



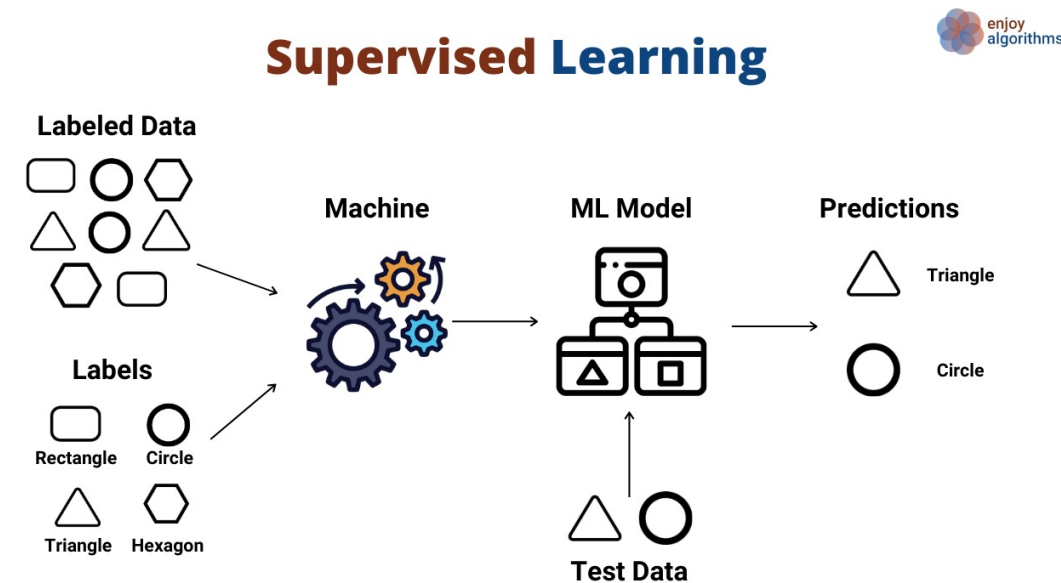
Bài giảng môn học:  
**Học máy (Machine Learning)**

# CHƯƠNG 3: HỌC CÓ GIÁM SÁT (Supervised Learning)

Giảng viên: Đặng Văn Nam  
Email: [dangvannam@hmg.edu.vn](mailto:dangvannam@hmg.edu.vn)

