**Makine Öğrenmesi Nedir?**

Her şeyden önce makine öğrenmesinin yakıtı olan veri nedir sorusuna yanıt bulmaya çalışalım.

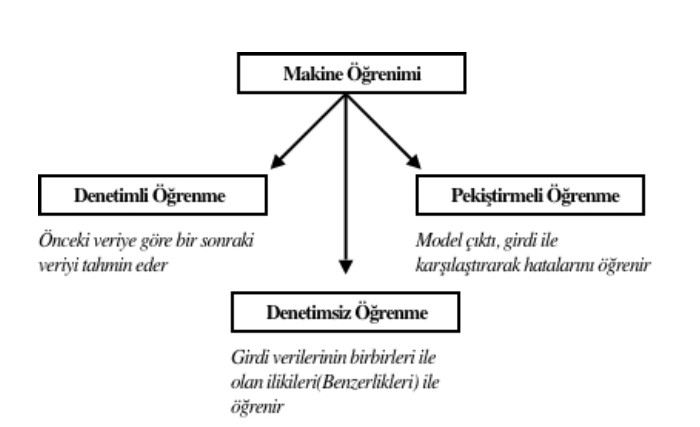
**Veri;**

Faydalı bilgiler içerebilen bilgi topluluğudur. Bu bilgiler ham bir şekilde bulunur ve ismini veriden alan veri bilimciler tarafından bu bilgi yığınını işlenip veriden faydalı bilgi çıkartılabilir, bunu yaparken birçok alandan faydalanırlar biz bu yazımızda elimizdeki veriden yeni verileri tahmin edebilmemizi veriyi gruplamamızı sağlayan makine öğrenmesini işleyeceğiz.

**Makine Öğrenmesi;**

Yapay zekanın bir alt dalı olan makine öğrenmesi bilgisayara veri öğrenimini mümkün kılan bizim için elimizdeki verilerden istatistiki bilgiler yardımıyla tahmin ya da sonuçlar çıkarabilen bir yöntemdir. Bunu yaparken tıpkı insanın öğrenme biçimi gibi önce eldeki bilgilerden bir öğrenim edinir ardından eldeki verilerden yeni bilgiler ortaya çıkartır. Geçmişteki deneyimler gelecekte performansın iyi olmasını sağlar.

Birkaç makine öğrenmesi örneğiyle demek istediğimizi açıklayalım örneğin ünlü satranç ustası Garri Kasparov ile IBM’in geliştirmiş olduğu Deep Blue’nun yapmış olduğu satranç maçlarında Deep Blue Kasparov’un hamlelerin algoritmasını öğrenerek gelecekte oynayacağı maçlarında ne yapacağını tahmin edebiliyor ve ünlü satranç ustasını yenebilecek hale geliyor. Bir başka örneğimiz daha gündelik hayattan olsun örneğin maillerimize gelen spam mesaj olabilir uyarısı alıyoruz gelen mesajın spam olup olmadığını anlarken daha önceden spam mesajları gören makine onları etiketliyor ve daha sonra gelen mesajlarda bu etiketlerin olup olmamasına göre bize spam mesaj uyarısı gönderiyor bir başka örnek Youtube Spotify Instagram gibi dev firmalar sizin daha önceki faaliyetleriniz üzerine öneriler denilen sizin dikkatinizi çekebilecek sizin daha önce dinlemiş izlemiş olduğunuz müziklerden videolardan size bir takım öneriler sunarken aslında tamamen makine öğrenmesi algoritmalarını kullanıyor. Bunları yaparken farklı türde makine öğrenmesi türleri kullanıyoruz. Şimdi makine öğrenmesi türleri nelerdir onlara kısaca bakalım.

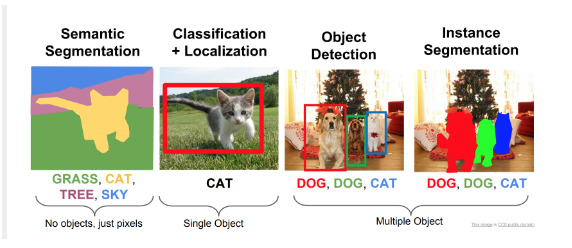
**Makine Öğrenmesi Türleri;**

3 temel türden oluşan makine öğrenmesi Denetimli Denetimsiz ve Pekiştirmeli öğrenmeden oluşur şimdi bu türleri detaylı olarak inceleyip aralarındaki farklara bakalım.

