

W.1.3 Veri madenciliğine Giriş ve Temel İstatistiksel Terimler

Veri madenciliğine Giriş

Veri Madenciliğinin Önemi

1. Büyük Veri Setlerinde Gizli Bilgileri Ortaya Çıkarır

Veri madenciliği, büyük ve karmaşık veri setlerinde insan gözüyle fark edilemeyen desenleri, eğilimleri ve korelasyonları tespit eder. Bu, karar verme süreçlerinde derinlemesine içgörü sağlar.

2. Tahmin ve Sınıflandırma Yeteneği Sunar

Makine öğrenmesi algoritmalarıyla desteklenen veri madenciliği, gelecekteki olayları tahmin etmeye ve verileri sınıflandırmaya yardımcı olur. Örneğin:

- Müşterinin ürün satın alıp almayacağını tahmin etme (pazarlama)
- Hastalığın teşhisinde riskli grupları belirleme (sağlık)

3. Karar Destek Sistemlerini Güçlendirir

Yönetim ve iş stratejilerinde kullanılabilecek öngörüler sunarak karar destek sistemlerine katkı sağlar. Böylece daha verimli ve stratejik kararlar alınabilir.

4. Veri Temelli Rekabet Avantajı Sağlar

Veri madenciliğiyle elde edilen bilgiler, şirketlere rakiplerine karşı önemli avantajlar sağlar. Örneğin, müşteri davranışlarını analiz ederek daha etkili kampanyalar düzenlenebilir.

5. Hataları ve Anomalileri Ortaya Çıkarır

Finans, güvenlik ve sağlık gibi alanlarda, sıra dışı veya şüpheli durumları tespit etmek için kullanılır. Bu, sahtekarlık tespiti ya da sistem hatalarının önceden görülmesini sağlar.

Temel İstatistiksel Terimler

Ortalama (Mean / Aritmetik Ortalama)

Tanım: Bir veri setindeki tüm değerlerin toplamının, veri setindeki toplam eleman sayısına bölünmesiyle elde edilen değerdir. En yaygın kullanılan merkezi eğilim ölçüsüdür.

Özellikleri:

- Tüm veri noktalarını hesaba katar.
- *Aykırı değerlerden (outliers) **çok etkilenir**. Örneğin, 1, 2, 3, 4, 100 veri setinin ortalaması 22'dir. 100 değeri ortalamayı oldukça yukarı çekmiştir.
- Genellikle **simetrik dağılımlar** (örneğin, normal dağılım) için en iyi merkezi eğilim ölçüsüdür.

Medyan (Median / Ortanca)

Tanım: Bir veri seti küçükten büyüğe veya büyükten küçüğe sıralandığında tam ortada yer alan değerdir. Veri setini iki eşit yarıya böler; değerlerin yarısı medyandan küçük, diğer yarısı medyandan büyüktür.

Nasıl Hesaplanır?

- Veri seti küçükten büyüğe (veya büyükten küçüğe) sıralanır.
- Eğer veri setindeki eleman sayısı (n) tek ise:** Medyan, tam ortadaki değerdir.
- Eğer veri setindeki eleman sayısı (n) çift ise:** Medyan, ortadaki iki değer aritmetik ortalamasıdır.

Özellikleri:

- Aykırı değerlerden ortalama kadar etkilenmez.** Bu nedenle, çarpık (skewed) veya aykırı değerler içeren dağılımlar için daha **robust (sağlam)** bir merkezi eğilim ölçüsüdür.

- Veri setinin ortasındaki "tipik" değeri daha iyi temsil edebilir.

Mod (Mode / Tepe Değer)

Tanım: Bir veri setinde **en sık tekrar eden değerdir**.

Nasıl Hesaplanır?

Veri setindeki her bir değerın kaç kez tekrarlandığı sayılır ve en yüksek frekansa sahip değer (veya değerler) mod olarak belirlenir.

Özellikleri:

- Hem **sayısal (kantitatif)** hem de **kategorik (nitel)** veriler için kullanılabilir. (Örneğin, en popüler renk, en çok tercih edilen ürün gibi).
- Aykırı değerlerden **etkilenmez**.
- Bir veri setinde birden fazla mod olabilir veya hiç mod olmayabilir.



