



BATCH : **B 84 Data Science**
LESSON : **Machine Learning**
DATE : **17.03.2022**
SUBJECT : **Supervised Learning**



techproeducation



techproeducation



techproeducation



techproeducation



techproedu



MACHINE LEARNING - 4



Makine Öğrenmesi – 4



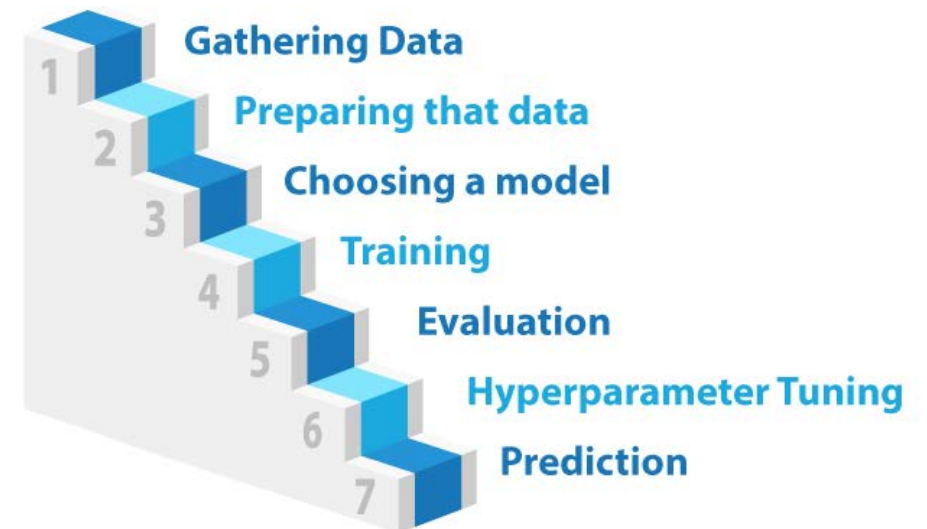
Overall Table of Contents



General Content

- ✓ Supervised Learning Algorithm - **Classification**
- ✓ Supervised Algorithm practices Python application
- ✓ Projects Solutions