**Denetimli Öğrenme (Supervised Learning)**

Denetimli Öğrenme algoritmasındaki en önemli nokta etiketli bir veri kümesi (labeled dataset) kullanılmasıdır. Bu sayede elimizdeki verileri istediğimiz ölçülerde etiketleyip bir sonraki adımda etiketlemiş olduğumuz verilerden yola çıkarak makinenin daha önce görmediği bilgileri işleyip bizim için birer tahmin yapmasını isteriz. Geliştirdiğimiz algoritma etiketli eğitim verilerini kullanarak optimal bir test verisi tahmin etmeye çalışır. Yani kısaca denetimli bir öğrenme, bilinen bir girdi seti ile bunlara denk gelen çıktıları alıp algoritmanın daha önce hiç görmediği (eğitimde kullanılmayan) yeni verilere en uygun çıktıları üretmek için kullanılan bir makine öğrenmesi modelidir. İki tekniği vardır bunlardan ilki sınıflandırma bir ötekisi de lineer regresyondur. Denetimli öğrenme için birkaç örnek verecek olursak;

Herhangi bir siteye girdiğimizde bir işlem onaylamak için bize makine olup olmadığımızı soruyor ve bunun testi için aşağıdakilerden hangisi köpek hangisi kedi vb. sorular soruyor peki makine o resimlerin köpek veya kedi olduğunu nasıl ayırt edebiliyor, bunu yaparken daha önceden makineye verilmiş olan etiketleri kullanarak tıpkı insana kedi denildiğinde akılda oluşan kedi imgesi gibi bilgisayarda o etiketler yardımıyla onun kedi olduğunu anlıyor bu sınıflandırma türü bir örnektir.



Başka bir denetimli öğrenme örneği verecek olursak Youtube premium yakın zamanda bir fiyat güncellemesi aldı ve hiç kimse daha fazla para ödemek istemez ama Youtube müşteri kaybetmeyi göze alıp bunu neden yaptı? Daha önceki fiyat artışlarından yola çıkan Youtube yeni fiyat artış modeline göre uygun eski fiyat algoritmaları ile yeni fiyat belirlemesi yapmaktadır. Bu da regresyon modeline en basit örneklerden biri olmaktadır.

**Denetimli Öğrenme yöntemleri iki grupta incelenir;**

* Sınıflandırma Yöntemi (Classification Method)
* Regresyon Yöntemi (Regression Method)

# Sınıflandırma Yöntemi (Classification Method)

Eğer sistem, hangi verinin, hangi koşullarda, hangi sınıfa ait olacağı bilgisi ile sınıflandırılarak eğitilirse, yeni veri setindeki veriyi de öğrendiklerine benzer biçimde sınıflandırabilir.

Örneğin: 0-17 yaş aralığındaki kişileri çocuk, 18-25 yaş aralığındaki kişileri genç, 26 yaş ve üstündeki kişileri yetişkin sınıfıyla sınıflandırmak,

**Sınıflandırma Yöntemleri de aşağıdaki gibi kategorize edilir:**

1. **Binary Classification (İkili Sınıflandırma):** İki olası sonuç ile sınıflandırma.

**Örn:** Cinsiyet sınıflandırması (Erkek / Kadın)

1. **Multi Class Classification (Çoklu Sınıf Sınıflandırma):** İkiden fazla sınıfı sınıflandırma. Bir sınıfa ait birden fazla farklı veri varsa bu farklı veriler tespit edilir ve her biri tek bir etikete atanır.

**Örn:** Bir hayvan sınıfında kedi ya da köpek olabilir ancak ikisi birlikte bir sınıfta olamaz kendi içinde sınıflara bölünmelidir.

1. **Multi Label Classification (Çoklu Etiket Sınıflandırma):** Bir veri birden fazla sınıfla ilişkilendirilebilir.