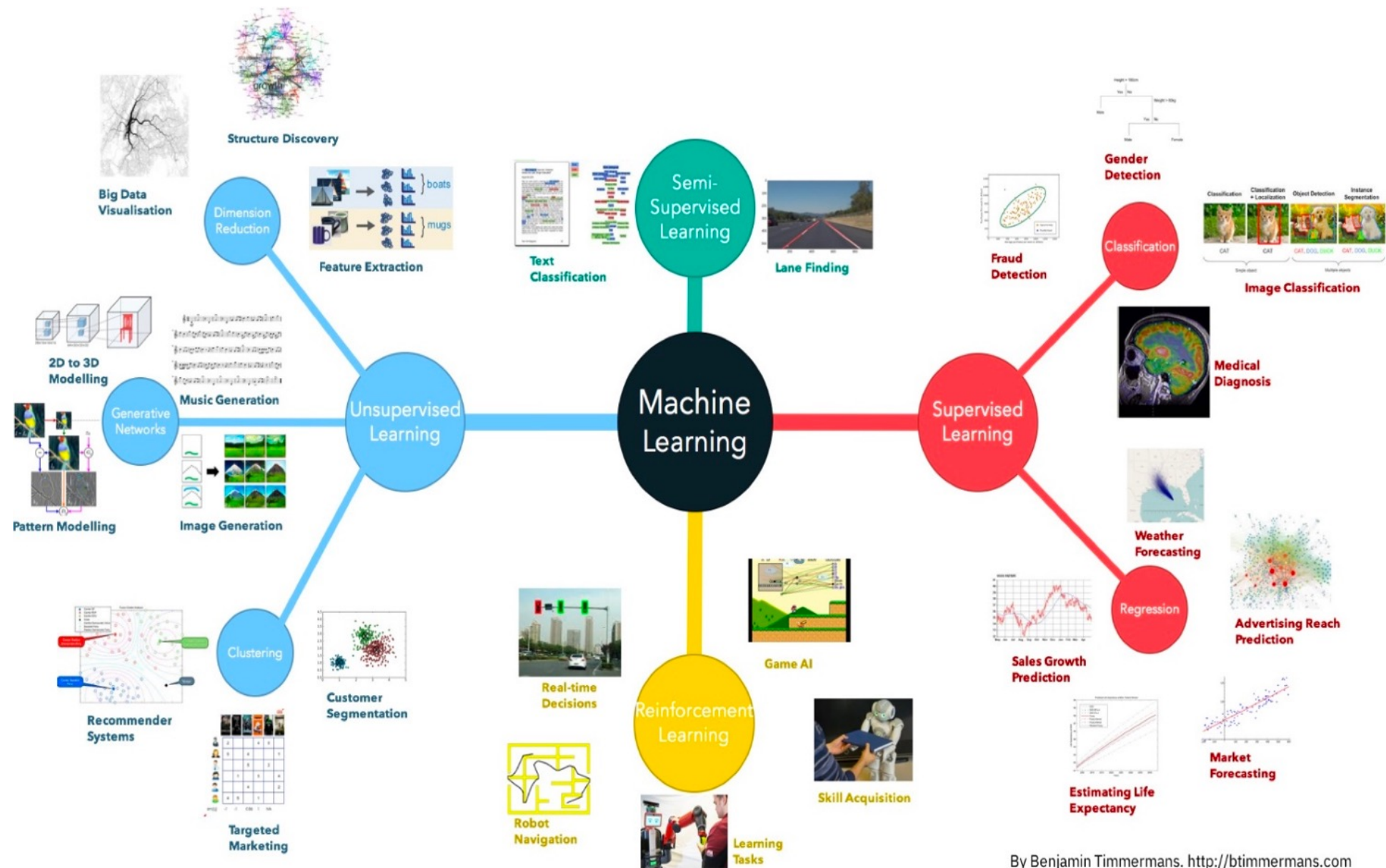
# Nội dung chương 3

1. Giới thiệu học có giám sát (Supervised Learning)
2. Phân loại học có giám sát (Classification – Regression)
3. Một số thuật toán quan trọng



# 1. Giới thiệu học có giám sát (supervised learning)

# Học có giám sát là gì?



# Học có giám sát là gì?

Một thuật toán học máy được gọi là học có giám sát (supervised learning) nếu việc xây dựng mô hình dự đoán **mối quan hệ giữa đầu vào (X) và đầu ra (y)** được thực hiện dựa trên các cặp (X - y) đã biết trong tập huấn luyện.

**Tập dữ liệu học (Training data) bao gồm các quan sát (X), mà mỗi quan sát được gắn kèm với một giá trị đầu ra mong muốn (Label - y)**

Sample



Label



dog

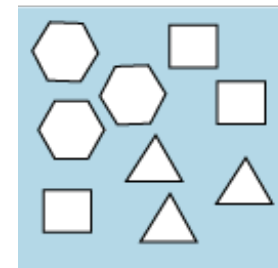


cat



horse

Labeled Data



Labels



Đây là nhóm thuật toán **phổ biến nhất trong các thuật toán Học máy**.

# Học có giám sát là gì?

Supervised Learning (Học có giám sát) là một nhóm thuật toán sử dụng **dữ liệu được gán nhãn** nhằm mô hình hóa mối quan hệ giữa biến đầu vào ( $X$ ) và biến đầu ra ( $y$ )

Dataset sẽ bao gồm:

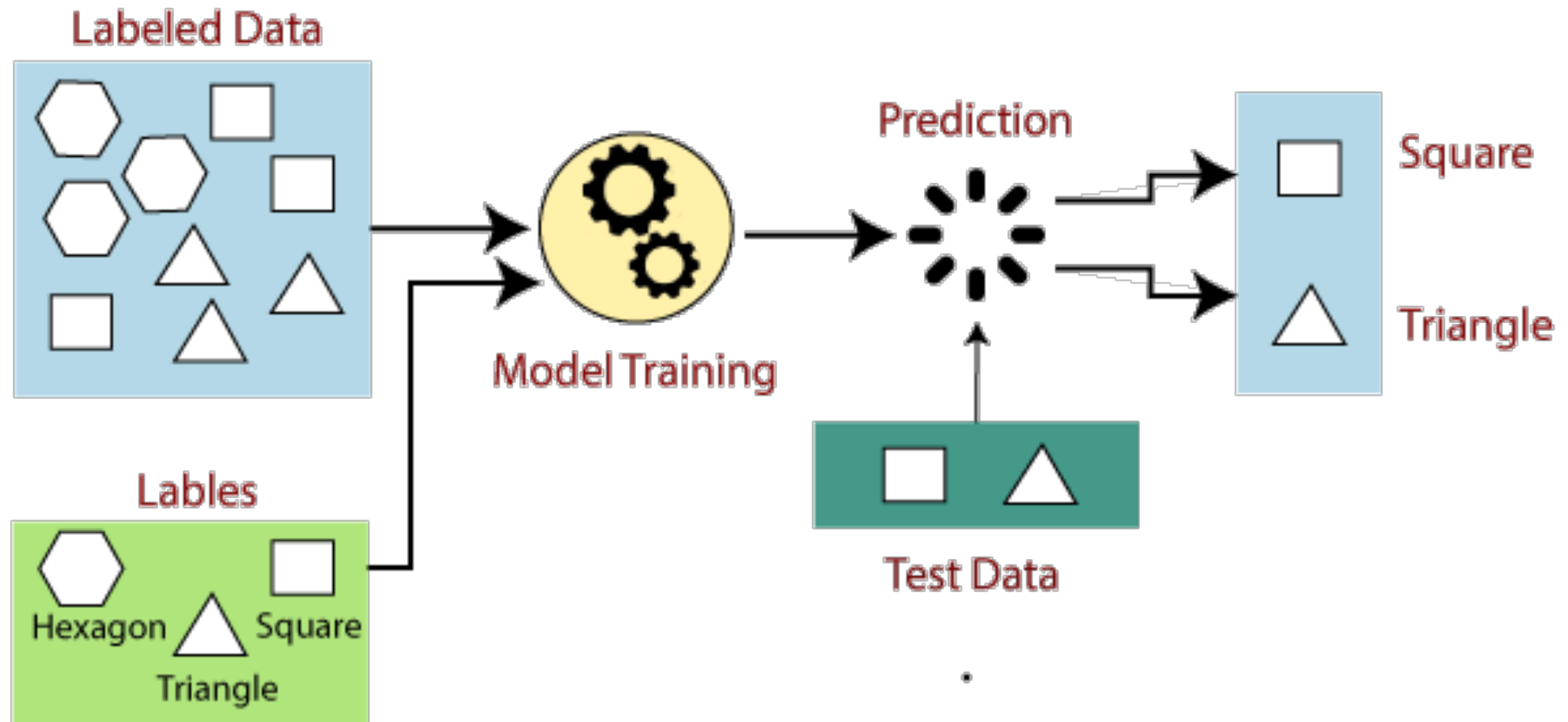
- Các thuộc tính đầu vào ( $X$  - Biến độc lập) – Features (inputs)
- Thuộc tính đầu ra ( $y$  - Biến phụ thuộc) – Target (label)

