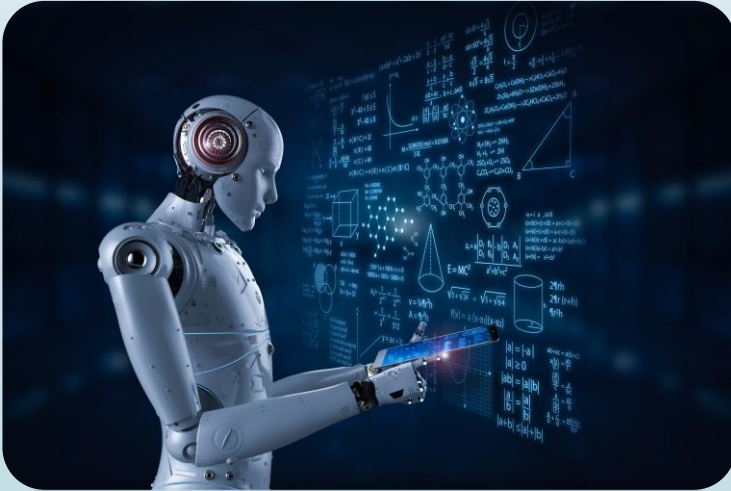


# Sınıflandırma Nedir?

1



Hazırlayan ve Sunan: **BEŞŞAR YASİN**

Tarih : 06/01/2023

Sürüm : v1

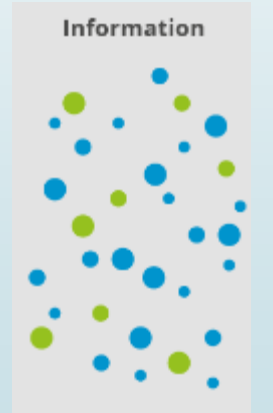
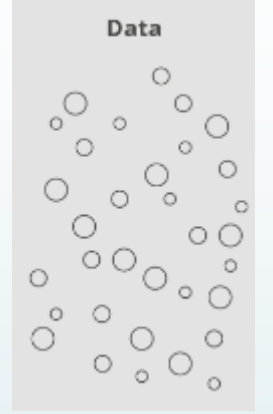
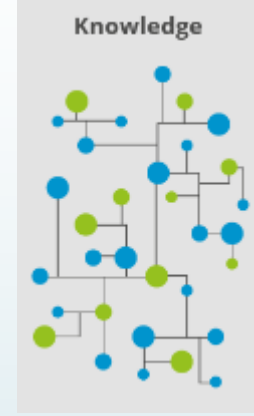
# İçindekiler

- Veri nedir?
- Veri Sınıflandırma nedir?
- Veri ve Bilgi Sınıflandırma İlkeleri
- Sınıflandırma nasıl yapılır?
- Sınıflandırma örneği -1
- Sınıflandırma Uygulaması -2
- Sonuç
- Kaynaklar



# Veri nedir?

- Veri, işlenmemiş ve yorumlanmamış bilgidir. Veri birçok farklı alandan elde edilebilir ve bilgisayar ortamında kullanımı önemlidir. Veriler nitel ve nicel veri toplama teknikleri ile elde edilir ve tek başlarına anlam ifade etmezler. İşlendikten sonra anlam kazanırlar ve belli bir bağlamı açıklayabilirler. Veriler gruplandırılır, sıralanır ve özetlenir, ve sonra elle veya bilgisayarla işlenir ve formüle edilir. Veri, araştırma, ölçüm, deney, ve gözlem gibi farklı yöntemlerle toplanır ve yapılandırılmış, yapılandırılmamış, ve yarı yapılandırılmış olabilir. Veri ayrıca ücretli ve ücretsiz, statik ve dinamik, açık ve büyük veri olarak da sınıflandırılabilir. Özetle, veri işlenmemiş bilgidir ve anlam kazanması için gruplandırılıp, sıralanıp, özetlenip ve işlenip formüle edilir.

