

# Makine Öğrenmesi



[istdatascience.com](http://istdatascience.com) | [info@istdsa.com](mailto:info@istdsa.com)



[tr.linkedin.com/school/istanbul-data-science-academy](https://tr.linkedin.com/school/istanbul-data-science-academy)



[medium.com/istanbuldatascienceacademy](https://medium.com/istanbuldatascienceacademy)

# ÖĞRENMEK NE DEMEK?

## ÖĞRENMEK:

Geçmişte öğrendiklerimizden ve tecrübelerimizden yararlanıp gelecekte alacağımız kararları geçmişten dersler çıkararak daha iyi bir şekilde verebilme süreci.

## MAKİNE ÖĞRENMESİ:

Makine öğrenmesi (Machine Learning), tükettikleri verilere göre öğrenen ya da performansı iyileştiren sistemler oluşturmaya odaklanan yapay zekanın (AI) bir alt kümesidir. Yapay zeka, insan zekasını taklit eden sistemler veya makineler anlamına gelen kapsamlı bir terimdir.

# NEDEN BÖYLE BİR YAKLAŞIMA İHTİYAÇ VAR?



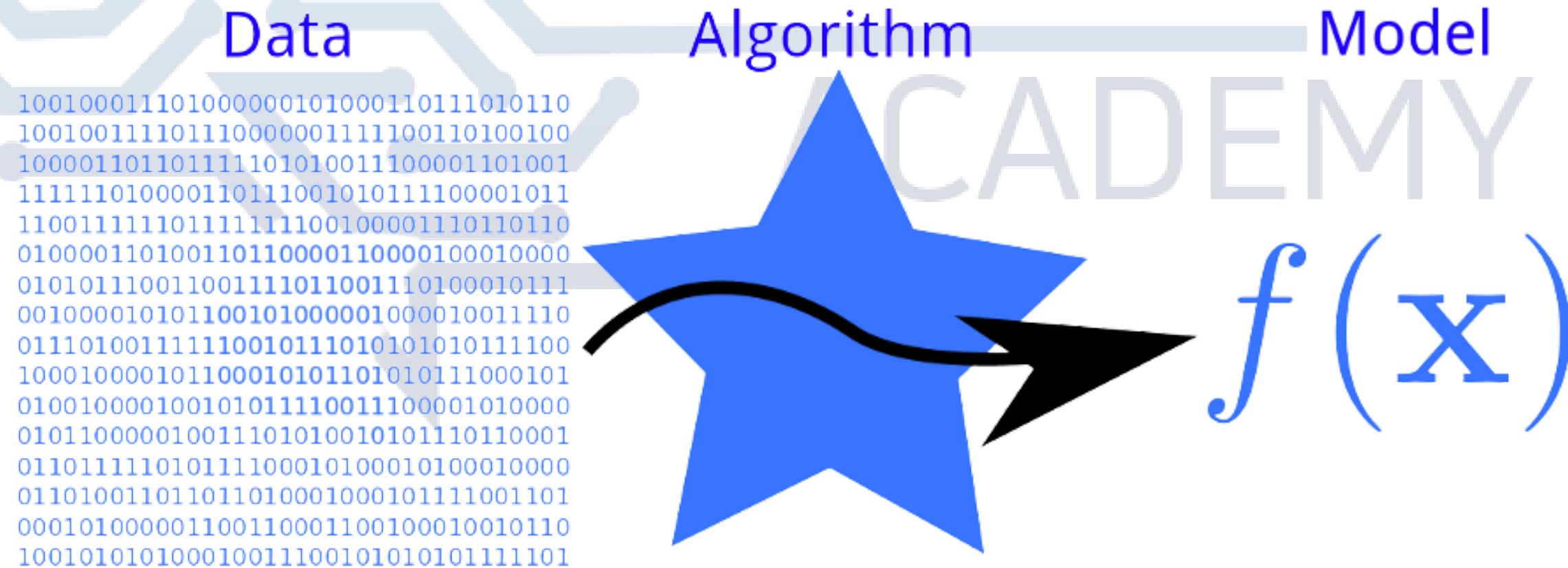
4x4	LECTURES			
	MATHEMATICS	PHYSICS	LITERATURE	GRADE
1	85	92	78	A
2	56	41	83	C
3	62	29	47	D
4	71	73	89	?

