

## BÜYÜK VERİ

Yapay zekâ teknolojisinin altında büyük veri (big data) yatıyor. Verinin gerçek zamanlı işlenebilmesi, bu datanın farklı kanallardan gelmesi, kendi kendine öğrenen teknolojilere sahip uygun ürünlere ulaşabilmesini sağlamak önemlidir. Artık kural bazlı statik teknolojiler yetmiyor. Yapay zekâyı bu yüzden teknoloji açısından değer yaratacak şekilde kullanıyoruz

Büyük Veri (Big Data), geleneksel veri işleme yazılımlarının işlemek, depolamak ve analiz etmekte zorlandığı büyüklükte ve karmaşıklıkta veri kümelerini ifade eder. Büyük veri, hacmi (volume), çeşitliliği (variety) ve hızı (velocity) bakımından geleneksel verilerden ayrılır.

Volume (Hacim): Büyük verinin miktarını ifade eder. Çok büyük veri kümelerinden bahsedilir.

Variety (Çeşitlilik): Farklı türde ve biçimlerde verilerin bulunmasını ifade eder. Yapılandırılmış, yapılandırılmamış ve yarı yapılandırılmış veriler bu kapsamdadır.

Velocity (Hız): Verinin ne kadar hızlı üretildiğini ve işlendiğini ifade eder. Gerçek zamanlı veri akışları bu kapsamda değerlendirilir.

Veracity (Doğruluk): Verinin güvenilirliğini ve doğruluğunu ifade eder. Verinin kalitesi ve güvenilirliği önemlidir.

Value (Değer): Verinin analiz edilmesiyle elde edilecek potansiyel değeri ifade eder. Büyük veri, doğru analiz edildiğinde önemli değerler sağlar.

Variability (Değişkenlik): Verinin tutarsızlığını ve değişkenliğini ifade eder. Verinin zaman içindeki değişimleri ve dalgalanmaları da dikkate alınmalıdır.

Visualization (Görselleştirme): Verinin anlaşılabilir ve analiz edilebilir hale getirilmesi için görselleştirilmesi önemlidir. Büyük veri genellikle karmaşıktır ve görselleştirme araçları kullanılarak daha anlaşılır hale getirilir.

Volatility (Değişkenlik/Dayanıksızlık): Verinin ömrünü ve ne kadar süre boyunca geçerli olduğunu ifade eder. Bazı veriler hızla eskir ve geçerliliğini yitirir.

Validity (Geçerlilik): Verinin doğruluğu ve iş için uygunluğunu ifade eder. Verinin geçerliliği, verinin amaca uygun olup olmadığını belirler.

## Büyük Verinin Kaynakları

1. Sosyal Medya: Facebook Twitter, Instagram gibi platformlar kullanıcıların oluşturduğu metin, görüntü, video gibi çeşitli veriler sağlar.
2. Sensörler ve 10T (Nesnelerin İnterneti): Akıllı cihazlar ve sensörler sürekli veri üretir. Örneğin, akıllı ev cihazları, sağlık izleme cihazları.
3. İşlem Kayıtları: Finansal işlemler, e-ticaret sitelerindeki alışveriş verileri.
4. Dijital Medya: Haber siteleri, bloglar, dijital kitaplar.
5. Kamu ve Kurumsal Kayıtlar: Hükümet verileri, sağlık kayıtları, müşteri ilişkileri yönetim (CRM) sistemleri.

## Büyük Verinin İşlenmesi ve Analizi

Büyük verinin işlenmesi ve analizi, özel teknolojiler ve yaklaşımlar gerektirir. İşte bazı temel teknolojiler ve yöntemler:

### Dağıtık Depolama ve İşleme:

Hadoop: Büyük veri depolama ve işleme için açık kaynaklı bir çerçevedir. HDFS (Hadoop Distributed File System) ve MapReduce gibi bileşenleri içerir.

Spark: Hızlı ve genel amaçlı dağıtık veri işleme sistemi. Hadoop'a göre daha hızlı ve hafızada işlem yapabilme yeteneğine sahiptir.

### Veri Tabanları:

NoSQL: MongoDB, Cassandra gibi NoSQL veritabanları, büyük verilerin esnek ve ölçeklenebilir bir şekilde depolanmasını sağlar.

NewSQL: Geleneksel SQL veritabanlarının gücünü ve NoSQL'in ölçeklenebilirliğini birleştiren veritabanlarıdır.

### Gerçek Zamanlı Veri İşleme:

Kafka: Gerçek zamanlı veri akışları için kullanılan dağıtık bir yayın-abone sistemi.

Flink: Gerçek zamanlı veri akışları ve toplu veri işleme için açık kaynaklı bir platform.

### Veri Analizi ve Görselleştirme:

R ve Python: Veri analizi ve makine öğrenmesi için yaygın olarak kullanılan programlama dilleri.

Tableau, Power BI: Veri görselleştirme ve iş zekası araçları.