NOT: Mean, Mode ve Median değerleri arasında most robust olan median değerleridir.



Veri analitiğinde "robust" (Türkçe karşılığı **sağlam** veya **dayanıklı**) terimi, bir istatistiksel yöntemin, modelin veya analizin, veri setindeki **aykırı değerlere (outliers)** veya **varsayımların ihlallerine karşı ne kadar dirençli olduğunu** ifade eder. Yani, verilerdeki küçük değişiklikler veya beklenmedik durumlar karşısında sonuçların ne kadar tutarlı ve güvenilir kaldığını belirtir.

Medyan ve Ortalama Farkının Anlamı

- **Ortalama > Medyan:** Bu durum genellikle **sağa çarpık (pozitif çarpık)** bir dağılıma işaret eder. Yani, veri setinde ortalamayı yukarı çeken, nispeten az sayıda yüksek değer (aykırı değerler) bulunur.
- **Ortalama < Medyan:** Bu durum genellikle **sola çarpık (negatif çarpık)** bir dağılıma işaret eder. Yani, veri setinde ortalamayı aşağı çeken, nispeten az sayıda düşük değer (aykırı değerler) bulunur.

Medyan ve Ortalama Farkı Fazla İse Ne Yapılır?

Medyan ve ortalama arasındaki büyük farkı tespit ettiğinizde, atmanız gereken adımlar şunlardır:

1. Veri Dağılımını Görselleştirin:

- **Histogram:** Verilerinizin şeklini, çarpıklığını ve olası aykırı değerleri görmek için en etkili yollardan biridir.
- **Kutu Grafiği (Box Plot):** Özellikle aykırı değerleri (noktalar veya yıldızlar olarak gösterilir) ve veri setinin çeyrekliklerini (Q1, Medyan, Q3) net bir şekilde gösterir.
- **Yoğunluk Grafiği (Density Plot):** Veri yoğunluğunun hangi aralıklarda toplandığını görselleştirir.

Bu görselleştirmeler, farkın neden kaynaklandığını anlamanıza yardımcı olacaktır.

2. Aykırı Değerleri Tespit Edin ve İnceleyin:

- **Tanım:** Aykırı değerler, veri setinin genel eğiliminden önemli ölçüde sapan gözlemlerdir.
- **İnceleme:** Bu aykırı değerlerin hatalı veri girişinden mi (örn. yanlış yazılmış bir sayı) yoksa gerçekten nadir ama geçerli olaylardan mı kaynaklandığını belirleyin.
- **Müdahale:**
 - **Hatalı İse:** Hatalı olan aykırı değerleri düzeltin veya analizden çıkarın.
 - **Geçerli İse:** Bu değerleri veri setinde tutmak önemlidir, ancak analiz yöntemlerinizi buna göre ayarlamanız gerekir.

3. Hangi Merkezi Eğilim Ölçüsünün Daha Uygun Olduğuna Karar Verin:

- **Medyan Kullanımı:** Veri setinizde aykırı değerler veya belirgin bir çarpıklık varsa, **medyan** genellikle verinin "tipik" değerini ortalamadan daha iyi temsil eder. Örneğin, bir ülkedeki ortalama geliri hesaplarken, çok zengin birkaç kişinin ortalamayı şişirmesini önlemek için medyan gelir daha anlamlıdır.

- **Ortalama Kullanımı:** Eğer veri dağılımı simetrikse ve aykırı değerler yoksa veya önemsizse, ortalama hala uygun bir ölçümdür.