50x50	LECTURES			
	MATHEMATICS	PHYSICS	LITERATURE	GRADE
1	85	92	78	A
2	56	41	83	C
3	62	29	47	D
...	...	...	...	...
50	71	73	89	?

- Model, aslında arka planda oldukça karmaşık olan büyük bir dünyayı kapsayan küçük bir kavramdır.
- İyi bir model, önemsiz olan detayları ortadan kaldırıp, her zaman gerçekten önemli olan detaylarla yola devam eder.
- İş hayatında çoğu şirkette model ve algoritma, birbirlerinin yerine kullanılan kavramlardır. Ancak doğru tanımlama şu şekilde olmalıdır;
  - "Makine öğrenmesindeki bir "algoritma", bir makine öğrenmesi "modeli" oluşturmak için veriler üzerinde çalıştırılan bir prosedürdür."

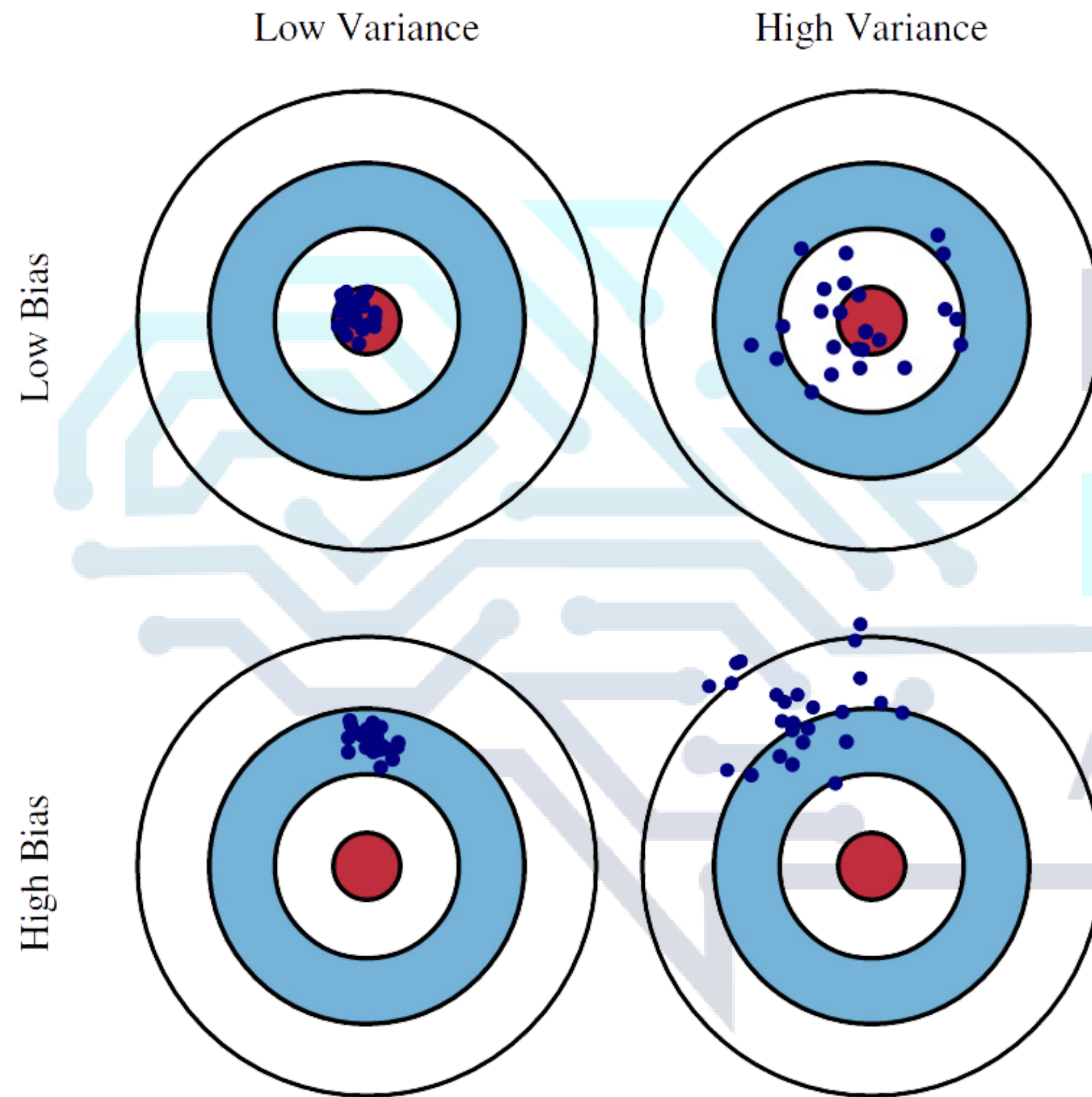
# ALGORİTMA VS MODEL

- Makine öğrenmesi algoritmaları elindeki verileri inceleyerek bir "örüntü tanıma" gerçekleştirir. Yani, algoritmalar verilerden "öğrenir".
- Makine öğrenmesindeki bir "model", veriler üzerinde çalışan bir makine öğrenmesi algoritmasının çıktısıdır. Bir model, bir makine öğrenmesi algoritması tarafından öğrenilenleri temsil eder.

