



# Regresyon Modeli: Veri Madenciliğinin Güçlü Aracı

Regresyon modelleri, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin matematiksel bir gösterimini sağlayarak, veri kümeleri içindeki kalıpları ve trendleri belirlemeye yardımcı olur. Bu sunum, regresyon modellerinin temel kavramlarını, türlerini, oluşturma adımlarını ve değerlendirme ölçütlerini kapsamlı bir şekilde inceleyecektir.

Dr. Öğr. Üyesi Alper Talha KARADENİZ

# Regresyon Nedir? Temel Kavramlar ve Amaçlar

## Temel Kavramlar

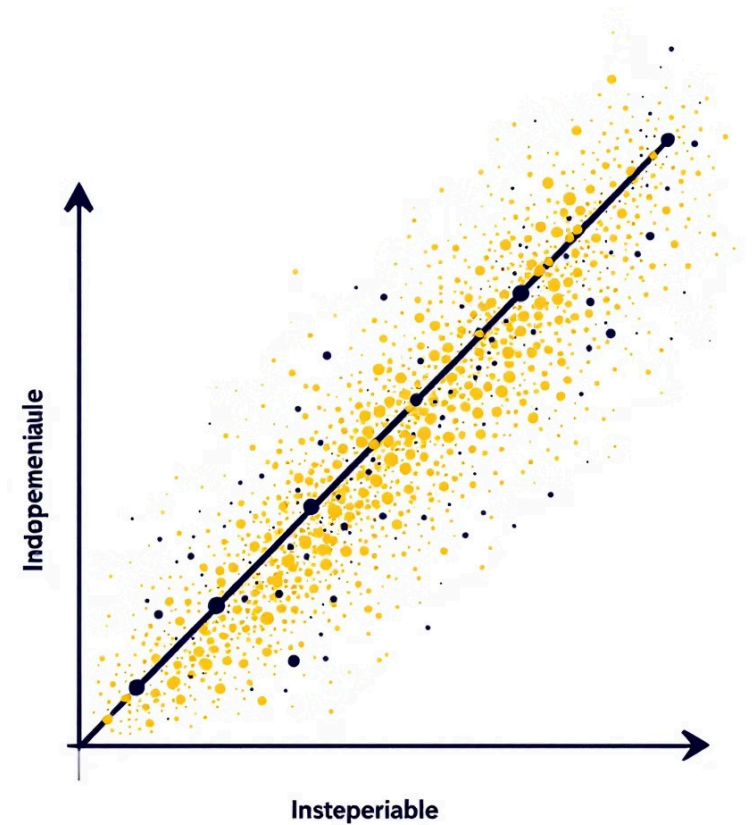
Regresyon, bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiyi analiz etmek için kullanılan istatistiksel bir tekniktir. Bağımsız değişkenler, bağımlı değişkeni etkileyen faktörlerdir. Örneğin, ev fiyatını tahmin ederken büyüklük, konum ve oda sayısı bağımsız değişkenler olabilir.

## Regresyon Tanımı

Regresyon analizi değişkenler arasındaki ilişkiyi inceleyen bir istatistiksel araçtır. Genellikle araştırmacı bir değişkenin bir diğeri üzerindeki nedensel etkisini ortaya çıkarmaya çalışır. Örnek olarak talebe göre fiyat artışı verilebilir.

## Amaçlar

Regresyon modelleri, veri kümeleri içindeki kalıpları ve trendleri belirlemeye yardımcı olur. Bu modeller, bağımlı değişkenin gelecekteki değerlerini tahmin etmek ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini ölçmek için kullanılır.



# Regresyon Modeli Türleri

## Doğrusal Regresyon

Bağımlı değişken ile bağımsız değişkenler arasında doğrusal bir ilişki varsayan en basit regresyon modelidir. Bu model, veri noktalarını en iyi şekilde temsil eden doğrusal bir çizgi bulmayı amaçlar.