7 steps of Machine Learning



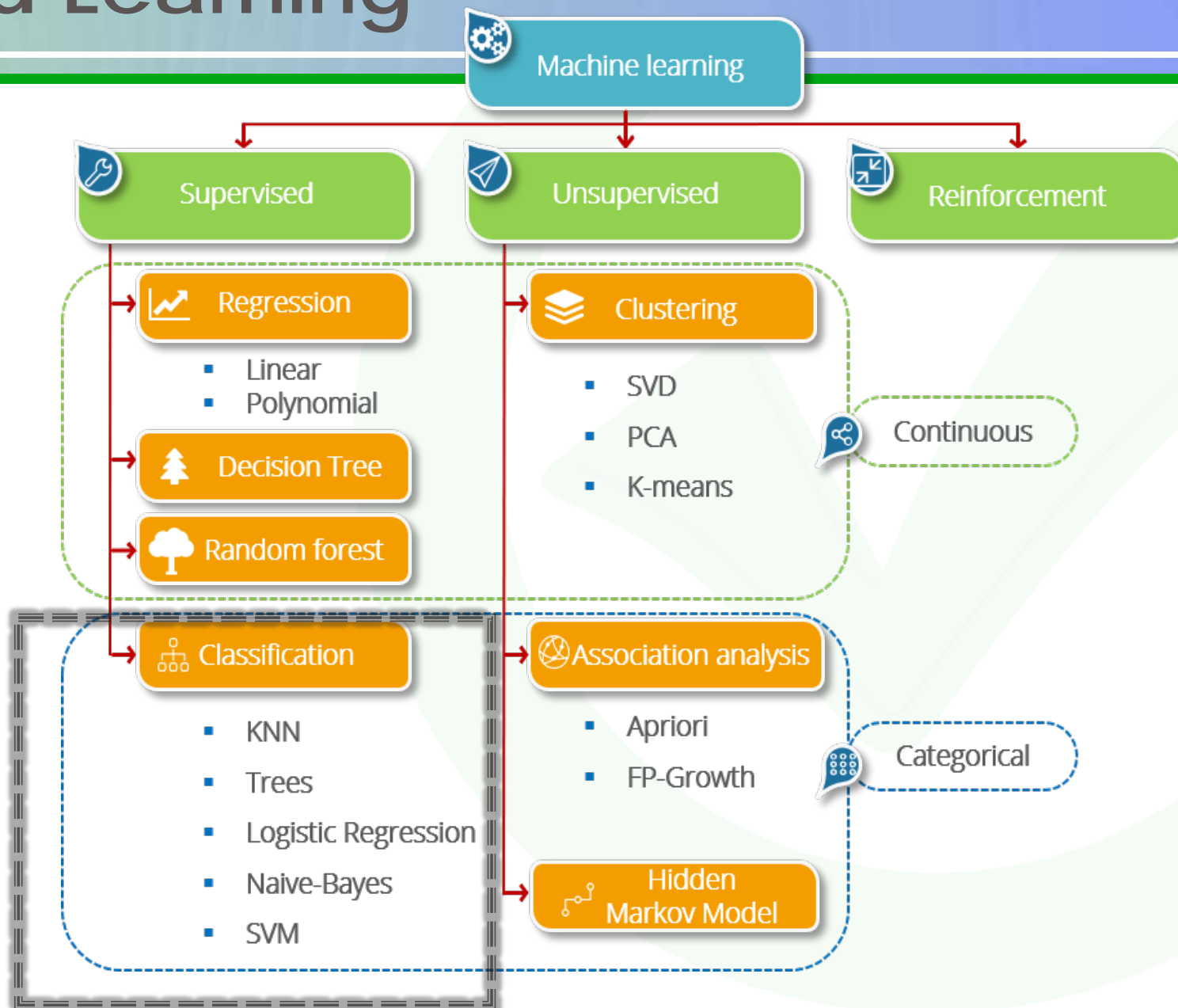


LOGISTIC REGRESSION

Supervised Learning



LOGISTIC REGRESSION

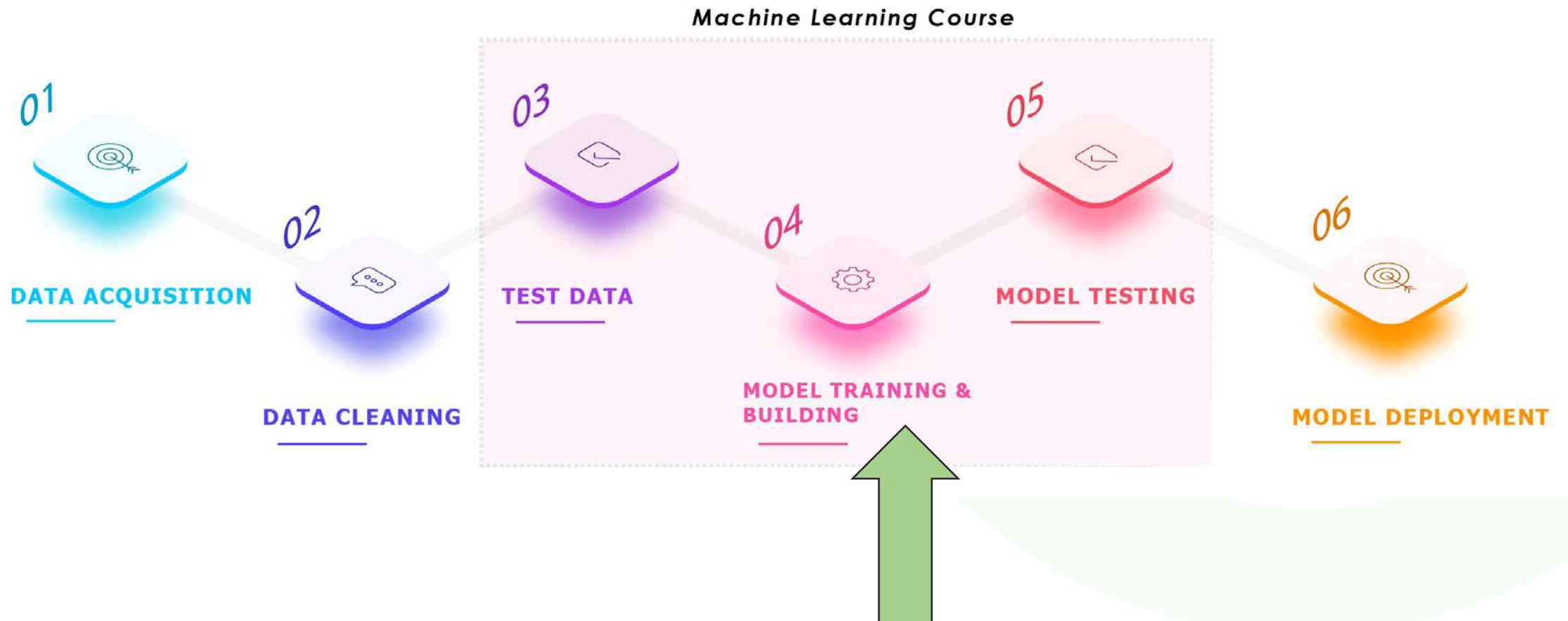




Run Navigation

Where are we?

DATA SCIENCE





Supervised Learning

LOGISTIC REGRESSION

- ✓ Titanic filmini hatırlayalım...
- ✓ İhtimaller...
- ✓ Adı regresyon ama.....
- ✓ Avantajları...



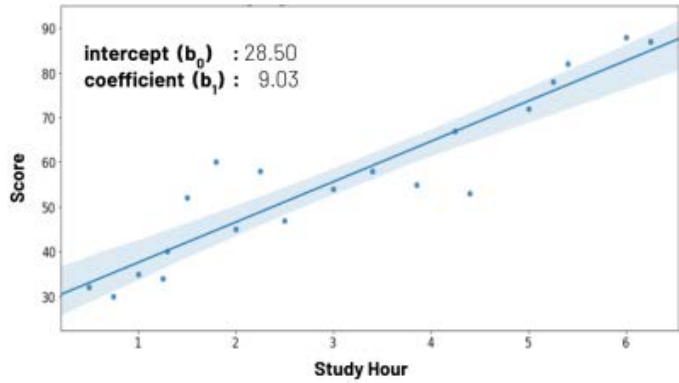
<i>İhtimaller</i>	<i>Rose'un Değerlendirmesi</i>
<i>Gemi batmadan varır.</i>	<i>Süper!</i>
<i>Gemi batar ama ikisi de kurtulur.</i>	<i>O da olur.</i>
<i>Gemi batar ve ikisi de kurtulamaz.</i>	<i>Birkaç dakika önce intihara yeltenen Rose için üzerinde durulacak bir ihtimal değil.</i>
<i>Gemi batar, Jack kurtulur, Rose kurtulmaz.</i>	
<i>Gemi batar, Rose kurtulur, Jack kurtulamaz.</i>	<i>Fecaat!</i>



Supervised Learning

LOGISTIC REGRESSION

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Study Hour	0.5	0.75	1.00	1.25	1.30	1.50	1.80	2.00	2.25	2.50	3.00	3.40	3.85	4.25	4.40	5.00	5.25	5.40	6.00	6.25
Score	32	30	35	34	40	52	60	45	58	47	54	58	55	67	53	72	78	82	88	87

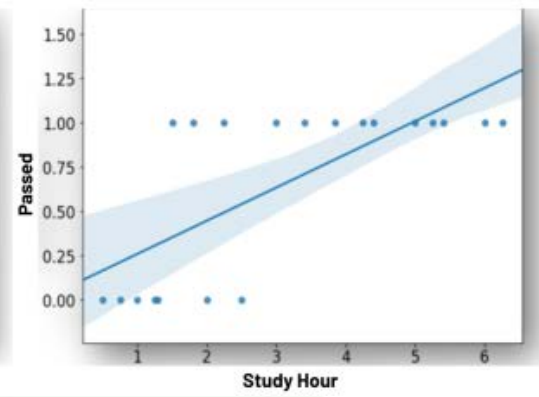
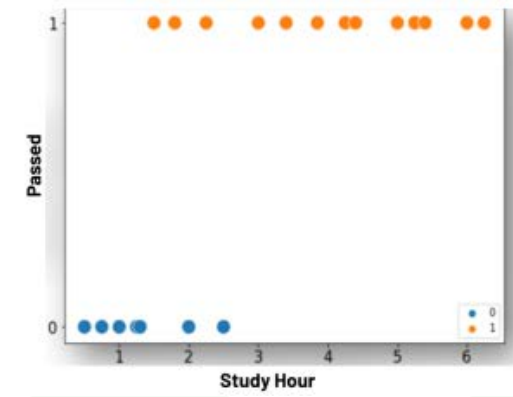