4. Robust İstatistiksel Yöntemleri Kullanın:

- **Tanım:** Robust yöntemler, aykırı değerlere ve varsayım ihlallerine karşı daha dirençli olan istatistiksel tekniklerdir.
- **Örnekler:**
 - **Kesilmiş Ortalama (Trimmed Mean):** Verinin en küçük ve en büyük belirli bir yüzdesini (örneğin %5 veya %10) dışarıda bırakarak ortalama hesaplar. Bu, aykırı değerlerin etkisini azaltır.
 - **Medyan Mutlak Sapma (Median Absolute Deviation - MAD):** Standart sapmaya robust bir alternatiftir ve veri setindeki değişkenliği ölçerken aykırı değerlerden daha az etkilenir.
 - **Robust Regresyon Yöntemleri:** Klasik En Küçük Kareler (OLS) regresyonu yerine, aykırı değerlerin regresyon çizgisini bozmasını engelleyen yöntemler (örn. Huber regresyonu, Tukey biweight regresyonu) kullanılabilir.
 - **Parametrik Olmayan Testler:** Verilerin belirli bir dağılıma (örn. normal dağılım) sahip olduğu varsayımını yapmayan testler (örn. Mann-Whitney U testi, Wilcoxon sıralı işaret testi) de aykırı değerlere karşı daha robust olabilir.

5. Veriyi Dönüştürmeyi Düşünün:

- Bazı durumlarda, verileri matematiksel olarak dönüştürmek (örn. logaritmik dönüşüm, karekök dönüşümü) dağılımı daha simetrik hale getirebilir ve bu da ortalama gibi parametrik yöntemlerin daha uygun hale gelmesini sağlar. Ancak dönüşüm, yorumlamayı zorlaştırabilir, bu yüzden dikkatli kullanılmalıdır.

Standart sapma (Standard Deviation - SD)

Tanım: bir veri setindeki değerlerin aritmetik ortalamadan ne kadar uzaklaştığını, yani **yayılımını veya değişkenliğini** ölçen en yaygın kullanılan istatistiksel ölçülerden biridir.

- **Düşük standart sapma:** Veri noktalarının ortalamaya yakın bir şekilde toplandığını, yani verilerin **homojen** ve tutarlı olduğunu gösterir.
- **Yüksek standart sapma:** Veri noktalarının ortalamadan geniş bir aralığa yayıldığını, yani verilerin **heterojen** ve değişken olduğunu gösterir.

Standart sapma, verilerin orijinal birimiyle ifade edildiği için yorumlaması kolaydır.

Standart Sapmanın Özellikleri

- **Pozitif veya Sıfır:** Her zaman ≥ 0 . Sıfırsa, tüm değerler eşittir; yayılım yoktur.
- **Aykırı Değerlere Duyarlı:** Uç değerlerden (aykırı değerler) kolayca etkilenir.
- **Aynı Birim:** Verilerin orijinal ölçü birimiyle ifade edilir, yorumlaması kolaydır.
- **Ortalamaya Bağlı:** Hesaplaması ortalamaya dayanır ve ortalamanın değişimiyle değişir.
- **Cebirsel İşlemlere Uygun:** İleri istatistiksel analizlerde yaygın olarak kullanılır.
- **Orijin Değişiminden Etkilenmez:** Her değere sabit ekleme/çıkarma SD'yi değiştirmez.
- **Ölçek Değişiminden Etkilenir:** Her değeri sabit bir sayıyla çarpma/bölme SD'yi aynı oranda değiştirir.
- **Dağılım Bilgisi:** Verilerin ortalama etrafındaki yayılımını ve dağılımın genişliğini gösterir. Özellikle normal dağılımda çok anlamlıdır.

Varyans

Tanım: bir veri setindeki değerlerin **aritmetik ortalamadan ne kadar yayıldığını, ne kadar değişken olduğunu** gösteren temel bir istatistiksel ölçüdür. Daha spesifik olarak, her bir veri noktasının ortalamadan olan farkının karelerinin ortalamasıdır.

Kısaca varyans, verilerinizin ne kadar **dağınık** olduğunun bir göstergesidir.

- **Yüksek varyans:** Veri noktaları ortalamadan uzakta, geniş bir aralığa yayılmış demektir. Bu, veri setinin **daha değişken** ve heterojen olduğunu gösterir.
- **Düşük varyans:** Veri noktaları ortalamaya yakın bir şekilde toplanmış demektir. Bu da veri setinin **daha az değişken** ve homojen olduğunu gösterir.