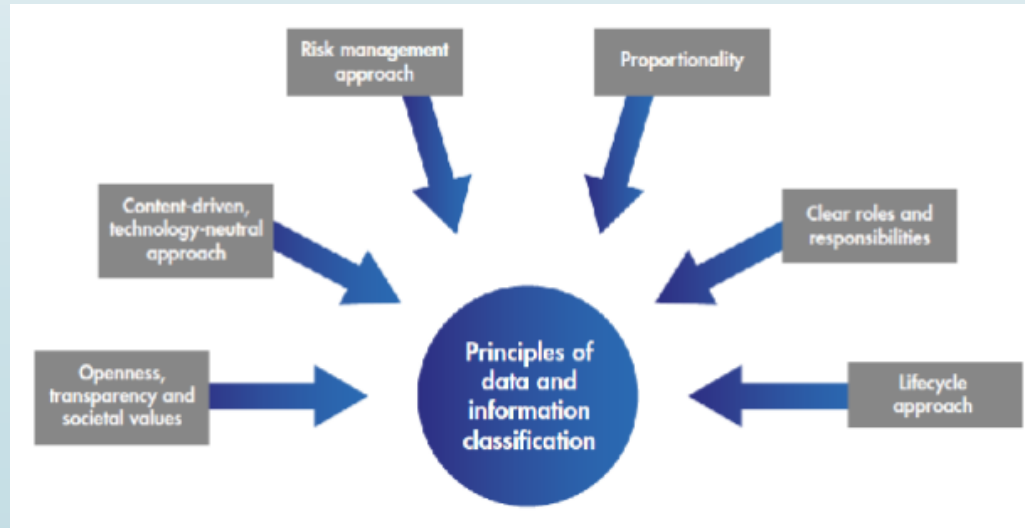


# Veri sınıflandırma nedir?

- Veri sınıflandırması, geniş anlamda, verilerin daha verimli kullanılabilmesi ve korunabilmesi için ilgili kategorilere göre düzenlenmesi süreci olarak tanımlanmaktadır. Temel düzeyde, sınıflandırma süreci verilerin bulunmasını ve alınmasını kolaylaştırır. Risk yönetimi, uyumluluk ve veri güvenliği söz konusu olduğunda veri sınıflandırması özellikle önemlidir. Veri sınıflandırması hassas verilerin uygun biçimde kullanılmasını sağlamada önem teşkil etmektedir. Veri sınıflandırmasının rolü büyük önem taşımaktadır. İlk olarak güvenliğini sağlamaya çalıştığımız veri ya da verileri belirlememiz gerekmektedir. Veri sınıflandırma önemli bir adımdır verinin niteliğini, hassasiyetine, değerini belirlemede önem taşımaktadır. Bilgiler, ortak bir riski paylaşan önceden tanımlanmış gruplara bölünür. Her grubu kendi içinde güvenceye almak için güvenlik kontrolleri tanımlanır. Hassas verilerin işlenmesi ve işlenmesini iyileştirmek için sınıflandırma araçları kullanılabilir. Dikkat çekici renklerde kullanılan uyarı etiketleri verinin gizli, tehlikeli vb. olduğunu göstermektedir. Tüm bunların kullanımı bizi daha dikkatli kullanmaya, davranmaya yönlendirmektedir. Veri sınıflandırması gizli, hassas verilerin korunmasını ve uyarı etiketleri ile karşılaşılabilecek olası sorun ve tehlikelere karşı önem niteliği taşımaktadır. Bunun yansıya gizli, hassas verilerin korunmasını, üçüncü şahısların eline geçmesine engel olma niteliği taşımaktadır. Veri sınıflandırması günlük rutin haline getirildiğinde güvenlik seviyesi artar ve ortaya çıkabilecek riskler azalır. Bunun yansıya metadata değeri yükselir ve güvenlik ihlalleri minimum seviyeye iner.