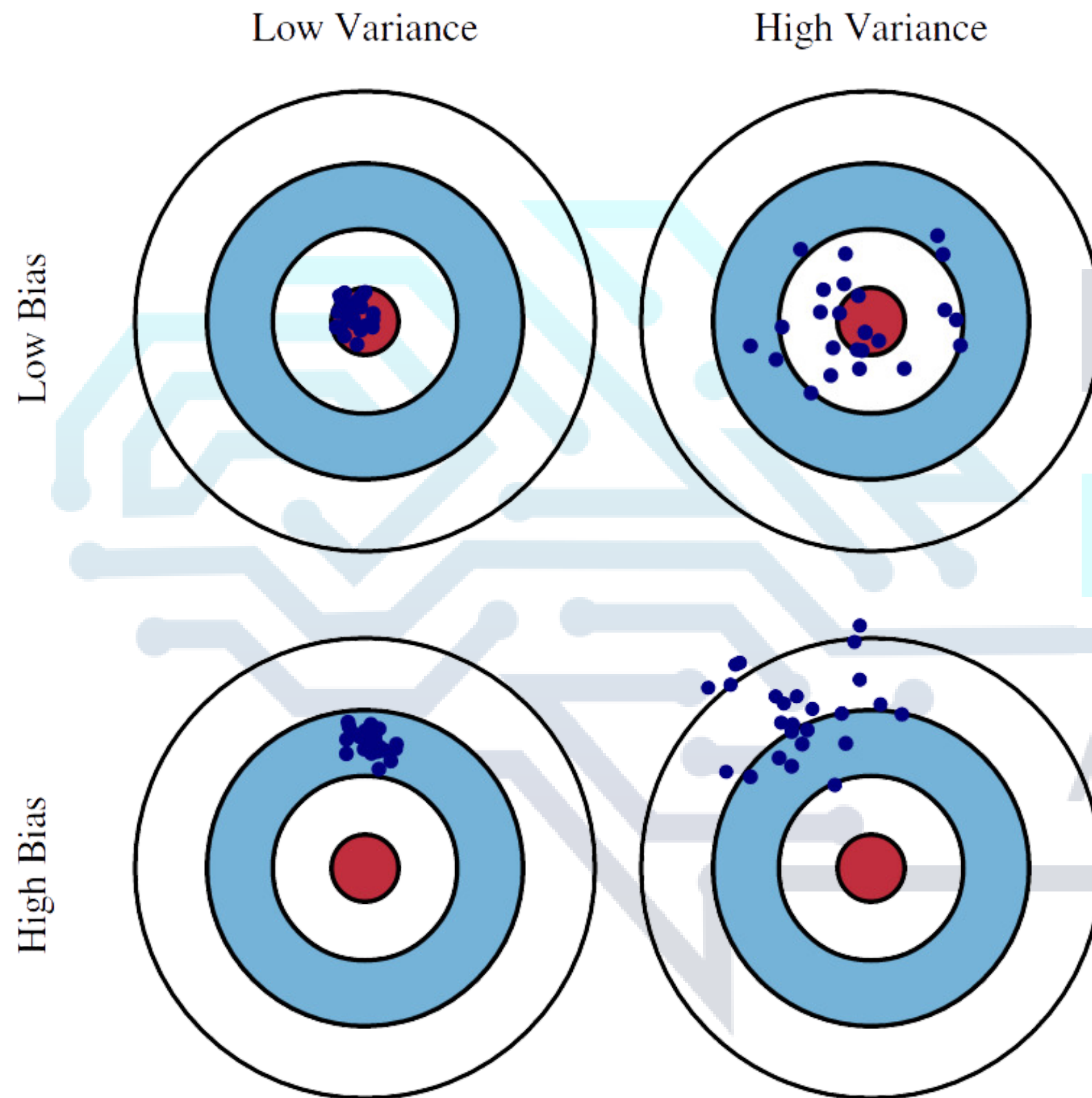




# BIAS & VARIANCE (ÖNYARGI & VARYANS)



# BIAS & VARIANCE (ÖNYARGI & VARYANS)

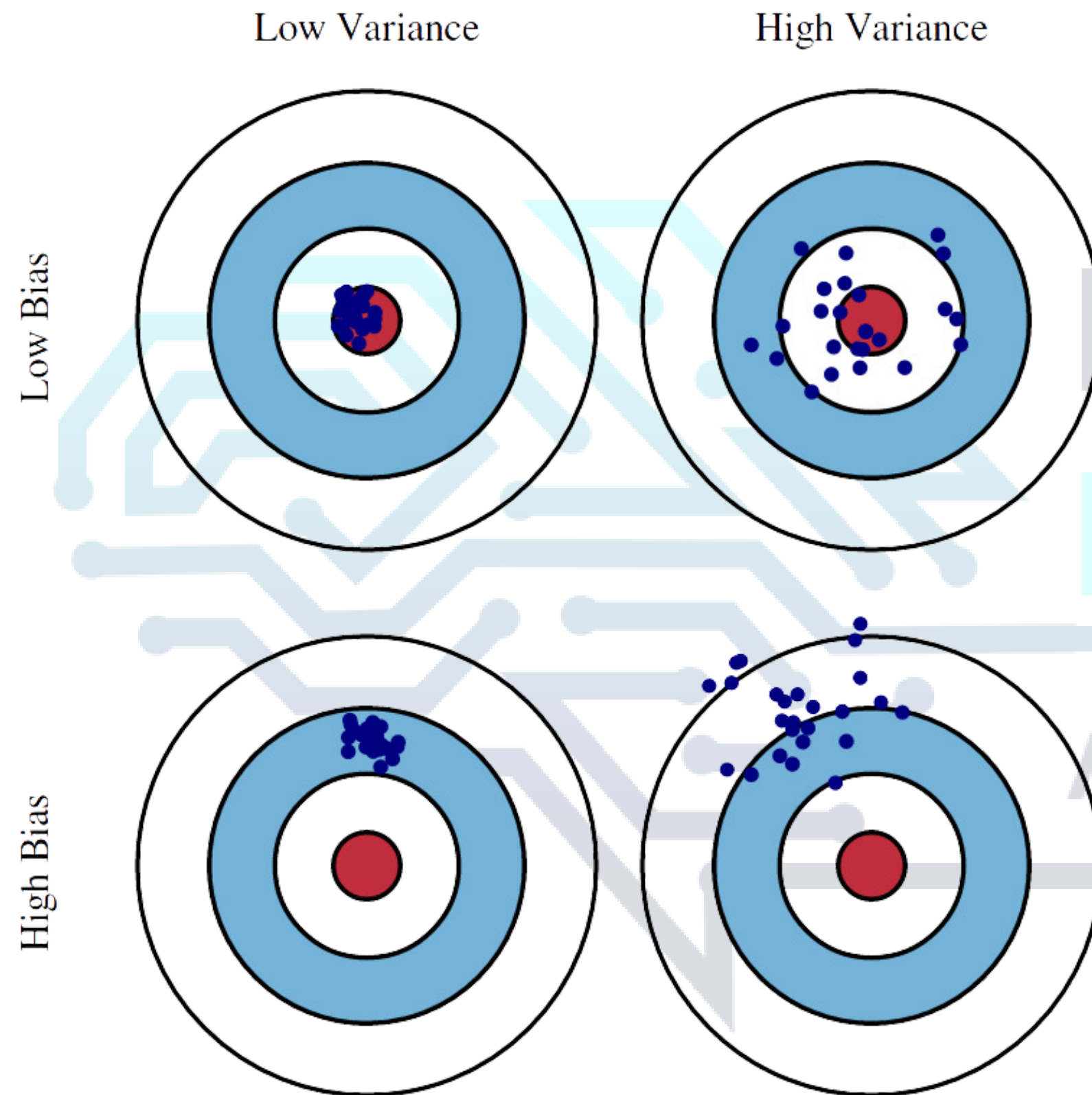


- **HIGH BIAS**





# BIAS & VARIANCE (ÖNYARGI & VARYANS)

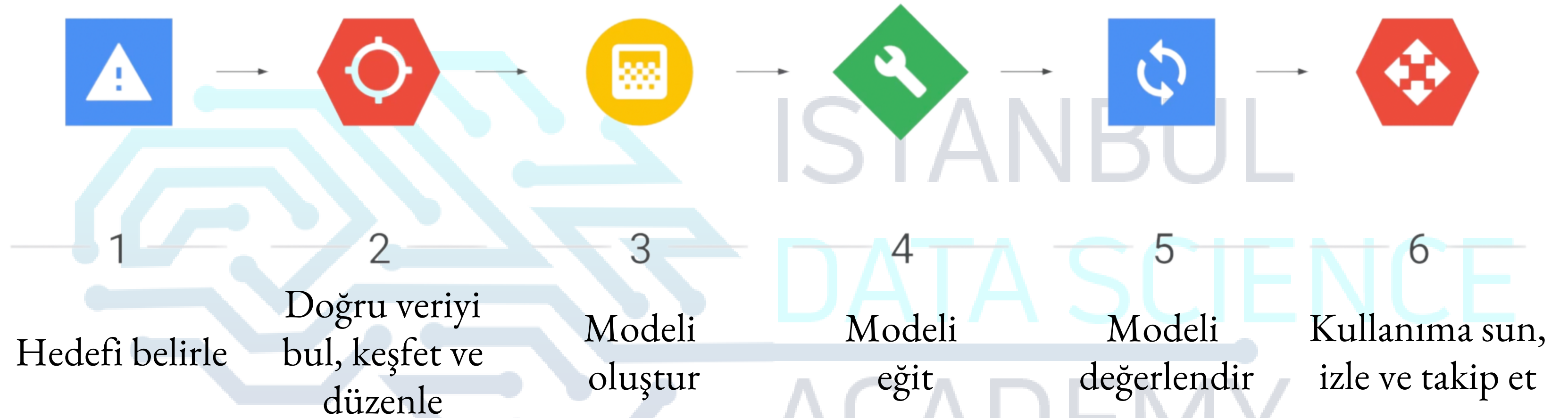


- HIGH BIAS

- HIGH VARIANCE



# MAKİNE ÖĞRENMESİ SÜRECİ



## X (Input - Girdi):

- **Observations (Gözlemler):** Kullandığımız veri seti içerisindeki satırlar
- **Features (Özellikler):** Kullandığımız veri seti içerisindeki sütunlar
- İlk projede kullandığımız observationlara ve featurelara ait bir kesit :)

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	02/27/2021	03:00:00	REGULAR	7537255	2570434
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	02/27/2021	07:00:00	REGULAR	7537260	2570445
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	02/27/2021	11:00:00	REGULAR	7537285	2570500
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	02/27/2021	15:00:00	REGULAR	7537354	2570546
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	02/27/2021	19:00:00	REGULAR	7537501	2570582