Biến độc lập  
(Features)
Biến phụ thuộc  
(Target) - Label

id	Age	Gender	Type	Blood_pressure	Cholesterol	Heartbeat	Thalassemia	Result
Patient_01	63	Male	Typical angina	145	233	150	6	0
Patient_02	67	Male	Asymptomatic	160	286	108	3	1
Patient_03	67	Male	Asymptomatic	120	229	129	7	1
Patient_04	37	Male	Non-anginal pain	130	250	187	3	0
Patient_05	41	Female	Atypical angina	130	204	172		0
Patient_16	56	Male	Atypical angina	120	236	178	3	0
Patient_07	62	Female	Asymptomatic	140	268	160	3	1
Patient_08	57	Female	Asymptomatic	120	354	163	3	0
Patient_19	63	Male	Asymptomatic	130	254	147	7	1
Patient_10	53	Male	Asymptomatic	140	203	155	7	1
Patient_110	57	Male	Asymptomatic	140	192	148	6	0
Patient_120	56	Female	Atypical angina	140	294	153	3	0
Patient_130	56	Male	Non-anginal pain	130	256	142	6	1
Patient_140	44	Male	Atypical angina	120	263	173	7	0
Patient_150	52	Male	Non-anginal pain	172	199	162	7	0
Patient_160	57	Male	Non-anginal pain	150	168	174	3	0
Patient_170	48	Male	Atypical angina	110	229	168	7	1
Patient_180	54	Male	Asymptomatic	140	239	160	3	0
Patient_190	48	Female	Non-anginal pain	130	275	139	3	0
Patient_200	49	Male	Atypical angina	130	266	171	3	0
Patient_21	64	Male	Typical angina	110	211	144		0