Varyansın Özellikleri

- **Her Zaman Pozitif veya Sıfır:** Negatif olamaz. Sıfırsa, tüm veri noktaları aynıdır; yayılım yoktur.
- **Aykırı Değerlere Duyarlı:** Uç değerlerden (aykırı değerler) kolayca etkilenir.
- **Ölçü Birimi Farklı:** Orijinal veri biriminin **karesidir** (örn. metre için metrekare). Bu durum, yorumlamayı zorlaştırabilir.
- **Standart Sapmanın Karesi:** Standart sapmanın karesidir; dolayısıyla matematiksel hesaplamalarda ve istatistiksel teoride temel bir rol oynar.
- **Cebirsel İşlemlere Uygun:** Matematiksel işlemlere ve daha karmaşık istatistiksel modellere daha uygundur.
- **Orijin Değişiminden Etkilenmez, Ölçek Değişiminden Etkilenir:** Verilere sabit bir sayı eklemek/çıkarmak varyansı değiştirmezken, sabit bir sayıyla çarpmak/bölmek varyansı o sayının karesiyle çarpar/böler.

Varyasyon Katsayısı (Coefficient of Variation - CV)

Tanım: bir veri setindeki **bağıl değişkenliği** veya **yayılımı** ölçen standartlaştırılmış bir istatistiksel ölçüdür. Standart sapmanın ortalamaya oranlanmasıyla hesaplanır ve genellikle yüzde (%) olarak ifade edilir.

Varyasyon katsayısı, farklı birimlerde veya çok farklı ortalamalara sahip veri setlerinin değişkenliğini **karşılaştırmak için** son derece kullanışlıdır.

Varyasyon Katsayısı Neden Önemlidir?

- **Karşılaştırma Yeteneği:** En önemli özelliğidir. Standart sapma, verilerin orijinal birimiyle ifade edildiği için, farklı birimlerdeki veya farklı büyüklükteki veri setlerini doğrudan karşılaştırmak zordur. Varyasyon Katsayısı, birimsiz bir değer (veya yüzde) olduğu için bu karşılaştırmayı mümkün kılar.
- **Bağıl Değişkenlik:** Verilerin ortalamalarına göre ne kadar yayıldığını gösterir. Yüksek bir CV(>%35), ortalamaya göre daha fazla değişkenlik olduğunu, düşük bir CV(<%35) ise ortalamaya göre daha az değişkenlik (veya daha fazla tutarlılık) olduğunu gösterir.
- **Risk Analizi:** Finans gibi alanlarda, farklı yatırımların risk-getiri oranlarını karşılaştırmak için kullanılır. Daha düşük bir CV, genellikle belirli bir getiri düzeyi için daha az risk anlamına gelir.

Histogram Dağılımı

Tanım: Veri setinizdeki sayısal verilerin dağılımını görselleştirmek için kullanılan güçlü bir grafik aracıdır. Sürekli verileri belirli aralıklara (sınıflara veya kutulara) ayırır ve her aralığa düşen veri noktalarının sayısını (frekansını) dikey çubuklarla gösterir.

Histogram Nedir ve Ne İşe Yarar?

Bir histogram temel olarak şunları gösterir:

- **X Eksen (Yatay):** Veri değerlerinin aralıkları (sınıfları). Örneğin, yaş aralıkları, fiyat aralıkları gibi. Bu çubuklar birbirine bitişiktir, çünkü sürekli bir değişkenin aralıklarını temsil ederler.
- **Y Eksen (Dikey):** Her bir aralığa düşen veri noktalarının **frekansı (sayısı)** veya yüzdesi. Çubukların yüksekliği, o aralıktaki veri yoğunluğunu gösterir.

Ne İşe Yarar?

- **Veri Dağılımının Şeklini Anlamak:** Verilerinizin simetrik mi, çarpık mı, tek tepeli mi yoksa çok tepeli mi olduğunu hızlıca görmeyi sağlar.
- **Merkezi Eğilimi ve Yayılımı Görmek:** Verilerin nerede yoğunlaştığını (merkezi eğilim) ve ne kadar geniş bir alana yayıldığını (yayılım) görsel olarak değerlendirmeyi sağlar.
- **Aykırı Değerleri Tespit Etmek:** Ana veri grubundan uzaklaşan, izole çubuklar aykırı değerlerin varlığına işaret edebilir.
- **Süreç Kabiliyeti:** Kalite kontrol gibi alanlarda, bir sürecin belirli spesifikasyon limitleri içinde ürün üretip üretmediğini analiz etmek için kullanılır.
- **Hipotez Oluşturma:** Verilerinizin hangi istatistiksel dağılıma uyabileceği hakkında ilk fikirleri edinmeyi sağlar (örneğin, normal dağılım).