Regression Equation

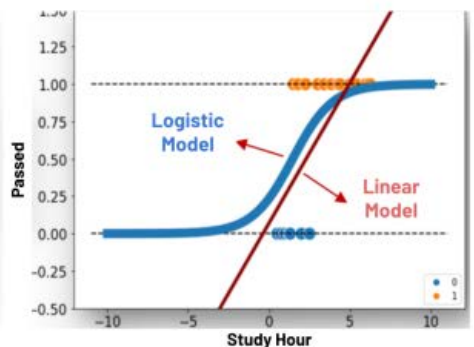
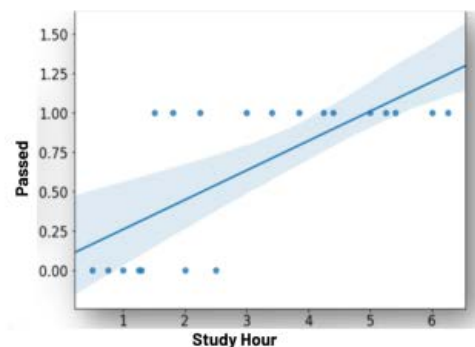
$$y = b_0 + b_1 * X$$

$$y = 28.50 + 9.03 * \text{Study hour}$$

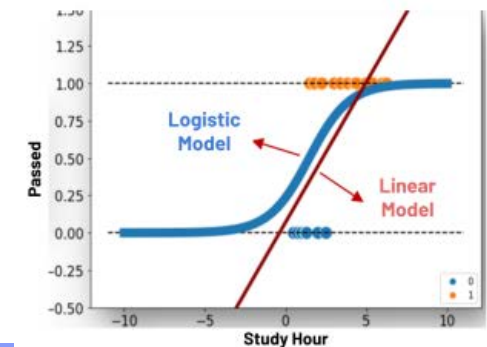
Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Study Hour	0.5	0.75	1.00	1.25	1.30	1.50	1.80	2.00	2.25	2.50	3.00	3.40	3.85	4.25	4.40	5.00	5.25	5.40	6.00	6.25
Passed	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1



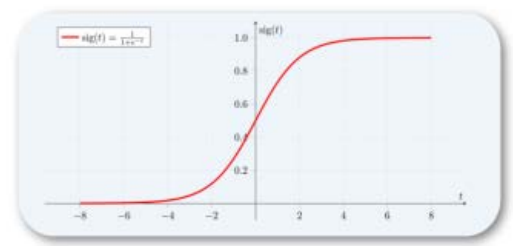
Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Study Hour	0.5	0.75	1.00	1.25	1.30	1.50	1.80	2.00	2.25	2.50	3.00	3.40	3.85	4.25	4.40	5.00	5.25	5.40	6.00	6.25
Passed	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1



Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Study Hour	0.5	0.75	1.00	1.25	1.30	1.50	1.80	2.00	2.25	2.50	3.00	3.40	3.85	4.25	4.40	5.00	5.25	5.40	6.00	6.25
Passed	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1