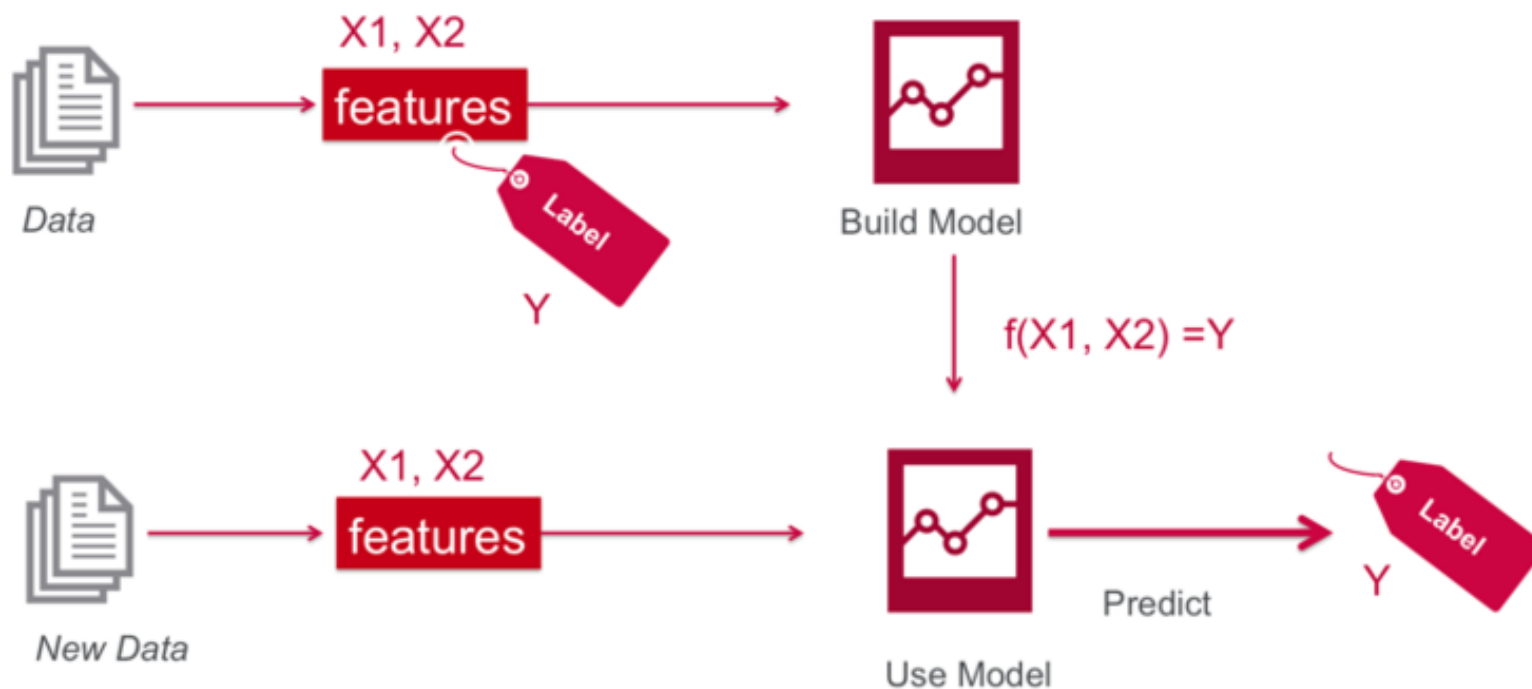
# Học có giám sát là gì?

Quá trình huấn luyện và kiểm thử với Mô hình học máy có giám sát



# Học có giám sát là gì?

- Bản chất của Supervised learning là học một hàm **f** phù hợp với tập dữ liệu hiện có và có khả năng tổng quát hóa cao.
- Hàm học được sau đó sẽ dùng để dự đoán cho các quan sát mới.





# Học có giám sát là gì?

- Ví dụ: Tập dữ liệu có gán nhãn

Biến phụ thuộc (Nhãn – y) chỉ là 1 trong 2 giá trị (Hữu hạn):

- **0: Không được cứu**
- **1: Được cứu**

Data_Titanic						
Survived	Pclass	Sex	Age	SibSp	Parch	Embarked
0	3	male	22.0	1	0	S
1	1	female	38.0	1	0	C
1	3	female	26.0	0	0	S
1	1	female	35.0	1	0	S
0	3	male	35.0	0	0	S
0	3	male		0	0	Q
0	1	male	54.0	0	0	S
0	3	male	2.0	3	1	S
1	3	female	27.0	0	2	S
1	2	female	14.0	1	0	C

# Học có giám sát là gì?

Biến phụ thuộc (Nhãn – y) chỉ là 1 trong 3 giá trị (Hữu hạn):

- **positive:** Comment có nội dung tích cực
- **neutral:** Comment có nội dung trung tính
- **negative:** Comment có nội dung tiêu cực

Data\_Comment

Comment	Sentiment
@VirginAmerica What @dhepburn said.	neutral
@VirginAmerica plus you've added commercials to the experience... tacky.	positive
@VirginAmerica I didn't today... Must mean I need to take another trip!	neutral
@VirginAmerica it's really aggressive to blast obnoxious "entertainment" in your guests' faces &an	negative
@VirginAmerica and it's a really big bad thing about it	negative
@VirginAmerica seriously would pay \$30 a flight for seats that didn't have this playing. it's really the only bad thing about flying VA	negative
@VirginAmerica yes, nearly every time I fly VX this “ear worm” won’t go away :)	positive
@VirginAmerica Really missed a prime opportunity for Men Without Hats parody, there. <a href="https://t.co">https://t.co</a>	neutral
@virginamerica Well, I didn't...but NOW I DO! :-D	positive
@VirginAmerica it was amazing, and arrived an hour early. You're too good to me.	positive
@VirginAmerica did you know that suicide is the second leading cause of death among teens 10-2	neutral
@VirginAmerica I &lt;3 pretty graphics. so much better than minimal iconography. :D	positive
@VirginAmerica This is such a great deal! Already thinking about my 2nd trip to @Australia &amp; I	positive
@VirginAmerica @virginmedia I'm flying your #fabulous #Seductive skies again! U take all the #stre	positive

Data\_Iris

sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa

Biến phụ thuộc (Nhãn – y) chỉ là 1 trong 3 giá trị (Hữu hạn):

- **Iris-setosa**
- **Iris-versicolor**
- **Iris-virginica**

# Học có giám sát là gì?

Biến phụ thuộc (Nhãn - y) có rất nhiều giá trị khác nhau (vô hạn):