Farklı Histogram Dağılım Şekilleri

Histogramın şekli, veri setiniz hakkında çok değerli bilgiler sağlar. İşte en yaygın dağılım şekilleri:

1. Simetrik (Çan Şekilli / Normal Dağılım)

- **Görünüm:** Grafiğin ortasında tek bir tepe noktası bulunur ve bu tepe noktasından her iki yana doğru simetrik olarak alçalır, bir çan şeklini andırır.
- **Özellikler:** **Ortalama, medyan ve mod** birbirine çok yakındır veya eşittir. Birçok doğal fenomeni (boy uzunlukları, IQ skorları) bu dağılıma uyar.
- **Anlamı:** Verilerin ortalama etrafında dengeli bir şekilde toplandığını gösterir.

2. Çarpık Dağılımlar (Skewed Distributions)

Verilerin simetrik olmadığı durumlardır ve kuyruklarının hangi yöne uzadığına göre adlandırılırlar:

- **Sağa Çarpık (Pozitif Çarpık):**
 - **Görünüm:** Tepe noktası grafiğin sol tarafında yer alır ve dağılımın uzun bir "sağ kuyruğu" vardır.
 - **Özellikler:** **Ortalama > Medyan > Mod** sıralaması görülür. Yüksek değerler, ortalamayı medyandan daha fazla yukarı çeker.
 - **Anlamı:** Veri setinde nispeten az sayıda yüksek değere sahip gözlemler (aykırı değerler olabilir) bulunur. Örnek: Gelir dağılımları (çoğu kişinin geliri düşük-orta seviyede, çok az kişinin geliri çok yüksek).
- **Sola Çarpık (Negatif Çarpık):**
 - **Görünüm:** Tepe noktası grafiğin sağ tarafında yer alır ve dağılımın uzun bir "sol kuyruğu" vardır.
 - **Özellikler:** **Mod > Medyan > Ortalama** sıralaması görülür. Düşük değerler, ortalamayı medyandan daha fazla aşağı çeker.
 - **Anlamı:** Veri setinde nispeten az sayıda düşük değere sahip gözlemler bulunur. Örnek: Sınav notları (çoğu öğrencinin notu yüksek, az sayıda düşük not).

3. Bimodal (Çift Tepeli)

- **Görünüm:** Histogramda belirgin iki tepe noktası bulunur.
- **Anlamı:** Veri setinin aslında iki farklı alt gruptan oluştuğunu ve her grubun kendi içinde bir tepe noktasına sahip olduğunu gösterebilir. Bu durum, veri setini daha derinlemesine inceleme ve alt grupları ayırma ihtiyacına işaret eder. Örnek: Bir sınıftaki öğrencilerin boyları eğer sınıfta hem ilkokul öğrencileri hem de lise öğrencileri varsa.