# Veri ve Bilgi Sınıflandırma İlkeleri

- Açıklık, şeffaflık ve toplumsal değerler göz önünde bulundurularak, veriler duyarlılık, değer ve kritikliğine göre sınıflandırılmalıdır. Erişim kısıtlamaları sadece bilginin ifşa edilmesinin kuruluşun, personelinin veya üçüncü şahısların meşru menfaatlerine ve yasal yükümlülüklerine zarar verebileceği durumlarda uygulanmalıdır. Bilgi içeriğine ve içeriğin tehlikeye atılma riskine göre sınıflandırılmalıdır ve bilginin formatı ve ortamına göre ayırım yapılmamalıdır. Bilgi koruma, bilginin duyarlılık, değer ve kritiklik düzeyine uygun olarak derecelendirilmiş bir yaklaşım ile yapılmalıdır. Bilgi uygun bir seviyede sınıflandırılmalıdır, ancak mümkün olduğu kadar düşük seviyede sınıflandırılır.

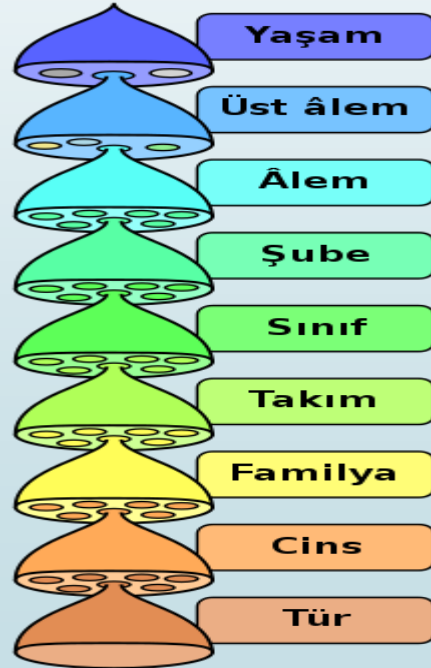


# Sınıflandırma nasıl yapılır?

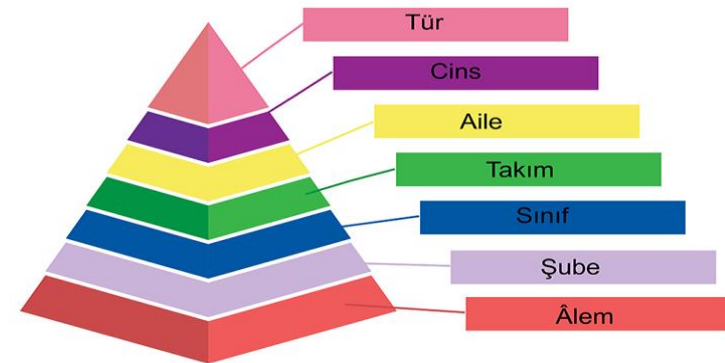
- Sınıflandırma, veri kümelerindeki öğeleri belirli sınıflara (kategorilere) göre ayırmak ve bu öğeleri sınıflarına göre grupelemek için kullanılan bir yöntemdir. Bu yöntem, veri madenciliği, makine öğrenimi ve benzeri alanlarda sıklıkla kullanılır.
- Sınıflandırma yapmak için öncelikle bir veri kümesi ve bu veri kümesinin sınıflandırılmasını istediğiniz özellikleri (öznitelikleri) belirlemek gerekir. Örneğin, bir kitap mağazasındaki kitap verilerini sınıflandırmak isteyebilirsiniz. Bu durumda, kitap verilerinin sınıflandırılmasını istediğiniz öznitelikler tür (öykü, roman, şiir vb.), yazar, yayınevi ve fiyat olabilir.
- Sonra, sınıflandırma yapmak için bir yöntem seçmelisiniz. Örneğin, destek vektör makinesi (SVM), k-en yakın komşu (k-NN) veya lojistik regresyon gibi popüler sınıflandırma algoritmalarından birini kullanabilirsiniz. Bu algoritmaları seçerken, veri kümenizin büyüklüğü, öznitelik sayısı ve veri kümenizdeki sınıf dağılımı gibi faktörleri de dikkate almalısınız.
- Son olarak, seçtiğimiz algoritmayı veri kümeniz üzerinde eğitim ve sınıflandırma işlemini gerçekleştirebiliriz. Eğitim sırasında, algoritma veri kümenizdeki öğeler arasındaki ilişkileri öğrenir ve bunları sınıflara göre grupelemek için kullanır. Eğitim tamamlandıktan sonra, algoritmayı test etmek için bir test veri kümesi kullanarak sınıflandırma performansını değerlendirebiliriz.