Data\_Housing

area	bedrooms	bathrooms	mainroad	guestroom	basement	parking	furnishingstatus	price
7420	4	2	yes	no	no	2	furnished	13300000
8960	4	4	yes	no	no	3	furnished	12250000
9960	3	2	yes	no	yes	2	semi-furnished	12250000
7500	4	2	yes	no	yes	3	furnished	12215000
7420	4	1	yes	yes	yes	2	furnished	11410000
7500	3	3	yes	no	yes	2	semi-furnished	10850000
8580	4	3	yes	no	no	2	semi-furnished	10150000
16200	5	3	yes	no	no	0	unfurnished	10150000
8100	4	1	yes	yes	yes	2	furnished	9870000
5750	3	2	yes	yes	no	1	unfurnished	9800000
13200	3	1	yes	no	yes	2	furnished	9800000
6000	4	3	yes	yes	yes	2	semi-furnished	9681000
6550	4	2	yes	no	no	1	semi-furnished	9310000

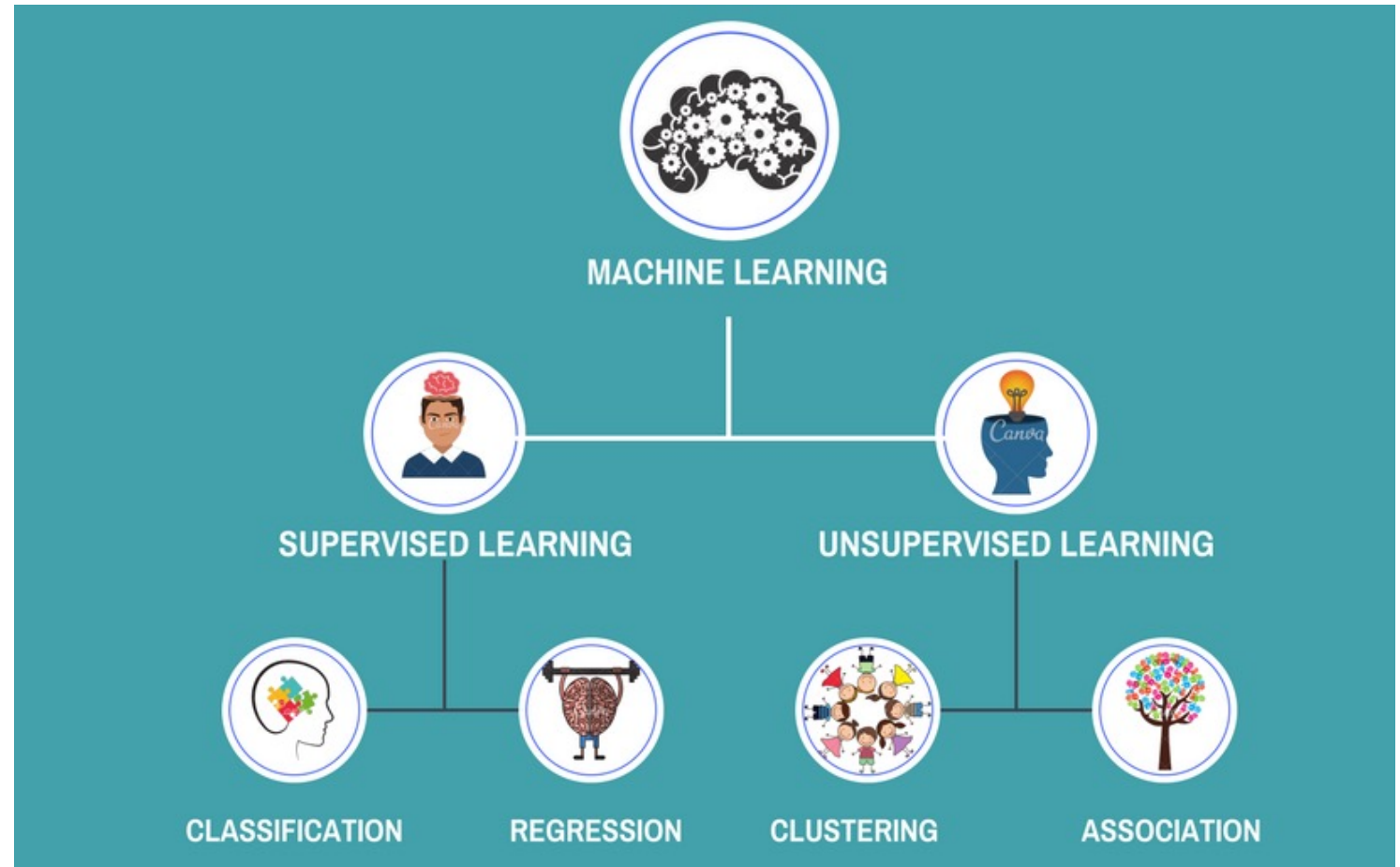
Data\_Diamond

Weight(Carat)	Price(USD)
0.23	484
0.31	942
0.2	345
1.02	4459
1.63	14022
1.14	4212
2.01	11925