Sigmoid Function

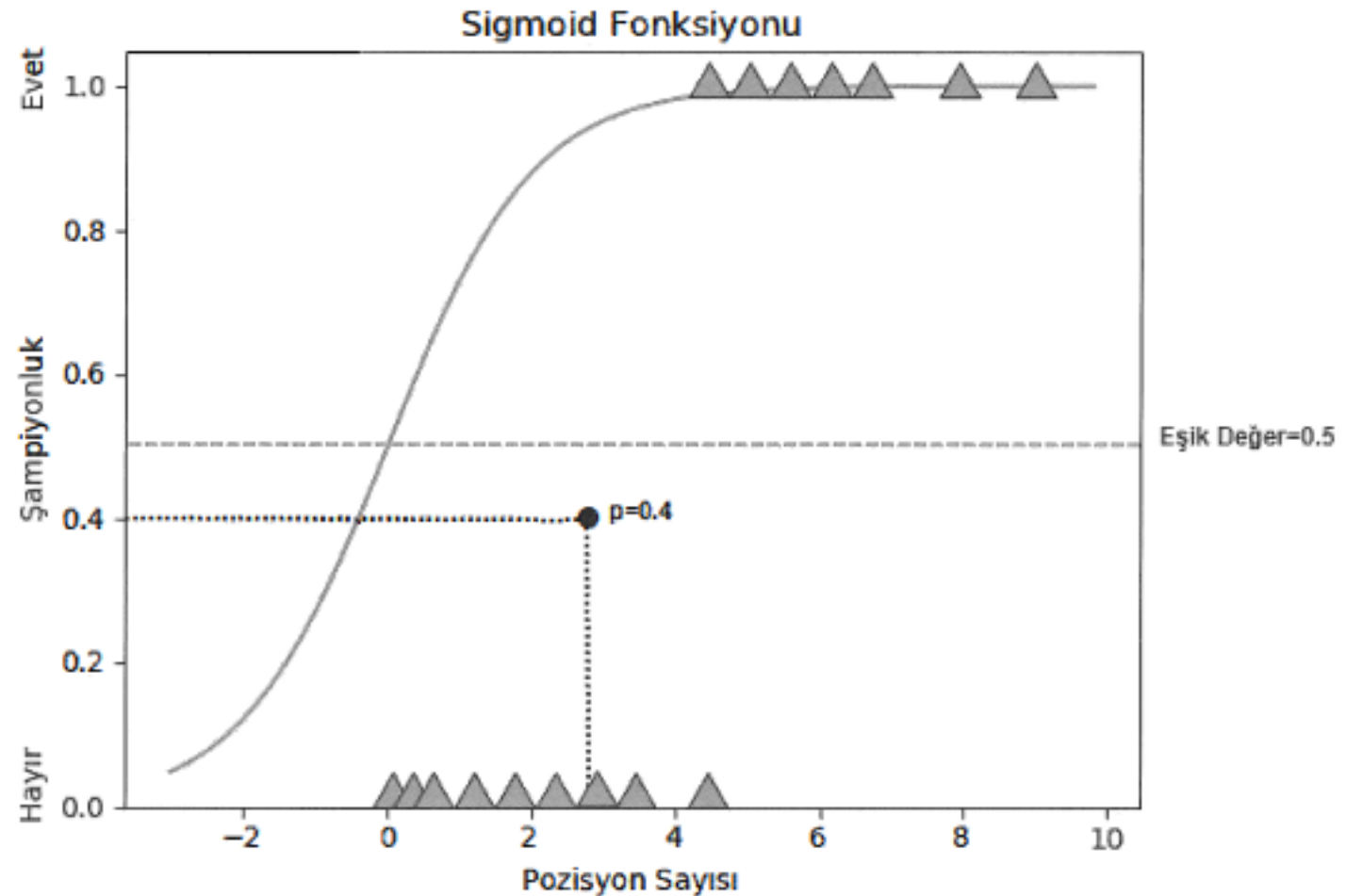




Supervised Learning

LOGISTIC REGRESSION

✓ Örnek

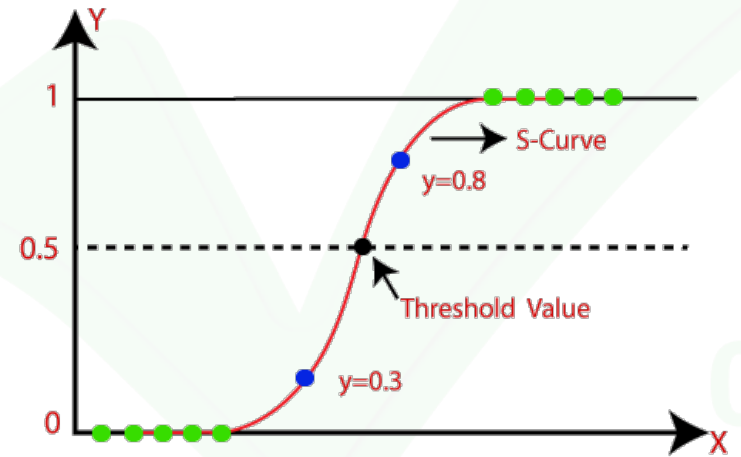
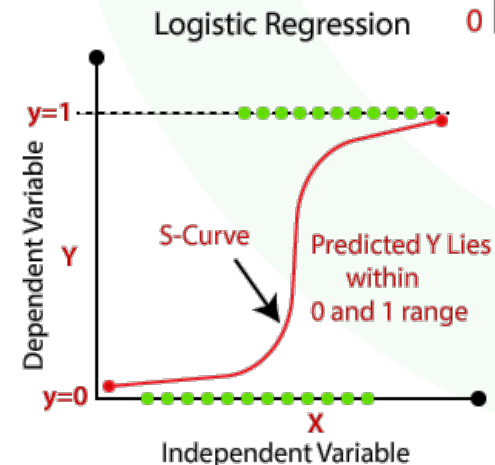
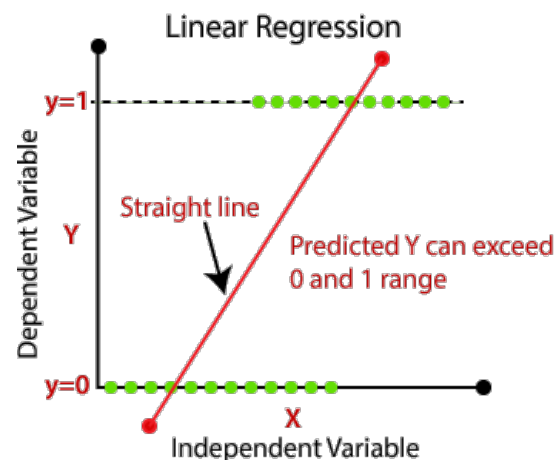
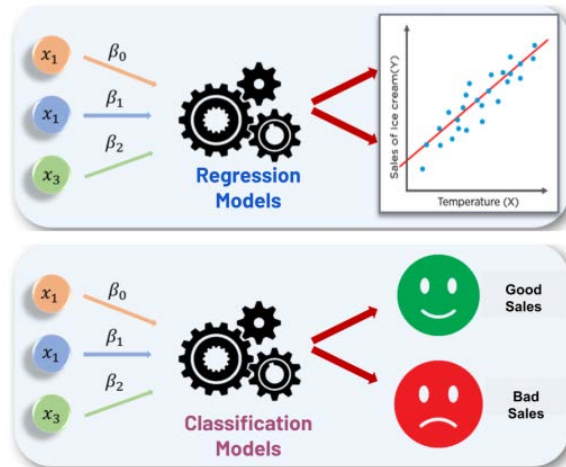
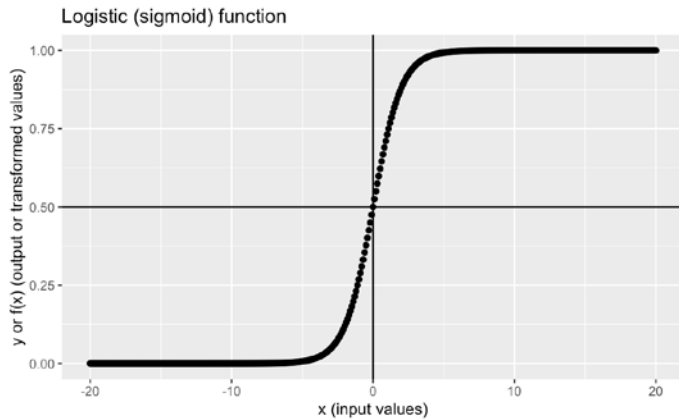




Supervised Learning

LOGISTIC REGRESSION

- ✓ Logit (sigmoid) fonksiyon
- ✓ Eşik değer katsayısı (discriminant threshold) -