# Sınıflandırma Örneği -1

- Aşağıdaki örneklerde, açıklayıcı ve basit bir sınıflandırma örneği: Bu görüntü, tür, cinsiyet,
- takım, aile, derece ve yeri gösterdiği için canlı organizmaları bölümlere ayırmak
- istediğimizde sınıflandırma sürecini temsil eder.



## BİYOLOJİK SINIFLANDIRMA





# Sınıflandıma kodu Örneği -2

- İlk önce sklearn aktarmak
- Veri yüklemek x ve y
- Veriler test
- Model tespiti
- Model ifade
- Sınıflandırma nasıl oldu
- Bu kod, scikit-learn kütüphanesini kullanarak bir veri kümesinde Rastgele Orman Sınıflandırıcısını eğitmektedir. Veri kümesi iki parçaya bölünmüştür: eğitim kümesi ve test kümesi. Rastgele Orman Sınıflandırıcısı eğitim kümesinde eğitilir ve test kümesinde tahminler yapmak için kullanılır. Modelin doğruluk oranı sonra sklearn.metrics modülünden accuracy\_score fonksiyonu kullanılarak hesaplanır.

İlk satır sklearn.ensemble modülünden RandomForestClassifier sınıfını içerir.

- İkinci satır sklearn.model\_selection modülünden train\_test\_split fonksiyonunu içerir. Bu fonksiyon veriyi eğitim kümesi ve test kümesi olarak bölmek için kullanılır.
- Üçüncü satır sklearn.metrics modülünden accuracy\_score fonksiyonunu içerir. Bu fonksiyon modelin doğruluk oranını hesaplamak için kullanılır.

```
C:\Users\hp> OneDrive\ Masaüstü > yeni 2.py > ...  
1 # Import necessary libraries  
2 from sklearn.ensemble import RandomForestClassifier  
3 from sklearn.model_selection import train_test_split  
4 from sklearn.metrics import accuracy_score  
5  
6 # Load the data  
7 X = # features  
8 y = # labels  
9  
10 # Split the data into a training set and a test set  
11 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  
12  
13 # Define the model  
14 model = RandomForestClassifier(n_estimators=100)  
15  
16 # Train the model  
17 model.fit(X_train, y_train)  
18  
19 # Make predictions on the test set  
20 y_pred = model.predict(X_test)  
21  
22 # Calculate the accuracy  
23 accuracy = accuracy_score(y_test, y_pred)  
24  
25 print("Accuracy:", accuracy)
```



## Uygulama Örneği -2 (devam)

- Kod, model için verileri yükler, X özellik kümesi ve y etiketleridir.
- Sonra veri X\_train, X\_test, y\_train, y\_test'e eğitim ve test için oranı %80:%20 olarak train\_test\_split fonksiyonu ile bölünmüştür. random\_state 42 olarak ayarlanmıştır, bu her seferinde kod çalıştırıldığında aynı bölünme elde edileceği anlamına gelir.
- Rastgele Orman Sınıflandırıcısı tanımlanır, 100 estimator ile. Ormanın ağaçlarının sayısı, sonra model eğitim kümesi ve etiketleri ile fit metodu kullanılarak eğitilir.
- Model sonra test kümesinde tahminler yapmak için predict metodu kullanılır. Tahminler y\_pred değişkeninde saklanır.
- Son olarak, modelin doğruluk oranı accuracy\_score fonksiyonu kullanılarak hesaplanır, bu fonksiyon tahmin edilen etiketleri (y\_pred) gerçek etiketlere (y\_test) karşılaştırarak 0 ile