## X (Input - Girdi):

- **Observations (Gözlemler):** Kullandığımız veri seti içerisindeki satırlar
- **Features (Özellikler):** Kullandığımız veri seti içerisindeki sütunlar

## Y (Output - Çıktı):

- **Y:** Kullandığımız veri seti içerisinde bulunan "Output" sütunu
- **$Y_p$  :** Model geliştirme aşamasından sonra tahmin edilen "Output" sütunu
- Makine öğrenmesi sonucunda tahmin ettiğimiz **output sütununun türüne göre** makine öğrenmesini **temel olarak 2 gruba** ayırabiliriz.



## X (Input - Girdi):

- **Observations (Gözlemler):** Kullandığımız veri seti içerisindeki satırlar
- **Features (Özellikler):** Kullandığımız veri seti içerisindeki sütunlar

## Y (Output - Çıktı):

- **Y:** Kullandığımız veri seti içerisinde bulunan "Output" sütunu
- **$Y_p$  :** Model geliştirme aşamasından sonra tahmin edilen "Output" sütunu
- **Regression (Regresyon):** Output sütunumuz sayısal değer
  - Ürün Fiyatı Tahmini, Evin  $M^2$  Tahmini, Maaş Tahmini, Kilo Tahmini, ...



## X (Input - Girdi):

- **Observations (Gözlemler):** Kullandığımız veri seti içerisindeki satırlar
- **Features (Özellikler):** Kullandığımız veri seti içerisindeki sütunlar

## Y (Output - Çıktı):

- **Y:** Kullandığımız veri seti içerisinde bulunan "Output" sütunu
- **$Y_p$  :** Model geliştirme aşamasından sonra tahmin edilen "Output" sütunu
- **Classification (Sınıflandırma):** Output sütunumuz kategorik değer
  - Müşteri Davranışı Tahmini, Hasta Durumu Tahmini, Kupaya Hangi Takım Sahip Olacak Tahmini, ...

## X (Input - Girdi):

- **Observations (Gözlemler):** Kullandığımız veri seti içerisindeki satırlar
- **Features (Özellikler):** Kullandığımız veri seti içerisindeki sütunlar