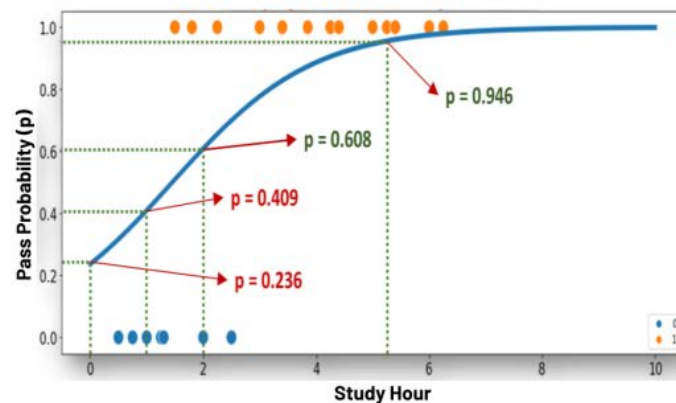
Supervised Learning

LOGISTIC REGRESSION

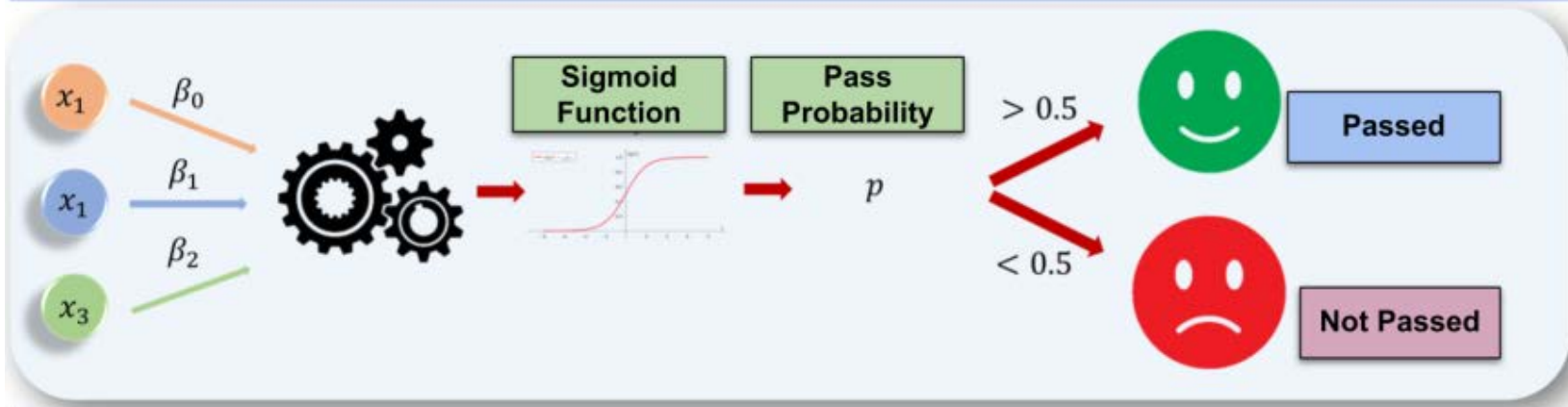


Logit (sigmoid) fonksiyon

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Study Hour	0.5	0.75	1.00	1.25	1.30	1.50	1.80	2.00	2.25	2.50	3.00	3.40	3.85	4.25	4.40	5.00	5.25	5.40	6.00	6.25
Passed	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1



Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Study Hour	0.5	0.75	1.00	1.25	1.30	1.50	1.80	2.00	2.25	2.50	3.00	3.40	3.85	4.25	4.40	5.00	5.25	5.40	6.00	6.25
Passed	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1

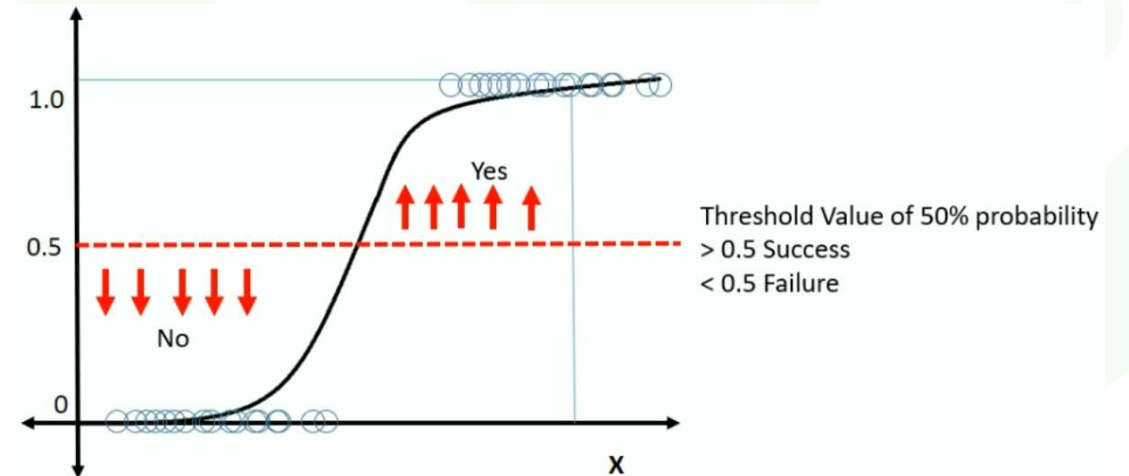
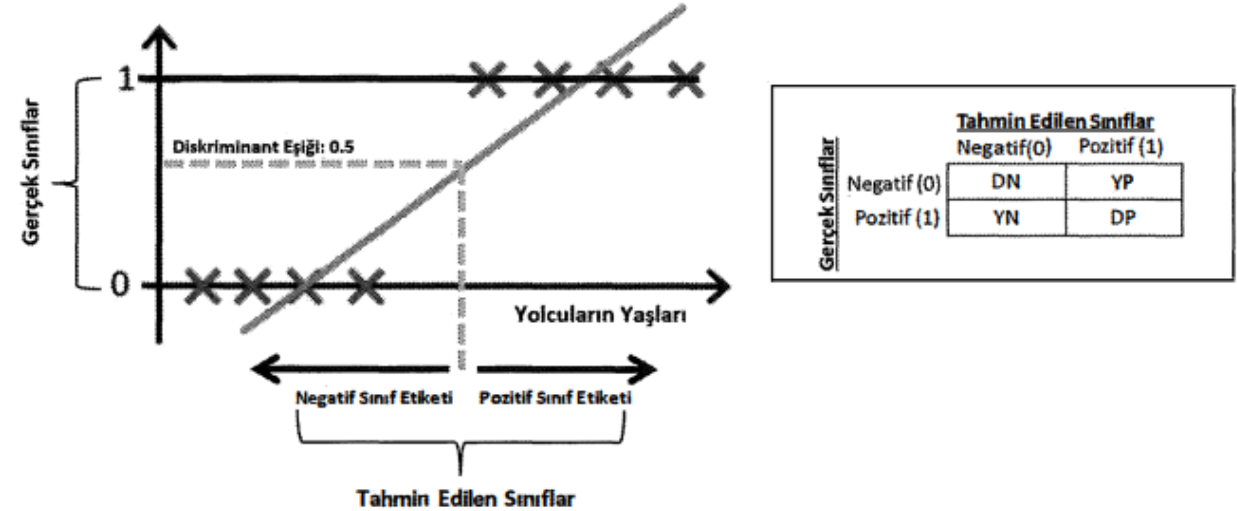
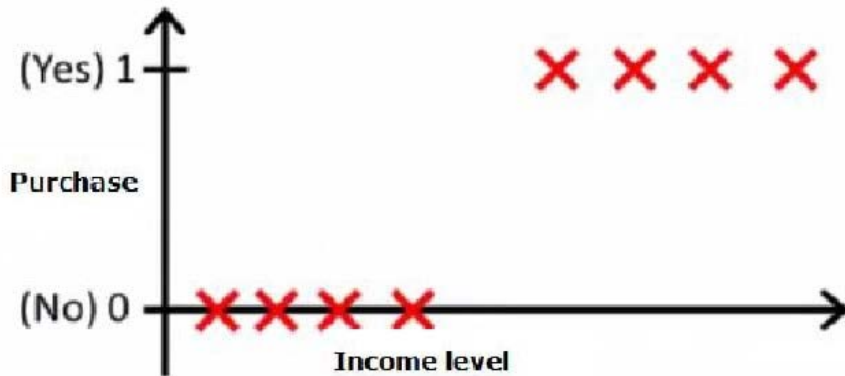