# Sonuç

10

Sınıflandırma, hayatta birçok alanda önemli bir rol oynamaktadır, bunlar arasında:

**Tıbbi tanı:** Tıbbi profesyoneller, hasta verilerine dayanarak hastalıkları tanımak ve sonuçları tahmin etmek için sınıflandırma algoritmalarını kullanırlar.

**Finans:** Bankalar ve diğer finans kurumları, sınıflandırma yöntemlerini kullanarak dolandırıcı işlemleri tanımlamak, kredi riskini değerlendirmek ve müşteri davranışlarını tahmin etmek için kullanırlar.

**Pazarlama:** Şirketler, müşterileri segmente etmek ve pazarlama kampanyalarını belirli gruplara hedeflemek için sınıflandırma kullanırlar.

**Tarım:** sınıflandırma algoritmaları, bitkilerdeki hastalıkları, zararlıları ve ürün verimini etkileyebilecek diğer sorunları tanımlamak için kullanılabilir.

**Doğal Dil İşleme:** sınıflandırma, NLP'de bir metnin duygu analizi yapmak, bir belgenin konusunu tanımlamak ve daha fazlası için kullanılır.

**Resim ve Video analizi:** sınıflandırma, resim ve video analizinde nesneleri tanımlamak ve izlemek, yüzleri tespit etmek ve daha fazlası için kullanılır.

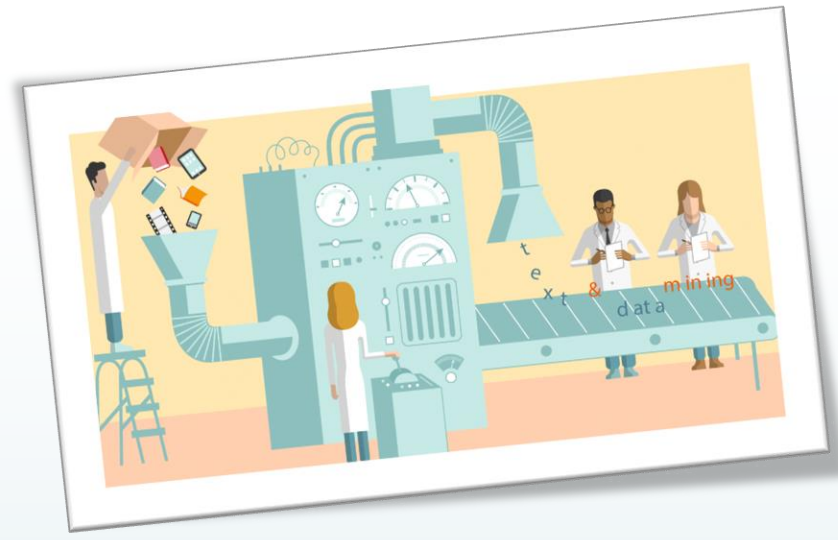
**Öz-sürücü arabalar:** sınıflandırma algoritmaları, arabanın çevresindeki nesneleri tanımlamak ve nasıl yol alacağına karar vermek için kullanılır.

Genel olarak, sınıflandırma, desenleri tanımlamak ve tahmin yapmak için önemli bir rol oynar, bu birçok farklı alanda karar vermeyi ve işlemleri otomatikleştirmeyi iyileştirmek için kullanılabilir.

# Kaynaklar

- <https://www.techtarget.com/searchdatamanagement/definition/data>
- <https://www.javatpoint.com/data>
- <https://www.datacamp.com/blog/classification-machine-learning>
- <https://machinelearningmastery.com/types-of-classification-in-machine-learning/>





İlginiz için teşekkürler...