemi biçiminde ortaya konulur. Bu ifadeye sadece bir bağımsız değişken dahil edildiği için basit regresyon doğrusu adı verilir.

## 3.12 Zaman Serileri Analizi

Zaman serisi verileri, aynı değişkenin belirli bir zaman içindeki seyrini gösteren veri kümeleridir. Bir başka deyişle zaman serileri istatistiki verilerin oluş zamanları esas alınarak sıralanması sonucu elde edilen verilerdir. Zaman serilerinin hazırlanmasındaki temel amaç, bir değişkenin geçmişte gösterdiği eğilimleri açıklamak ve bu bilgileri kullanarak gelecekteki davranışını tahmin etmektir.

## 3.13 Zaman Serisi Görselleştirme

Zaman serilerinde, zaman da bir değişken olarak dikkate alınır. Zaman bilgisi dışında, zaman içinde değişime uğrayan bir değişken kullanılır.

## 3.14 Veri Madenciliği

Kurumların ürettiği verinin boyutu dijitalleşmenin yaygınlaşması ile birlikte daha çok büyümeye devam etmektedir. Geleneksel istatistik veri analizleri ile büyük boyuttaki verinin analizi yeterli olarak yapılamamaktadır ve makine öğrenmesi adı verilen yöntemlerle çözümler aranmaktadır. Veri madenciliği, büyük ölçekli veri kümeleri içinden değerli olan bir bilgiyi elde etme süreci olarak tanımlanabilir. Veriler arasındaki ilişkileri ortaya koymak ve ileriye yönelik öngörüler geliştirmek üzere veri madenciliği yöntemlerinden yararlanılır.


**Şekil 1:** Veri Madenciliğinin Yapay Zeka ve Makine Öğrenmesi İle Olan İlişkisi

Yapay Zekâ, insan zekâsına özgü algılama, öğrenme, düşünme, sorun çözme ve öngörme gibi yeteneklerin bir yapay sisteme kazandırılma çalışmalarıdır. Makine Öğrenmesi, bir sistemin veriye dayalı olarak öğrenmesini sağlayan algoritmalar topluluğudur. Bu algoritmalar, sistemin eğitilmesini ve öngörü yapabilecek seviyeye gelmesini amaçlar. Veri Madenciliği, makine öğrenmesi algoritmalarını kullanarak büyük veri kümelerindeki gizli örüntüleri açığa çıkarma ve öngörüler geliştirme sürecidir. Büyük Veri üzerinde uygulanan veri madenciliği süreçlerine Büyük Veri Madenciliği denmektedir.

## 4. SONUÇLAR

Hizmet alan kitlelerin yaşam kalitesinin artırılmasında gerek şehir genelinde bulunan algılayıcılardan gerekse diğer veri sağlayıcılardan toplanan verilerin analiz edilmesinin önemli bir yeri vardır. Bu çalışma, akıllı şehirler için önemli bir yer tutan modern veri analizlerine dikkat çekmek ve bir farkındalık yaratmak amacıyla hazırlanmıştır.

## KAYNAKLAR

**IDC,** (2020), "Data Age 2025", (https://www.seagate.com/files/wwwcontent/our-story/trends/files/dataage-idc-report-final.pdf),

**Press, G.,** (2020), "6 Predictions About Data In 2020 And The Coming Decade",

**T.C. Çevre Şehircilik ve İklim Değişikliği Bakanlığı,** 2021. Veri Analiz Yöntemleri / https://www.akillisehirler.gov.tr/egitim-veri-analiz-yontemleri/

**T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı,** (2020), "T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı 2020 -2023 Ulusal Akıllı Şehirler Stratejisi ve Eylem Planı", Ankara.

**Zuhair, M.,** (2016), "Büyük Veri: Analytics'te Kullanılan Veri Türleri", (http:// blog.agroknow.com/?p=4690)