## Lojistik Regresyon

Bağımlı değişkenin kategorik (evet/hayır) bir değişken olduğu durumlarda kullanılır. Bu model, olasılığı tahmin etmek için bir sigmoid fonksiyonunu kullanır.

## Polinomial Regresyon

Bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiyi bir polinom fonksiyonu ile temsil eden bir modeldir. Bu model, veri noktalarını daha karmaşık eğrilerle temsil edebilir.

Özellik	Doğrusal Regresyon	Lojistik Regresyon
Amaç	Sayısal (sürekli) bir değeri tahmin eder.	Kategorik (sınıflandırma) bir sonucu tahmin eder.
Hedef Değişken (Y)	Gerçek sayılar (örneğin: 50.0, 120.3)	Kategorik değerler (örneğin: 0 ya da 1)
Çıktı Aralığı	$(-\infty, \infty)$ — yani her türlü sayı olabilir	$(0, 1)$ aralığında bir olasılık değeri
Model Denklemi	$y = \beta_0 + \beta_1 x + \dots + \beta_n x_n$	$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x + \dots + \beta_n x_n)}}$
Kayıp Fonksiyonu	Ortalama Kare Hata (MSE)	Lojistik Kayıp (Log Loss / Binary Cross Entropy)
Sonuç Yorumu	Gerçek değer tahmini (örneğin: ev fiyatı)	Olasılık tahmini ve sınıf kararı (örneğin: hastalık var/yok)
Kullanım Alanları	Tahmin problemleri: gelir, fiyat, sıcaklık	Sınıflandırma: spam/ham, hasta/sağlıklı, evet/hayır

# Regresyon Modeli Oluşturma Adımları

## Veri Hazırlama

Regresyon modeli oluşturmanın ilk adımı veri hazırlamaktır. Bu adımda veri temizlenir ve eksik değerler tamamlanır. Veri kalitesi, modelin performansını doğrudan etkiler.

## Model Seçimi

Veri hazırlandıktan sonra, uygun bir regresyon modeli seçilmelidir. Doğrusal, lojistik veya polinomial regresyon gibi çeşitli regresyon modelleri arasından veri yapısına en uygun olan seçilir.

## Model Eğitimi

Seçilen model, veri kümesi kullanılarak eğitilmelidir. Eğitim sürecinde, modelin parametreleri veri kümesine uydurulur ve en uygun değerler belirlenir.

## Model Değerlendirmesi

Eğitim tamamlandıktan sonra, modelin performansı değerlendirilmelidir. Bu, modelin tahmin doğruluğunu ve genel performansını ölçmek için gerçekleştirilir.

## Tahmin

Model değerlendirildikten sonra, yeni veri noktaları için tahminlerde bulunmak üzere kullanılabilir. Bu, modelin eğitildiği aynı veri kümesinde veya yeni bir veri kümesinde gerçekleştirilebilir.

# Regresyon Modeli Değerlendirme ve Performans Ölçütleri



## Ortalama Kare Hata (MSE)

MSE, modelin tahminlerinin gerçek değerlerden ne kadar uzak olduğunu ölçer.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

MSE ne kadar küçük ise, modelin tahminleri gerçek değerlere o kadar yakındır. Kare alma işlemi, büyük hataları daha fazla cezalandırır.



## Ortalama Mutlak Hata (MAE)

MAE, modelin tahminlerinin gerçek değerlerden ortalama olarak ne kadar uzak olduğunu ölçer.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

MAE ne kadar küçük ise, model tahminleri gerçek değerlere ortalamada o kadar yakın demektir.