4. Düzgün (Uniform)

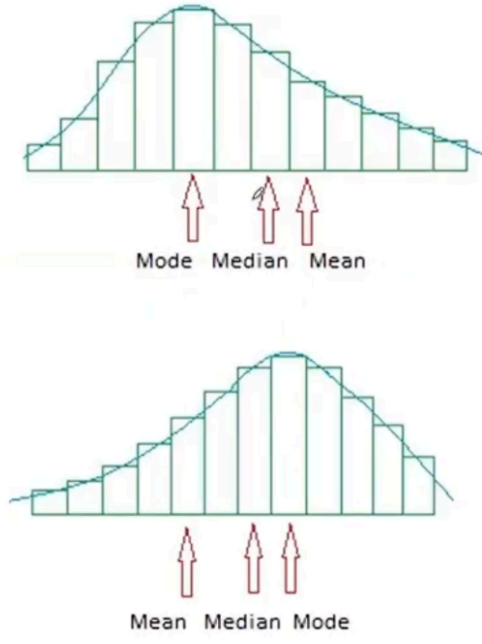
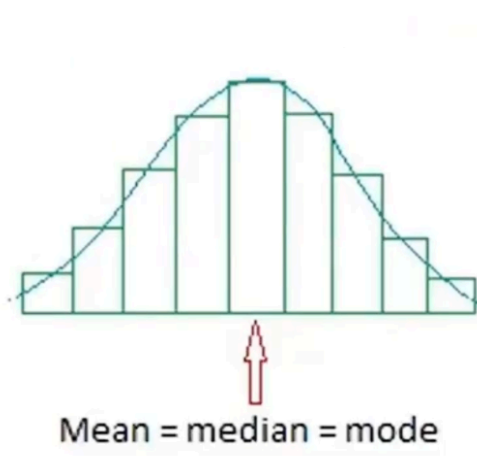
- **Görünüm:** Çubukların yükseklikleri yaklaşık olarak aynıdır, bu da grafik boyunca nispeten düz bir çizgi oluşturur.
- **Anlamı:** Verilerin belirli bir aralıkta eşit olasılıkla dağıldığını gösterir. Her bir aralıkta yaklaşık olarak aynı sayıda gözlem vardır. Örnek: Adil bir zarın atılışları.

5. Düzensiz / Çok Tepeli

- **Görünüm:** Belirgin bir desen veya ana tepe noktası yoktur, birden fazla küçük tepe noktası olabilir.
- **Anlamı:** Verilerin rastgele bir dağılıma sahip olduğunu veya ölçümde sorunlar olabileceğini gösterebilir. Genellikle veri setinde birden fazla nedeni olan karmaşık bir değişkenliği işaret eder.



Histogramlar, özellikle büyük veri setleriyle çalışırken, verilerinizi anlamanın ve sonraki analiz adımlarını belirlemenin ilk ve en önemli yollarından biridir. Hangi merkezi eğilim ölçüsünün (ortalama mı, medyan mı) daha uygun olduğuna karar vermede de kritik rol oynar.



En tepe nokta MOD; MEDIAN her zaman ortada; MEAN kuyruğa doğru.

Korelasyon

Tanım: iki veya daha fazla değişken arasındaki doğrusal ilişkinin yönünü ve gücünü anlamak için kullanılan temel ve çok önemli bir istatistiksel yöntemdir.

Korelasyonun Rolü ve Önemi

1. İlişki Keşfi:

- **Keşifsel Veri Analizi (EDA):** Korelasyon, bir veri setine ilk bakışta hangi değişkenlerin birbiriyle ilgili olabileceğini gösterir. Bu, ileri analizler (örneğin regresyon modeli oluşturma) için hangi değişkenlerin potansiyel olarak önemli olduğunu belirlemeye yardımcı olur.
- **Örnek:** Müşteri demografik verileri ve satın alma alışkanlıkları arasındaki korelasyonu inceleyerek, belirli yaş gruplarının belirli ürünleri daha fazla satın alma eğiliminde olup olmadığını anlayabiliriz.

2. Yön ve Güç Belirleme:

- **Yön:** Değişkenler aynı yönde mi (biri artarken diğeri de artıyor) yoksa zıt yönde mi (biri artarken diğeri azalıyor) hareket ediyor?
- **Güç:** Bu ilişki ne kadar güçlü? Yani bir değişkenin değişimi diğeri ne kadar tutarlı bir şekilde etkiliyor? Korelasyon katsayısı (-1 ile +1 arası) bu gücü ve yönü sayısal olarak ifade eder.
 - **Korelasyon Katsayısı Yorumlamaları (genel kabul görmüş, ancak bağlama göre değişebilir):**
 - **0.00 – 0.19:** Çok zayıf ilişki
 - **0.20 – 0.39:** Zayıf ilişki
 - **0.40 – 0.59:** Orta düzeyde ilişki
 - **0.60 – 0.79:** Güçlü ilişki
 - **0.80 – 1.00:** Çok güçlü ilişki

3. Örüntü ve Eğilim Tespiti:

- Korelasyon matrisleri veya saçılım grafikleri (scatterplot) aracılığıyla, veri setindeki genel örüntüleri ve eğilimleri görselleştirmek ve analiz etmek mümkündür.