## 2. Phân loại học có giám sát

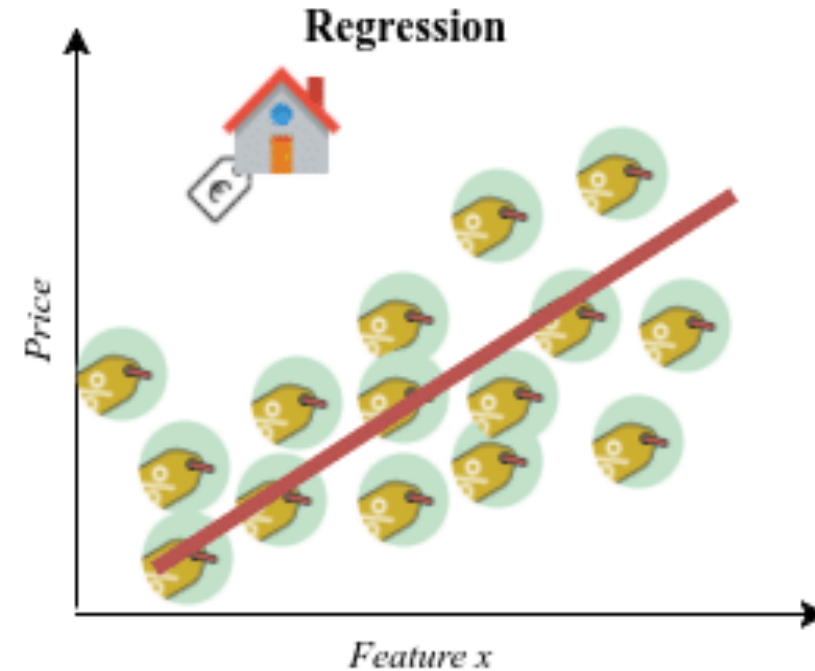
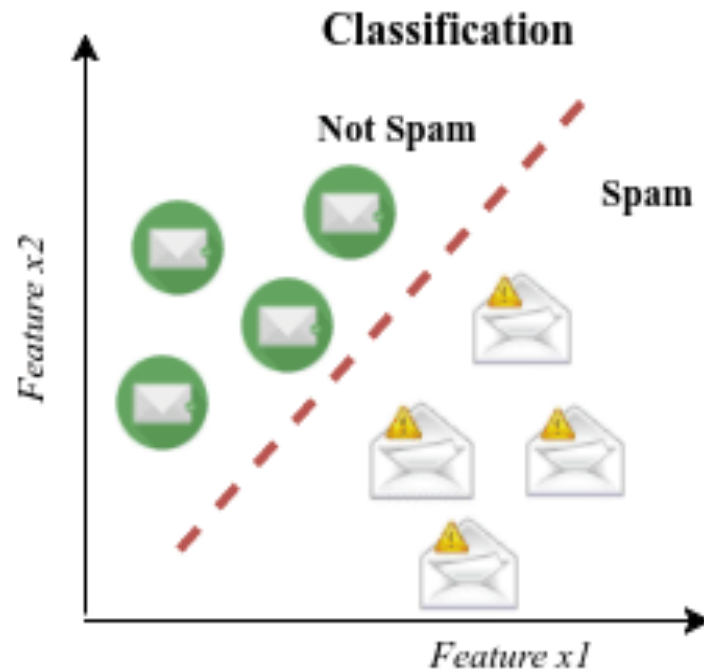
# Các loại học có giám sát

- Học có giám sát bao gồm 2 loại:
  - Phân loại (Classification)
  - Hồi quy (Regression)