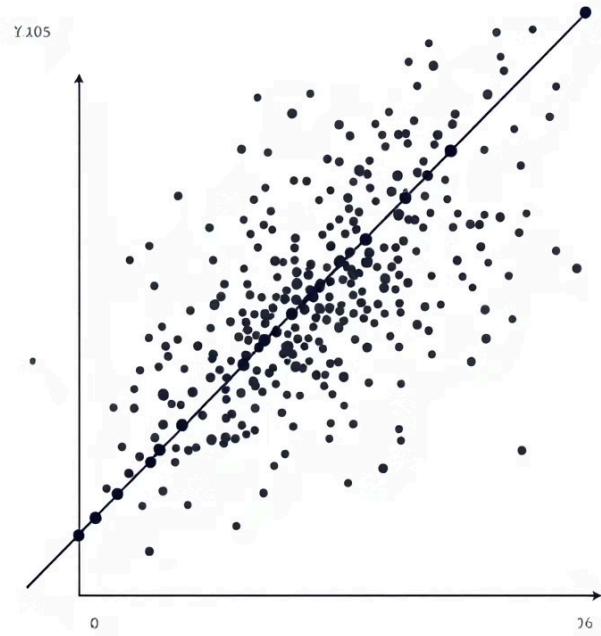
## R-kare Değeri

R-kare değeri, modelin verileri ne kadar iyi açıkladığını gösterir.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- $R^2=1$  Model tüm veriyi mükemmel açıklar.
- $R^2=0$  Model hiçbir şeyi açıklayamıyor demektir (ortalama kadar iyi).
- $R^2<0$  Model, ortalama değer tahmininden bile kötü.

# Basit Doğrusal Regresyon



$f(x)$

## Temel Denklem

Stokastik (Olasılıklı) bir model olan ve ana kütledeki ilişkiyi gösteren basit doğrusal regresyon denklemi şöyle ifade edilir:  $y = B_0 + B_1X + \varepsilon$

## Parametreler

$B_0$ : Doğrunun y-eksenini kestiği yer ve regresyon sabitidir.  $B_1$ : Doğrunun eğimi veya regresyon katsayısıdır.  $\varepsilon$ : Rastgele (Tesadüfi-Şans) hata değeridir.

## Uygulama

Basit doğrusal regresyon; bağımsız değişken (X) ile bağımlı değişken (Y) deki değişimi açıklamayı, bağımsız değişkendeki bir birimlik değişimin bağımlı değişken üzerindeki etkisini ölçmeyi amaçlar.



# Regresyon Modeli Varsayımları ve Tahmin

## Doğrusallık

Doğrusal regresyon modelleri, bağımsız değişkenler ile bağımlı değişkenler arasında doğrusal bir ilişki olduğunu varsayar. Bu varsayım, modelin tahmin doğruluğunu etkileyebilir.

## Normal Dağılım

Regresyon modelleri, hata terimlerinin normal bir dağılıma sahip olduğunu varsayar. Bu varsayım, modelin tahmin doğruluğunu etkileyebilir ve istatistiksel testlerin geçerliliği için önemlidir.



## Bağımsızlık

Regresyon modelleri, hata terimlerinin birbirinden bağımsız olduğunu varsayar. Bu varsayım, modelin tahmin doğruluğunu etkileyebilir ve zaman serisi verilerinde özellikle önemlidir.

## Sabit Varyans

Regresyon modelleri, hata terimlerinin tüm bağımsız değişken değerleri için aynı varyansa sahip olduğunu varsayar. Bu varsayım, modelin tahmin doğruluğunu etkileyebilir.

# Regresyon Modelinin Avantajları ve Sınırları

## Avantajları

- Veri içindeki trendleri ve kalıpları belirleme
- Gelecekteki değerleri tahmin etme
- Bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini ölçme
- Kolay uygulanabilirlik ve yorumlanabilirlik

## Sınırları

- Doğrusal olmayan ilişkilere uyum sağlamakta zorluk
- Çoklu değişkenli modellerde değişken seçimi zorluğu
- Veri kalitesine duyarlılık
- Varsayımların ihlali modelin doğruluğunu etkileyebilir



# Vaka Çalışması: E-Ticaret Satış Analizi

1.50

## Regresyon Katsayısı

Örnek hesaplamada bulunan B1 değeri, bağımsız değişkendeki bir birimlik artışın bağımlı değişkende 1.50 birimlik artışa neden olduğunu gösterir. Bu, örneğin, pazarlama harcamalarındaki her 1000 TL'lik artışın satışlarda ortalama 1500 TL'lik bir artışa yol açabileceği anlamına gelir. Bu katsayı, pazarlama stratejilerinin etkinliğini değerlendirmek için kritik bir göstergedir.