Supervised Learning

LOGISTIC REGRESSION

✓ Yaş ve hayatta kalma ihtimaliyetini inceleyelim





Supervised Learning

LOGISTIC REGRESSION

- ✓ sigmoid fonksiyonu → logit fonksiyonu → logistik regresyon süreci
- ✓ Titanic veri seti hakkında..

Sigmoid Fonksiyonu → Logit Fonksiyonu → Logistik Regresyon

$$y = \alpha_0 + \beta_1 x$$

$$p(t) = \frac{1}{1 + e^{-t}}$$

$$p(x) = \frac{1}{1 + e^{-(\alpha_0 + \beta_1 x)}}$$

$$g(x) = \ln \frac{p(x)}{1 - p(x)} = \alpha_0 + \beta_1 x = y$$

Doğrusal regresyonda y hedef değişkeni x öznitelik değişkeni ile tahmin edilir. β katsayısı özniteliğin hedef değişken üzerindeki etkisini ya da hedef değişkendeki değişimlere katkısını gösterir.

Sigmoid fonksiyonu bir öznitelik değişkeninin değerlerini bir olayın olma olasılığına indirger. Değişkenin gözlem değerleri ne olursa olsun sigmoid fonksiyonu bunları 1 ve 0'lara indirger.

Doğrusal regresyon fonksiyon kalıbının sigmoid fonksiyonu ile geliştirilmesi **Logistik fonksiyonu**, $p(x)$, verir. Buna göre hedef değişkene konu olan bir olayın olması öznitelik değişkeninin hedef değişkene katkı yapma olasılığına bağlıdır.

Logit fonksiyonu, $g(x)$, ise aslında logistik fonksiyonun tersidir. Buna göre bir olayın olma olasılığı (hedef değişkenin doğru pozitif olması) özniteliklerin doğrusal kombinasyonuna bağlıdır.

Veri Setinin Açıklamaları:

PassengerId: Yolcu numarası

Survived: Kaza sonucu yolcunun kurtulup kurtulmadığı ile ilgili sınıf etiketi. 1: kurtuldu; 0: kurtulamadı. Bu etiket sadece eğitim setinde mevcuttur.

Pclass: Bilet Sınıfı. Aynı zamanda sosyoekonomik statü indikatörü (1: Üst Sınıf, 2: Orta Sınıf, 3: Alt Sınıf).

Names: Yolcunun adı ve soyadı

Sex: Yolcunun cinsiyeti (female: kadın; male: erkek)

Age: Yolcunun yaşı

SibSp: Yolcunun beraber seyahat ettikleri ailesindeki kişi sayısı. Kardeşler ve eşlerin toplam sayısı.

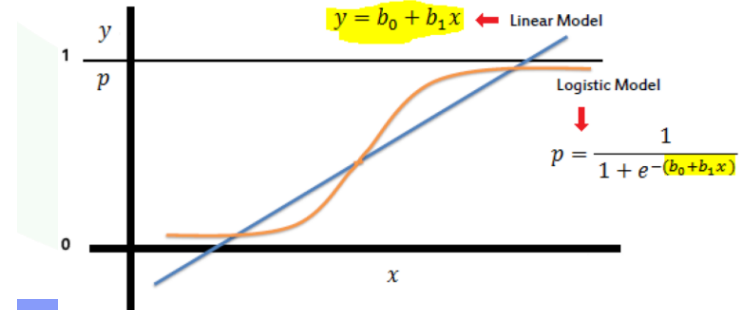
Parch: Yolcunun beraber seyahat ettikleri ailesindeki kişi sayısı. Annesi, babası, çocukları. Eğer çocuk dadısı ile seyahat ediyorsa parch=0 olarak kaydedilmiştir.

Ticket: Bilet numarası

Fare: Bilet ücreti

Cabin: Kabin numarası

Embarked: Yolcunun gemiye bindiği liman (C: Cherbourg, Q: Queenstown, S: Southampton)