# Các loại học có giám sát

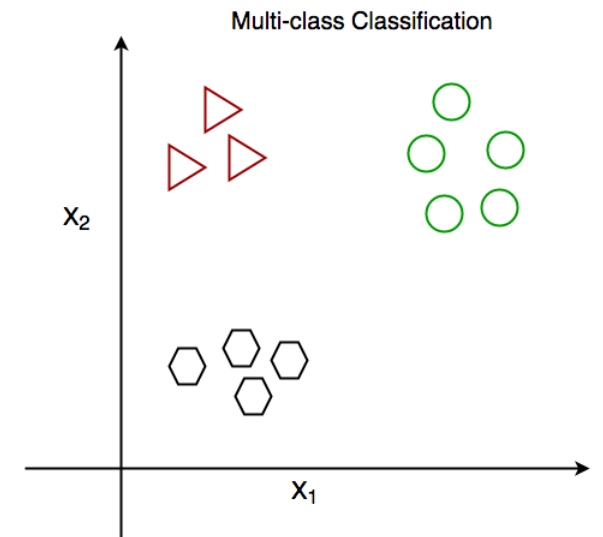
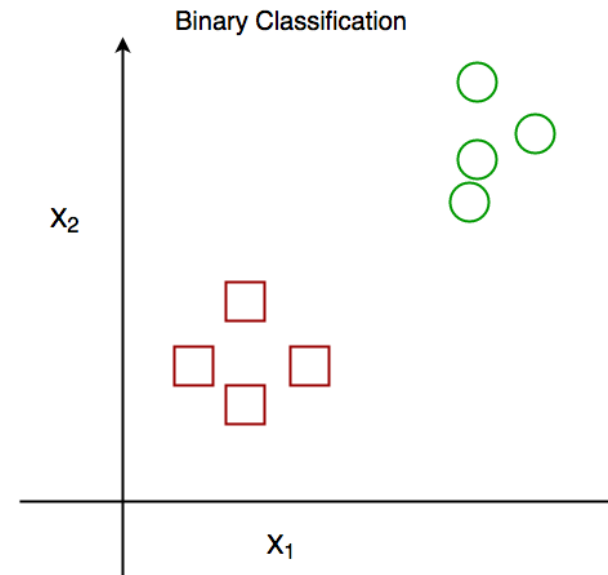
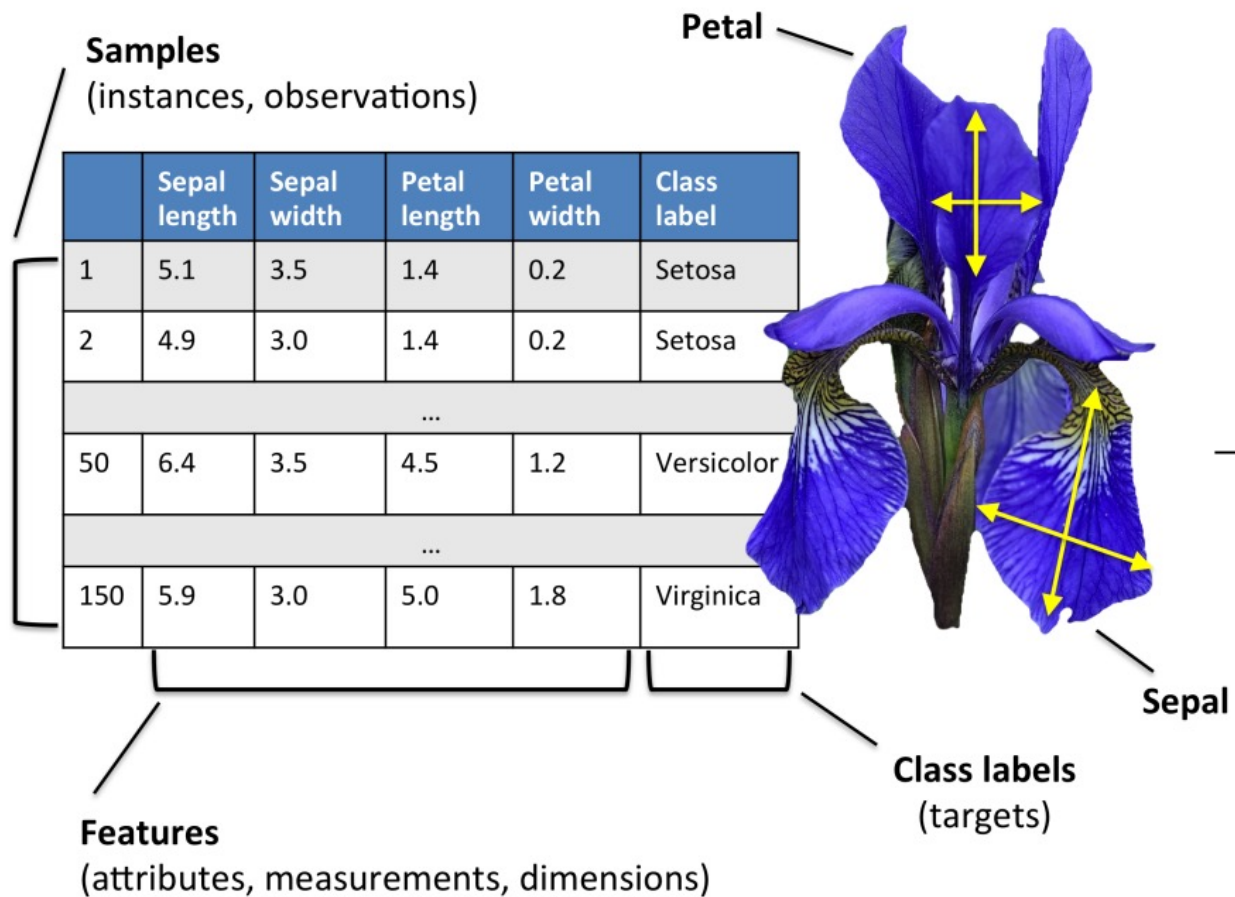
## Phân loại (Classification) & Hồi quy (Regression)



Hai nhóm bài toán cơ bản trong học có giám sát là classification (phân loại) và regression (hồi quy), trong đó biến đầu ra của bài toán phân loại có các giá trị rời rạc trong khi biến đầu ra của bài toán hồi quy có các giá trị liên tục.