0.33

## Regresyon Sabiti

Örnek hesaplamada bulunan B0 değeri, bağımsız değişken sıfır olduğunda bağımlı değişkenin 0.33 değerini alacağını gösterir. Bu, temel satış düzeyini temsil eder; yani, pazarlama veya fiyatlandırma gibi faktörler olmasa bile, şirketin ortalama olarak elde edeceği satış miktarıdır. Bu değer, işin sürdürülebilirliği için önemlidir.

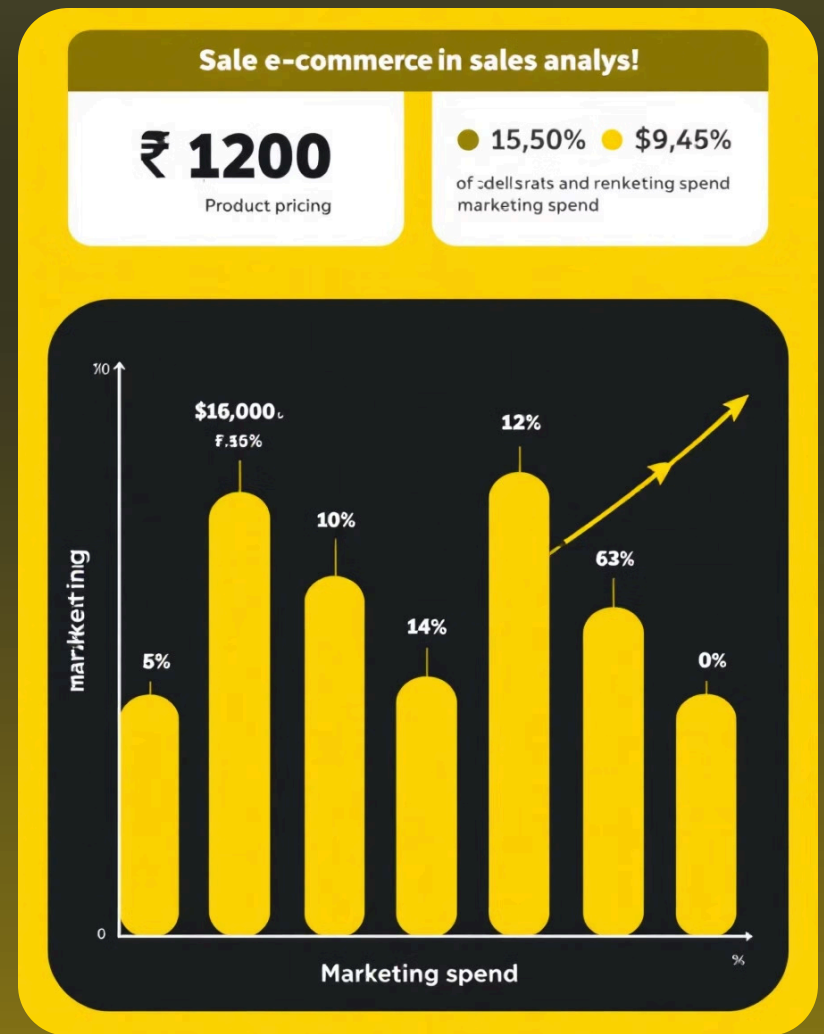
31%

## Fiyat Etkisi

Vaka çalışmasında, ürün fiyatının satışlar üzerindeki etkisi %31 olarak belirlenmiş, bu da fiyatın en güçlü belirleyici faktör olduğunu göstermiştir. Bu, fiyatlandırma stratejilerinde yapılacak değişikliklerin satışlar üzerinde önemli bir etkiye sahip olabileceği anlamına gelir. Örneğin, fiyat indirimleri kısa vadede satışları artırabilir, ancak uzun vadeli karlılığı etkileyebilir.