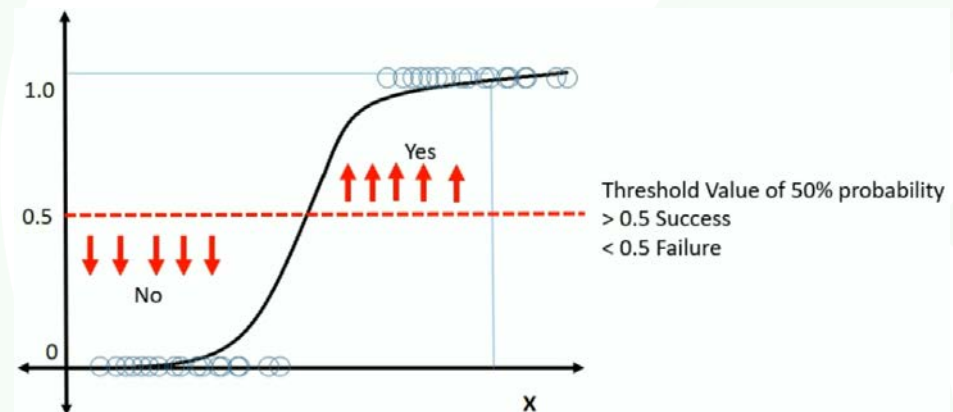
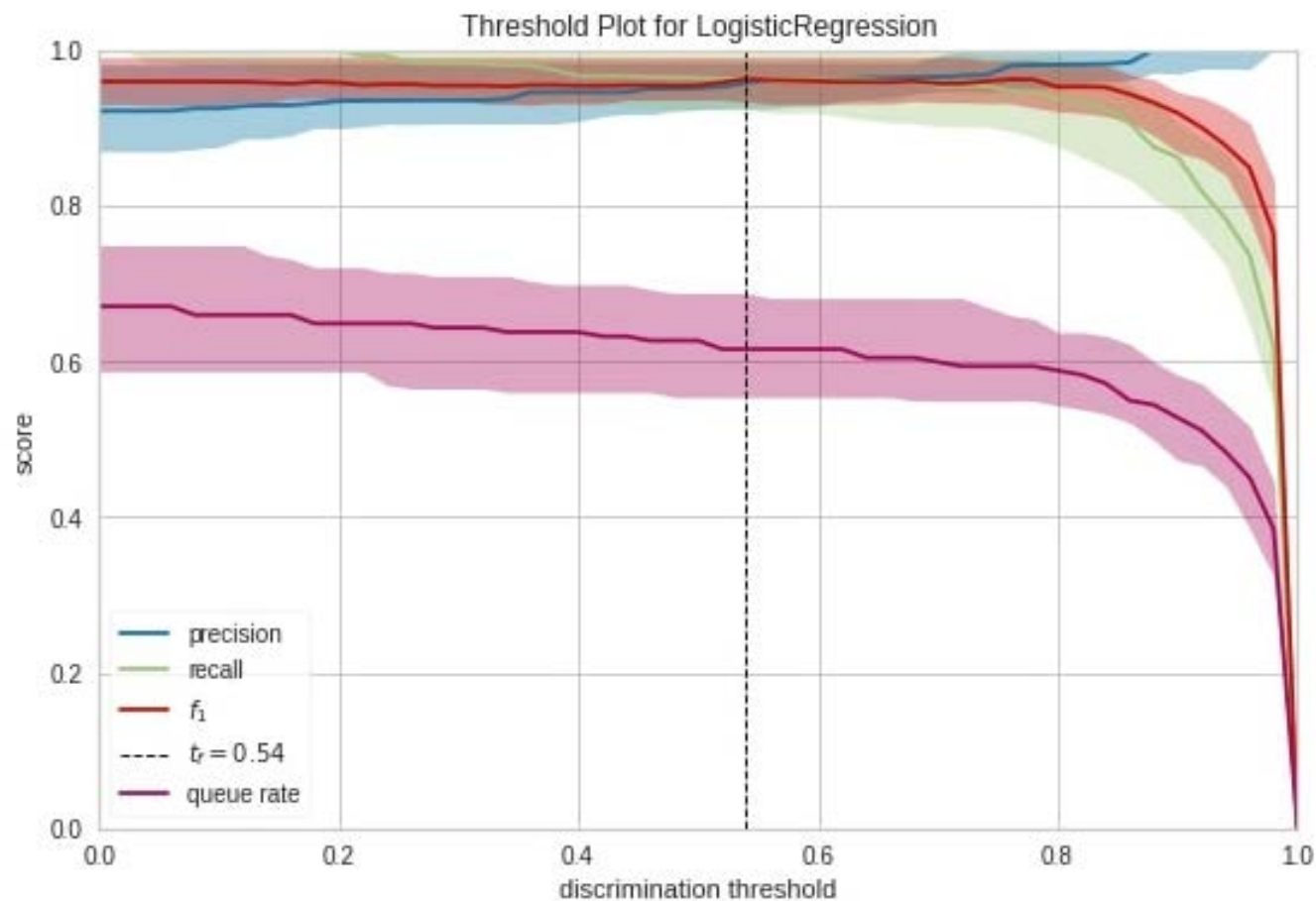