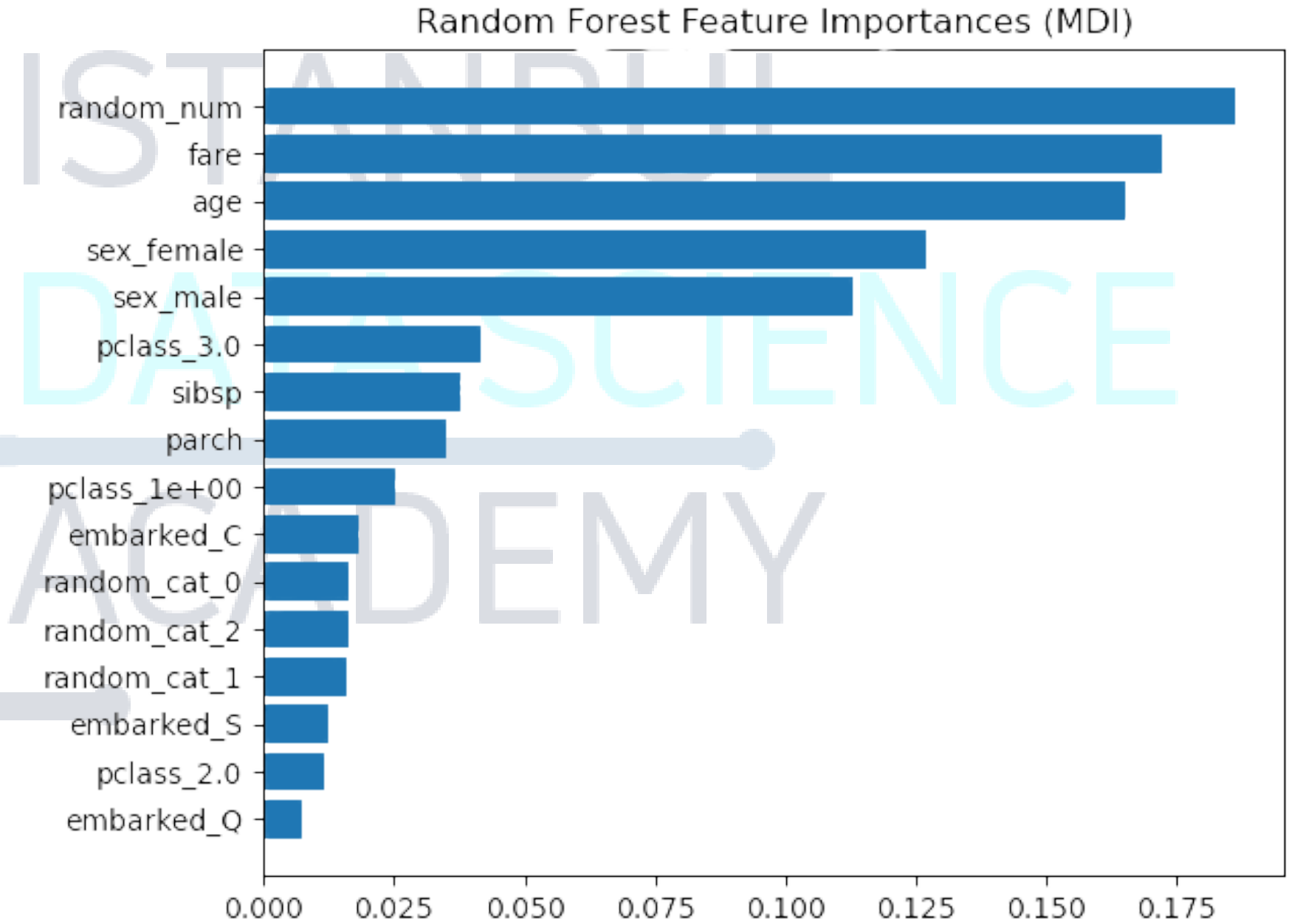
## Y (Output - Çıktı):

- **Y:** Bu sefer bu bilgi elimizde yok???
- **$Y_p$  :** Model geliştirme aşamasından sonra tahmin edilen "Output" sütunu
- **Clustering (Kümeleme):** Output sütunumuz **kategorik değer**
  - Müşterileri Gruplama, Video Hizmetleri, Sporcuları Gruplama, ...