**Örn:** Bir makale hem sağlık hem spor hem de insan ile ilgili olabilir.

# ****Sınıflandırma yönteminde en çok kullanılan algoritmalar:****

1. Naive Bayes: Verileri olasılık ilkeleri ile hasaplayarak sınıflandıran bir sınıflandırma algoritmasıdır. Basit bir ifadeyle, bir Naive Bayes sınıflandırıcı, bir sınıftaki belirli bir özelliğin varlığının başka herhangi bir özelliğin varlığına bağlı olmadığını varsayar. Örneğin, bir meyve kırmızı, yuvarlak ve çapı yaklaşık 3 inç ise bir elma olarak düşünülebilir. Bu özellikler birbirlerine veya diğer özelliklerin varlığına bağlı olsa bile, bu özelliklerin tümü, bu meyvenin bir elma olması olasılığına bağımsız olarak katkıda bulunur ve bu yüzden “Naif” olarak bilinir.
2. K-Nearest Neighbours (En Yakın Komşu)**:**Bu tip sınıflandırma, her bir noktanın en yakın komşularının basit çoğunluk oyu ile hesaplanması ile elde edilen sınıflandırmadır. Veri hangi veriye en çok yakındır? mantığı ile dallanır. Bu algoritmanın uygulanması kolaydır, gürültülü eğitim verisine (noisy training data) dayanıklıdır ve eğitim verileri büyükse oldukça etkilidir.
3. Decision Tree (Karar Ağacı): Veriler, sınıfları ile birlikte bu algoritmaya verildiğinde, algoritma verileri sınıflandırmak için kullanılabilecek bir dizi kural üretir. Karar düğümleri (decision node) ve yaprak düğümleri (leaf node) olan bir ağaç yapısına sahiptir. Hem sınıflandırma hem de regresyon yönteminde kullanılabilir.
4. Random Forest: Sınıflandırma işlemi sırasında birden fazla decision-tree kullanılarak sınıflandırma değerinin yükseltilmesi hedefleyen, sınıflama veya regresyon yönteminde kullanılabilen algoritmadır.
5. Support Vector Machine (Destekçi Vektör Makinesi) : Veri setinde birbirine benzeyen gruplar arasına birbirinden en uzak olan noktalardan sınırlar çizmeye yarayan algoritmadır.

**Regresyon Yöntemi (Regression Method)**

Regresyon problemleri, üretilen çıktının sürekli sayılardan oluştuğu durumlar için kullanılıyor.

Örneğin: bir öğrencinin okula geldiği gün sayısı ve sınavlardan aldığı puan ile bir regresyon algoritması kullanabilirsiniz.

**Regresyon yönteminde en çok kullanılan algoritmalar:**

1. **Linear Regression:**Sayısal girdi ve çıktılar arasındaki doğrusal ilişkiyi tespit etmeyi sağlar. Düzlemde yayılmış verinin modelini en iyi biçimde doğrusal olarak çıkartmaya çalışan yöntemdir.
2. **Logistic Regression:**Bir sonucu belirleyen bir veya daha fazla bağımsız değişken bulunan veri kümesini analiz etmek için düzlemde en iyi eğriyi yakalamaya çalışan istatistiksel bir yöntemdir. Sonuç, ikiye bölünmüş bir değişkenle ölçülür (sadece iki olası sonuç vardır). Bu algoritmanın uygulanması kolaydır, gürültülü eğitim verisine (noisy training data) dayanıklıdır ve eğitim verileri büyükse oldukça etkilidir.
3. **Multiple Linear Regression:** Birden fazla tahminleyici (predictor) değişken kullanarak tahminlemeye çalışılan doğrusal regresyonun adıdır.
4. **Polynomial Regression:**Veriler arası ilişki her zaman doğrusal olmayabilir. Optimum ilişkiyi bulmak için bir eğri gerekebilir. Tıpkı polinom fonksiyonlarında olduğu gibi bu yöntemde de bir terimin karesi veya küpü (veya terimin üssü herhangi bir sayı olabilir) alınarak doğrusal olmayan bir regresyon modeli oluşturulmak istenebilir. Bu gibi durumlarda kullanılabilen bir algoritmadır.
5. Support Vector Regression:Algoritmayı karakterize eden tüm ana özellikleri (maksimal marjı) koruyan bir regresyon yöntemi olarak da kullanılabilir. Support Vector Machine ile aynı ilkeleri kullanır. Ana fikir, hatanın en üst düzeye çıkarıldığı hiper düzlemi bireyselleştirerek hatayı en aza indirgemek, hatanın bir kısmının tolere edildiğini göz önünde bulundurmak.