Bir e-ticaret şirketinin ürün satışlarını etkileyen faktörleri analiz eden regresyon modeli, ürün fiyatı, pazarlama harcaması ve satış promosyonlarını bağımsız değişkenler olarak incelemiştir. Sonuçlar, ürün fiyatının satışlar üzerindeki en güçlü belirleyici olduğunu ortaya koymuştur. Model, ayrıca, pazarlama harcamalarının ve satış promosyonlarının da satışlar üzerinde önemli etkileri olduğunu göstermiştir, ancak fiyat kadar belirleyici değildir. Bu analiz, şirketin pazarlama ve fiyatlandırma stratejilerini optimize etmesine yardımcı olabilir.

Örneğin, regresyon analizi sonuçlarına göre, şirket, ürün fiyatlarını belirli bir oranda düşürerek satış hacmini artırabilir. Ancak, bu fiyat indiriminin karlılık üzerindeki etkisini dikkatlice değerlendirmesi gerekmektedir. Aynı şekilde, pazarlama harcamalarını artırarak da satışları artırma potansiyeli bulunmaktadır, ancak bu harcamaların getirisi, regresyon modeli tarafından sağlanan verilerle ölçülmelidir. Bu vaka çalışması, regresyon modelinin e-ticaret şirketleri için stratejik kararlar alırken ne kadar değerli bir araç olabileceğini göstermektedir.



# Örnek Kod ve Uygulama

ID	Reklam Bütçesi (x)	Satış (y)
1	1.0	2.0
2	2.0	3.0
3	2.5	3.5
4	3.0	4.0
5	4.0	5.0
6	5.0	7.0
7	6.0	8.0
8	8.0	10.0

Yukarıdaki veri seti kullanılarak Python'da LinearRegression modeli oluşturulabilir. Kod, veri hazırlama, model eğitimi, tahmin yapma ve performans metriklerini hesaplama adımlarını içerir. Sonuçta, reklam harcaması ile satışlar arasındaki ilişkiyi gösteren bir model elde edilir.

Model denklemi  $y = \beta_0 + \beta_1 x$  formunda olacak ve MSE, MAE, R-kare gibi performans metrikleri hesaplanarak modelin başarısı değerlendirilecektir. Bu pratik uygulama, regresyon modellerinin gerçek dünya verilerine nasıl uygulanabileceğini göstermektedir.

```
/// regression analysis
// full regression 4/56,55019));
// regressant of line Regractit);
// with Regression 601,28);
{
  // yuul regression factien lata points
  (egressit this usset (4612));
  (one {
    regrassion final on data point {
      (egressiön draltinn, fir55,4);
      eguln cohirtinn ncture anluurdedtin);
      (outes 19);
      (iolucl (/ension ins 1 2015,4,5, - 491,4,))
      (egression reogrinatiesitrouughat, retlures: (5;-1) );
      {
        cetttion;
        in luney;
        cetation;
      }
      its { {
        (upnffice/fevales /nactitor-/degreant wuthelohers/cattien "ragres.cora/tin");
        (action drl(");
        filye;;
        (ually regriön = "15,15");
        /ora dasat. (55,-13(");
        /as gotien:
        /etüt/cns((n (o8)
        (/eraet,dotive points _6, (45,4));
      });
      (unr dut/late ins 6((12,"1");
      //feerect/ron=>d (01,1),486,14);
      /se caup/calstiens);
    });
  });
}
//st =);
//cesion:
//electon(áll));
```