4. Tahmin ve Modelleme İçin Temel Oluşturma:

- Güçlü korelasyonlar, bir değişkeni kullanarak diğeri tahmin etmeye yönelik regresyon analizi gibi daha ileri istatistiksel modelleme çalışmaları için bir ön koşul olabilir. Yüksek korelasyon, bir değişkenin diğeri açıklama potansiyelinin yüksek olduğunu gösterir.

5. Risk ve Portföy Yönetimi (Finans):

- Finansal piyasalarda, farklı varlıkların (hisse senetleri, tahviller vb.) fiyat hareketlerinin birbiriyle nasıl ilişkili olduğunu anlamak için korelasyon kullanılır. Negatif korelasyonlu varlıkları portföye dahil etmek, riski çeşitlendirmeye ve azaltmaya yardımcı olabilir.

6. Değişken Seçimi (Feature Selection):

- Makine öğrenimi ve veri madenciliği projelerinde, bağımlı değişkenle yüksek korelasyona sahip bağımsız değişkenleri (özellikleri) seçmek, model performansını artırabilir. Ayrıca, bağımsız değişkenler arasındaki yüksek korelasyon (çoklu doğrusallık), modelde sorunlara yol açabileceği için bu durumun tespiti de önemlidir.

Korelasyon Uygulama Adımları (Genel)

Veri Hazırlığı: Analiz edilecek değişkenlerin sayısal (kantitatif) olması gerekir. Kategorik değişkenler için farklı korelasyon ölçümleri veya dönüşümler gerekebilir. Eksik değerlerin (NaN) yönetimi önemlidir.

Varsayımların Kontrolü (Pearson Korelasyonu İçin):

- **Doğrusallık:** Değişkenler arasında doğrusal bir ilişki beklentisi. Saçılım grafiği ile görsel kontrol edilir.
- **Normal Dağılım:** Pearson korelasyonu, değişkenlerin normal dağılıma yakın olmasını tercih eder. Ancak Spearman gibi parametrik olmayan korelasyonlar bu varsayımı gerektirmez.
- **Aykırı Değerler:** Aykırı değerler Pearson korelasyon katsayısını yanıltıcı şekilde etkileyebilir. Bu nedenle aykırı değerlerin tespiti ve yönetimi önemlidir.

Korelasyon Katsayısının Seçimi:

- **Pearson Korelasyonu (r):** En yaygın olanıdır. Doğrusal ilişkiyi ölçer, değişkenlerin sürekli ve yaklaşık normal dağılımlı olduğu varsayımını yapar.
- **Spearman Sıra Korelasyonu (rho):** Veriler sıralı (ordinal) ise veya normal dağılım varsayımı ihlal ediliyorsa kullanılır. Değişkenlerin sıraları arasındaki ilişkiyi ölçer.
- **Kendall's Tau:** Spearman'a benzer, özellikle küçük örneklem ve sıralı veriler için uygundur.

Korelasyon Hesaplaması: İstatistiksel yazılımlar (Python'da Pandas, R, SPSS, Excel vb.) kullanılarak kolayca hesaplanır. Genellikle bir korelasyon matrisi şeklinde sunulur.

Görselleştirme:

- **Saçılım Grafiği (Scatter Plot):** İki değişken arasındaki ilişkinin görsel olarak incelenmesi için olmazsa olmazdır. Korelasyonun yönünü, gücünü ve doğrusal olup olmadığını gösterir.
- **Isı Haritası (Heatmap):** Birden fazla değişken arasındaki korelasyonları bir matris şeklinde görselleştirmek için kullanılır. Renk yoğunluğu, ilişkinin gücünü gösterir.

Yorumlama: Elde edilen korelasyon katsayıları ve görselleştirmeler ışığında değişkenler arasındaki ilişkinin yönü, gücü ve istatistiksel anlamlılığı yorumlanır.

Nedensellik Uyarısı: Her zaman hatırlanmalıdır: Korelasyon, nedensellik anlamına gelmez. Daha fazla analiz (örneğin regresyon, deneysel çalışmalar) nedensel ilişkileri doğrulamak için gereklidir.