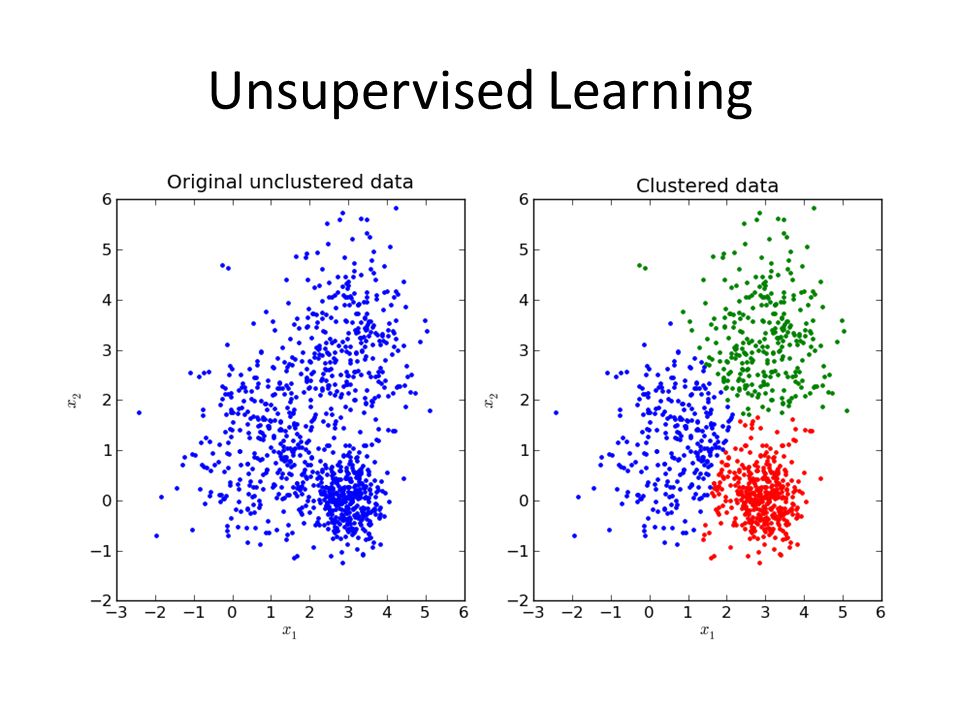
Örn: Bir personelin eğitim seviyesine göre maaşını tahmin eden model geliştirmek.

1. Decision Tree:Sınıflandırma yönteminde de kullanılan decision tree, regresyon yönteminde de aynı şekilde kullanılabilir. Bu algoritma, kök düğümden başlayarak, yukarıdan aşağıya inşa edilen node’lar (düğüm) ile verilerin, kendi içlerinde benzer değerlere (homojen) sahip olanlarının alt kümelere ayrılmasını sağlayan algoritmadır.

**Denetimsiz Öğrenme (Unsupervised Learining)**

Denetimsiz öğrenme herhangi bir algoritma verilemeden veriler arasındaki bağlantıyı ortaya çıkaran bir makine öğrenmesi tekniğidir. Elimizdeki verilerin herhangi bir etiketi veya tanımlayıcı özelliğinin olmadığı durumlarda kullanırız. Denetimsiz öğrenmede veriler yorumlanarak aralarında bir ilişki bulunur ki buna da İlişkilendirme (Association) problemleri denir veya kümeleştirme işlemi Kümeleme (Clustering) yapılır.

“İlişkilendirme algoritması, verileriniz arasındaki birlikteliği ortaya çıkarır. Bu algoritma genelde **Birliktelik Kuralı Madenciliği (Association Rule Mining)** olarak bilinir. İlişkisel veritabanları, işlem veritabanları ve diğer depo biçimleri gibi çeşitli veritabanlarında bulunan veri kümelerinden sık sık meydana gelen kalıpları, bağıntıları veya ilişkileri gözlemlemeyi amaçlayan bir prosedürdür.”



Bir örnek verecek olursak, bir alışveriş sitesinde gezerken bazı ürünlere bakmışızdır o ürünlere bakarken alt kısımlarda da o ürüne benzer ya da tamamlayıcı ürünler görürüz aslında bu denetimsiz öğrenmenin ilişkilendirmesinden dolayı karşımıza çıkan bir durumdur çünkü bizim gibi başka kullanıcılarda aynı ürün veya benzerlerine bakmıştır ve bizim baktığımız ürünler birer veri olur algoritmada bu ürünleri birbirleriyle ilişkilendirir. Makine öğrenmesi kısmında verdiğimiz sosyal medya örneği de aynı temele dayanır.

**Denetimsiz Öğrenme yöntemleri iki grupta incelenir;**

* Kümeleme (Clustering)
* Boyut Azaltma (Dimensionality Reduction)

# Kümeleme (Clustering)

Clusteringden önce önce verilerin standart hale getirilmesi gerekir (bununla ilgili bilgiler Veri Dönüşümü (Data Transformation) bölümünde bulunmaktadır). Kümeleme, bir nevi nokta mesafelerine göre kümeler oluşturma işlemidir.