Supervised Learning

LOGISTIC REGRESSION



Yellowbrick kullanımı



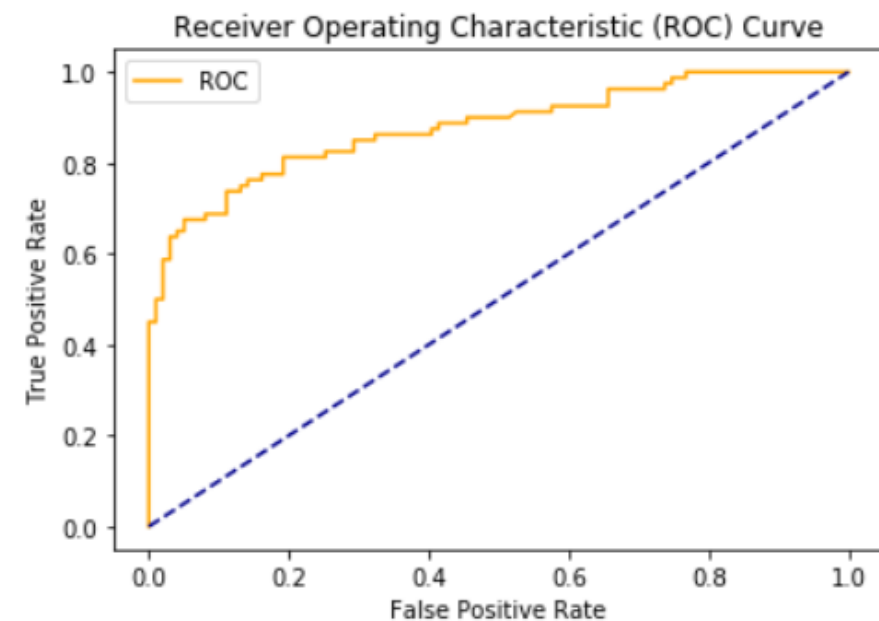
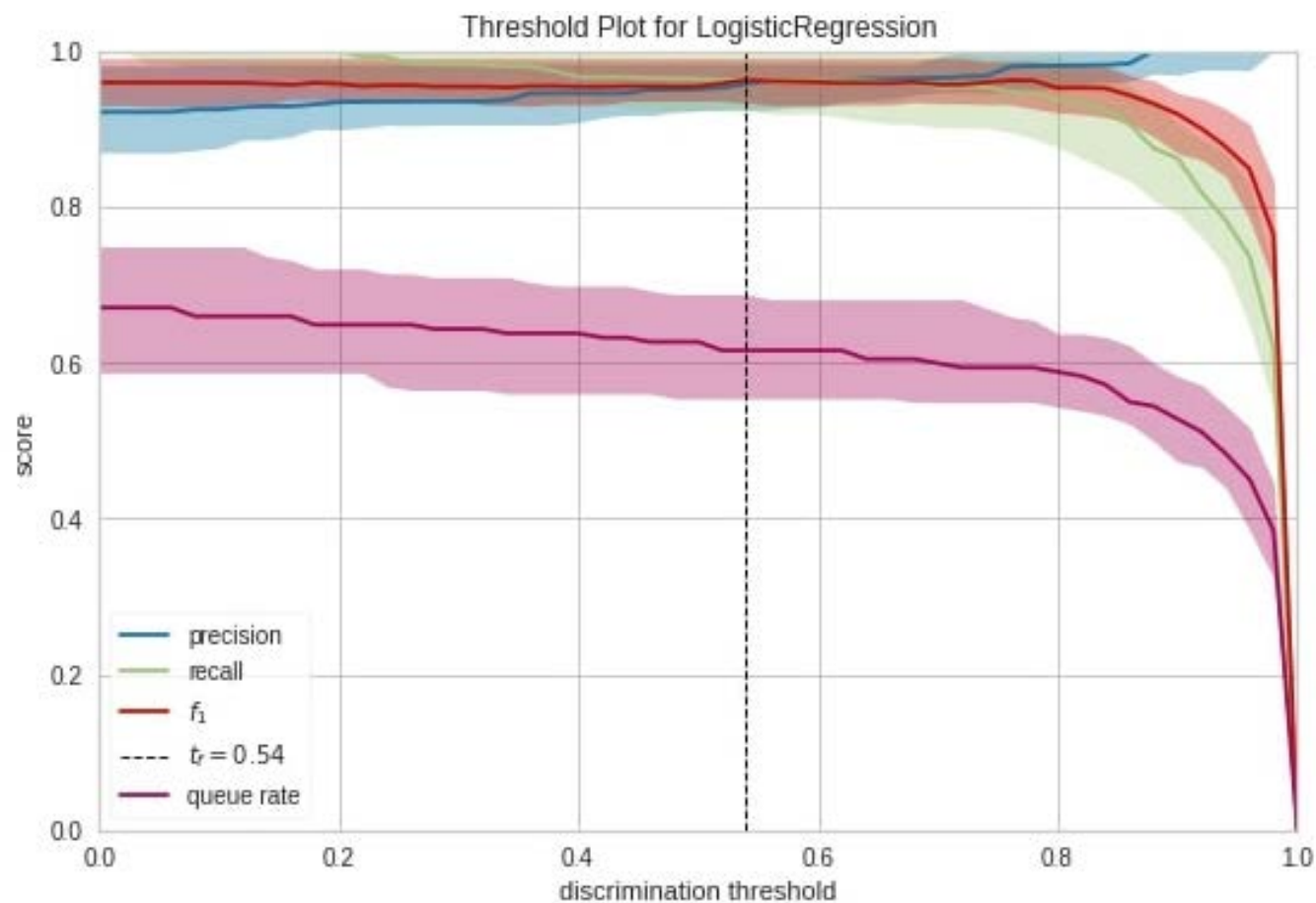


Supervised Learning

LOGISTIC REGRESSION



Precision-recall curve (PRC) grafiği anlamı



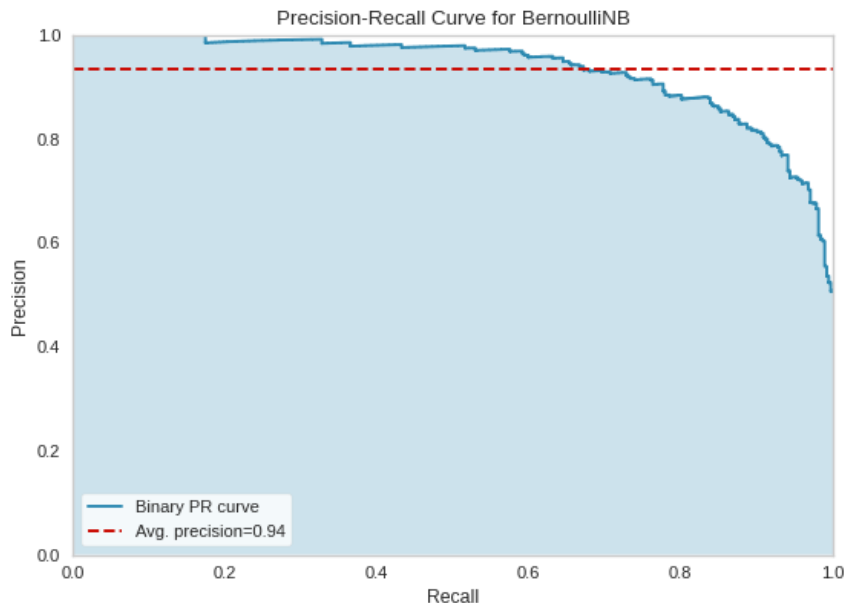
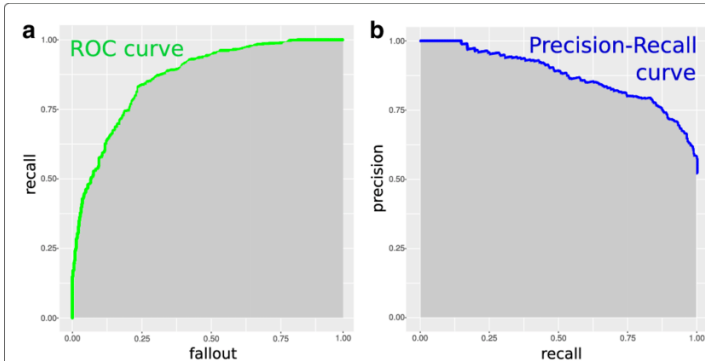


Supervised Learning

LOGISTIC REGRESSION



Precision-recall curve (PRC) grafiği anlamı



	Predicted class POSITIVE (spam ☒)	Predicted class NEGATIVE (normal ☐)	
Actual class POSITIVE (spam ☒)	TRUE POSITIVE (TP) ☒ ☒ 320	FALSE NEGATIVE (FN) ☒ ☐ 43	$Recall = \frac{TP}{TP + FN} = \frac{320}{320 + 43} = 0.882$
Actual class NEGATIVE (normal ☐)	FALSE POSITIVE (FP) ☐ ☒ 20	TRUE NEGATIVE (TN) ☐ ☐ 538	
	$Precision = \frac{TP}{TP + FP} = \frac{320}{320 + 20} = 0.941$		

ROC: Balanced Data
PRC: Unbalanced data



Where are we?

DATA SCIENCE