# Phân loại (Classification)

**Phân loại (Classification):** Nếu nhãn ( $y$  – Target) thuộc tập rời rạc và hữu hạn



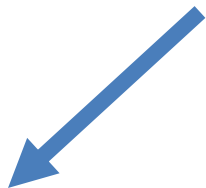


# Hồi quy (Regression)

Hồi quy (Regression): Nếu nhãn (y – Target) là biến liên tục (các số thực) ví dụ như dự báo nhiệt độ, giá nhà, mức tiêu thụ điện năng...



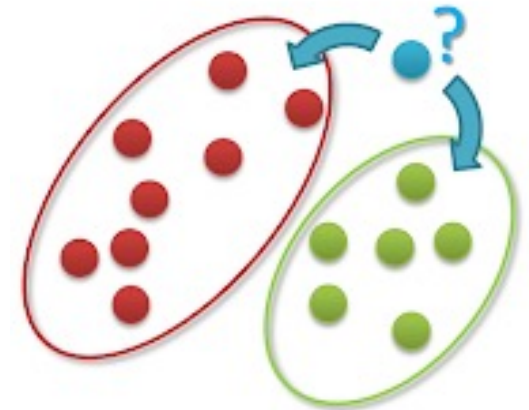
area	bedrooms	bathrooms	mainroad	guestroom	basement	parking	furnishingstatus	price
7420	4	2	yes	no	no	2	furnished	13300000
8960	4	4	yes	no	no	3	furnished	12250000
9960	3	2	yes	no	yes	2	semi-furnished	12250000
7500	4	2	yes	no	yes	3	furnished	12215000
7420	4	1	yes	yes	yes	2	furnished	11410000
7500	3	3	yes	no	yes	2	semi-furnished	10850000
8580	4	3	yes	no	no	2	semi-furnished	10150000
16200	5	3	yes	no	no	0	unfurnished	10150000
8100	4	1	yes	yes	yes	2	furnished	9870000
5750	3	2	yes	yes	no	1	unfurnished	9800000
13200	3	1	yes	no	yes	2	furnished	9800000
6000	4	3	yes	yes	yes	2	semi-furnished	9681000
6550	4	2	yes	no	no	1	semi-furnished	9310000





### 3. Một số thuật toán phân lớp (Classification)

Classification



# Top 10 Algorithms every Machine Learning Engineer should know

1. Naïve Bayes Classifier Algorithm

2. K Means Clustering Algorithm

3. Support Vector Machine Algorithm

4. Apriori Algorithm

5. Linear Regression Algorithm

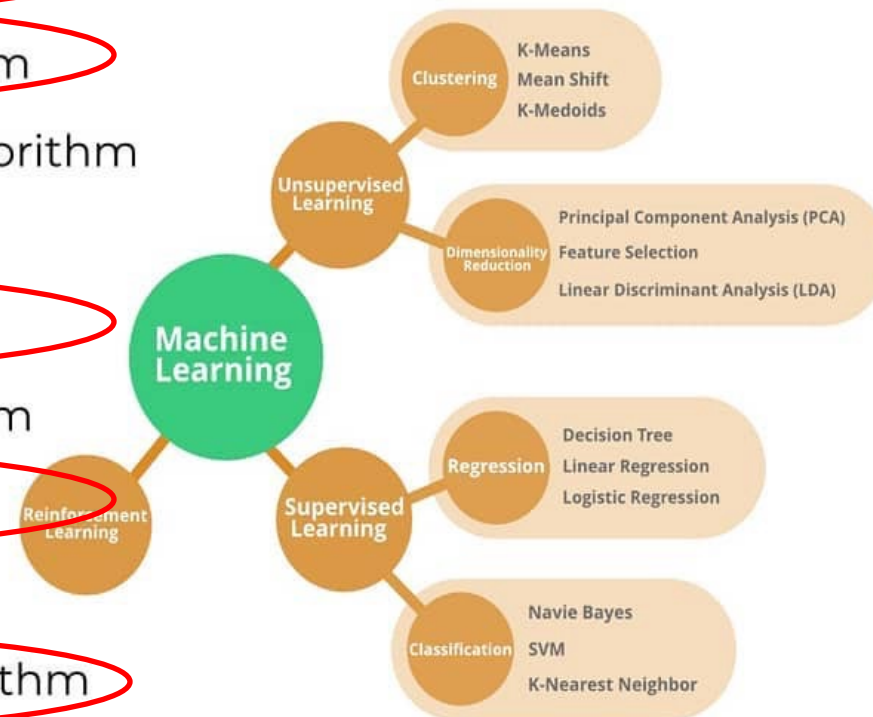
6. Logistic Regression Algorithm

7. Decision Trees Algorithm

8. Random Forests Algorithm

9. K Nearest Neighbours Algorithm

10. Artificial Neural Networks Algorithm





# Thank you!