# Boyut Azaltma (Dimensionality Reduction)

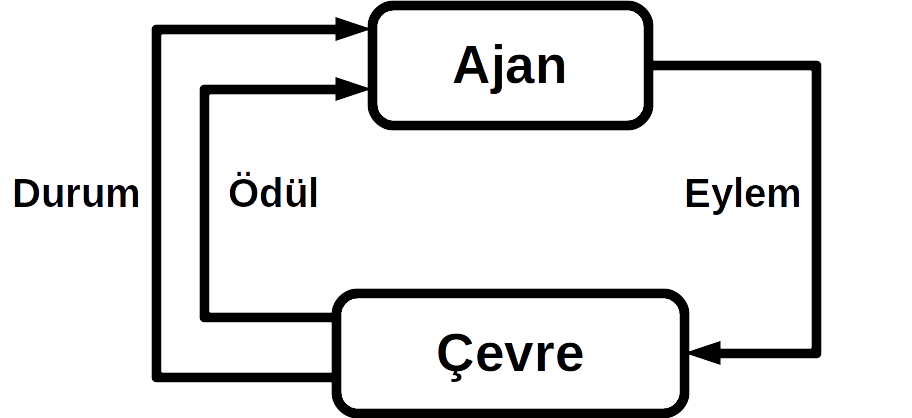
Boyutsal küçültme, her biri en fazla miktarda bilgi içerecek şekilde, verilerin azaltılması işlemidir. Yüksek boyutlu verilerin görüntülenip azaltılarak veya az ilişkili (low-correlation) özelliklerin kaldırılmasıyla makine öğrenme modellerinin hızlandırılması için kullanılır.

**Denetimli ve Denetimsiz Öğrenme Arasındaki Farklar**

İki öğrenme tekniği arasındaki en temel fark etiketli bir veri kümesinin kullanılmasıdır, hangi bilginin hangi bilgiyle eşelenileceği biliniyor ve algoritmanın daha önce görmediği bir veri için en uygun çıktıyı sağlıyor bu kısım denetimli öğrenmeydi. Denetimsiz öğrenmede ise etiketsiz veriler arasındaki ilişkiyi bulmaya çalışır. Denetimli öğrenme genellikle Regresyon ve Sınıflandırma problemlerine uygulanırken, denetimsiz öğrenme Kümeleme (Clustering) ve İlişkilendirme (Association) problemlerine uygulanır. Özetle Denetimli öğrenmede çıktının nasıl olması gerektiği bilinen etiketli veriler kullanılır. Fakat denetimsiz öğrenmede, çıktının nasıl olması gerektiği etiketsiz verilerle ile kendi kendine öğrenilir.

# Pekiştirmeli öğrenme (Reinforcement Learning)

Daha çok denetimli öğrenmeye benzese de bu öğrenme tekniği deneme yanılma yoluyla yapılır.  Belirli bir sorun için en iyi öneriyi ya da ilkeyi geliştirmek üzere başarılı sonuçlar dizisi pekiştirilir. Pekiştirmeli öğrenmede ajan (agent) adı verilen öğrenen makinemiz karşılaştığı durumlara bir tepki verir ve bunun karşılığında da sayısal bir ödül sinyali alır. Makine bu ödül mekanizmasıyla ödül puanını maksimum yapmaya yönelik çalışır. Pekiştirmeli Öğrenme, Markov karar süreci denilen bir model kullanmaktadır. Markov karar süreçlerinin en önemli 3 özelliği; algılama (sensation), eylem (action) ve hedef (goal). Pekiştirmeli öğrenme yaklaşımı makine öğrenmesinin denetimli/danışmanlı öğrenme (supervised learning) ve denetimsiz/danışmansız öğrenme (unsupervised learning) yaklaşımlarından farklıdır. Pekiştirmeli öğrenmede ajan puanını maksimum yaparken daha önce keşfettiği ödül yolundan giderken daha yüksek puan almak için de başka yollar keşfetmelidir. Bahsedilen ajan bir robot veya



**Faydalandığımız kaynaklar**

<https://www.ibm.com/tr-tr/cloud/learn/machine-learning>

<https://github.com/caniban/MEU_Makine_Ogrenmesi>

<https://medium.com/machine-learning-t%C3%BCrkiye/ad%C4%B1m-ad%C4%B1m-makine-%C3%B6%C4%9Frenmesi-b%C3%B6l%C3%BCm-1-makine-%C3%B6%C4%9Frenmesi-nedir-407a8ba21428>

https://www.kaggle.com/code/nusretgencal/understanding-machine-learning-mostly-turkish