# 2 TEMEL YAKLAŞIM TARZI

## 1. Featureların Önemini Anlamak

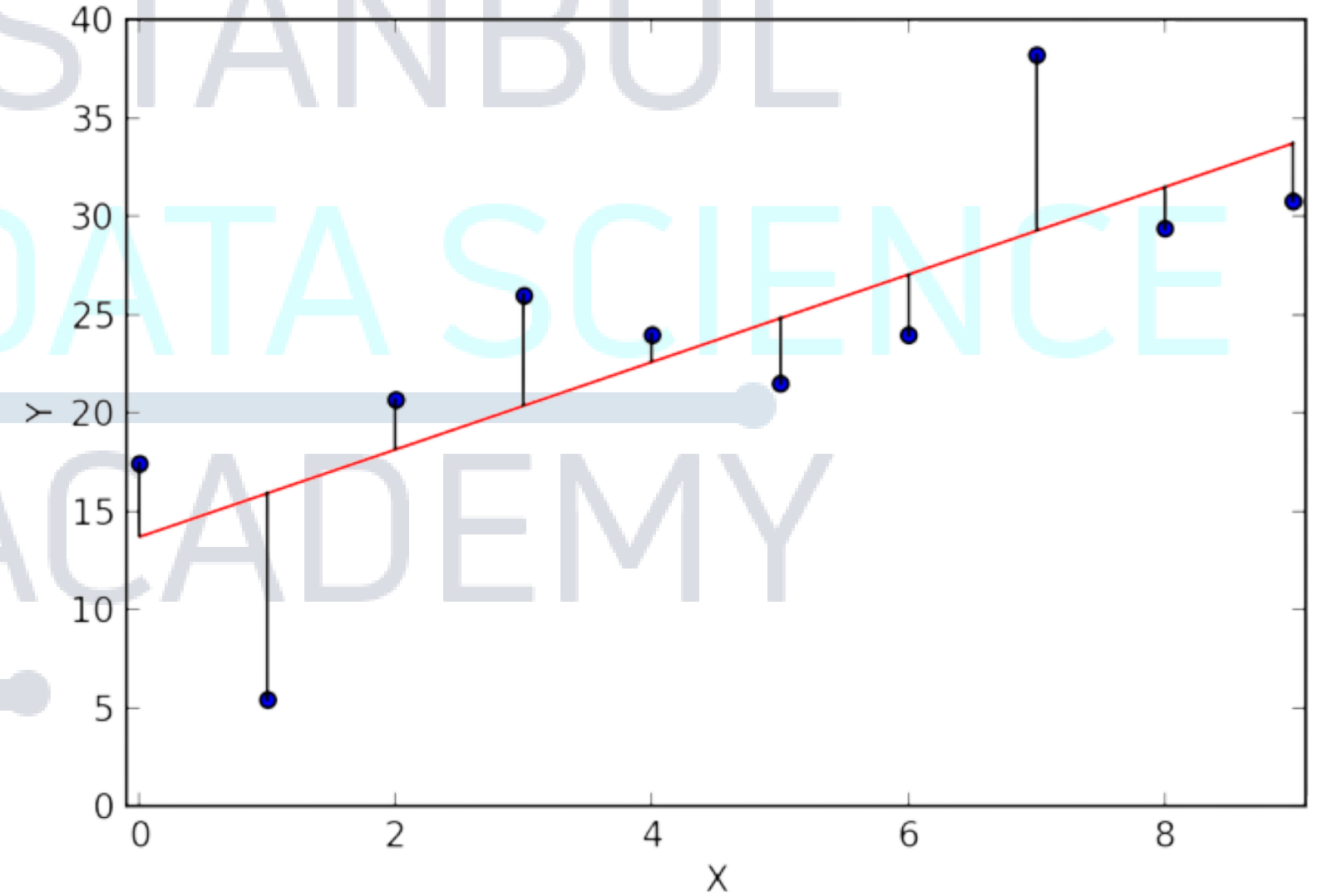
- Model geliştirme aşamasında genellikle ana amaç, en iyi tahmini yapabilmek için en uygun featureları belirleyerek onları amacımıza göre düzenlemektir.
- Ancak bazen de amaç, featureların output sütunu üzerinde yarattıkları etkiyi görmektir.



# 2 TEMEL YAKLAŞIM TARZI

## 2. En İyi Tahminleri Yapabilmek

- Bir çok çalışmada ana amaç, gerçeğe en yakın tahminler üreten modeli geliştirmektir. Ancak her zaman değil!
- Çalışmada ulaşmak istediğimiz sonuca göre bu durumda bazı değişiklikler olabilir. (Bunun ne anlama geldiğini ileride detaylı bir şekilde işleyeceğiz)



# 1. YAKLAŞIM + 2. YAKLAŞIM = GENEL YAKLAŞIM

## 3. Genel Yaklaşım Tarzı

- Bir çok çalışmada hem 1. yaklaşım tarzı hem de 2. yaklaşım tarzı beraber kullanılırlar.
- Bunun da temel sebebi, featurelar hakkında elde ettiğimiz bilgilerin model geliştirme aşamasında oldukça önemli bir rol oynamasıdır ki bizim de veri bilimciler olarak öncelikli amacımız bu :)



- Makine öğrenmesi, yapay zekanın bir alt kümesi olup, tahmine dayalı modeller geliştirmeye dayanan bir alandır.
- Makine öğrenmesi algoritmaları, geçmişteki deneyimlerden elde ettiği çıkarımlardan yararlanarak geleceğe yönelik tahminler yaparlar.
- Geliştirilen modeller 2 amaç uğruna kullanılırlar;
  1. Output sütunundaki değeri etkileyen featureların önemini kavramak
  2. Geleceğe yönelik tahminler yapmak