Analysis $\stackrel{?}{=}$ Analytics

Analysis

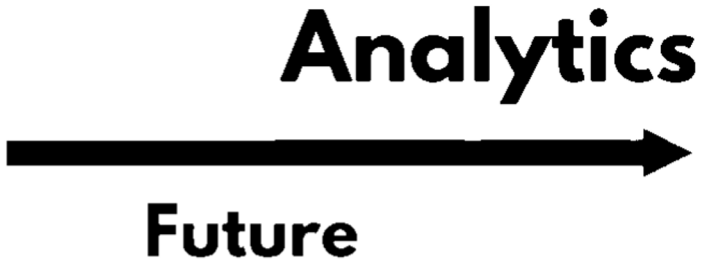
Past

Explain

How? Why?

STATS/BI

We can use different tools to explain the previous trends like, Power BI, Tableau, QlikView, MicroStrategy etc.



Explore potential future events

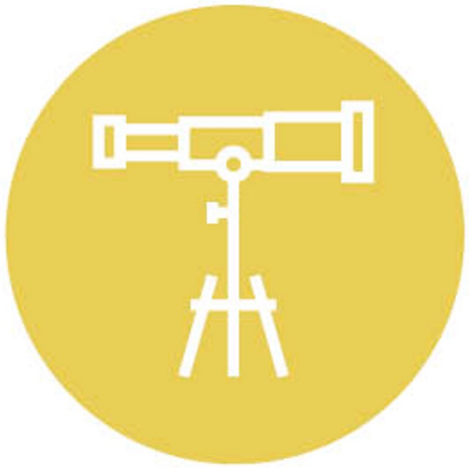
ML/AI

We can use different language packages and framework to implement ML/AI model.

Business Analytics

Descriptive
What has happened?

Diagnostic
Why did it happen?



Predictive
What will happen next?

Prescriptive
What should I do?

Looking back **Looking forward**

Case Study

A credit card company wants to reduce the number of customers defaulting on their payments. They gather historical data on customer transactions, payment history, credit scores, and demographic information. By analyzing this data, they aim to create a model that can forecast the likelihood of a customer defaulting on their next payment. The focus is on developing a predictive model that can identify early warning signs of potential defaults based on patterns and trends observed in the historical data. This analysis will enable the company to proactively intervene and offer assistance to customers at risk of default, thereby minimizing financial losses and maintaining a healthier customer base.

Case Study

A manufacturing plant experiences a sudden and significant drop in its production output. The operations team gathers data on machine performance, maintenance logs, and production schedules. By analyzing this data, they aim to pinpoint the exact factors that led to the production decline. The focus is on identifying any equipment malfunctions, breakdowns, or operational bottlenecks that might have contributed to the drop in output. This analysis will help the team determine the root causes of the issue and develop strategies to address the problem promptly, ensuring the plant returns to its optimal production levels.

Case Study

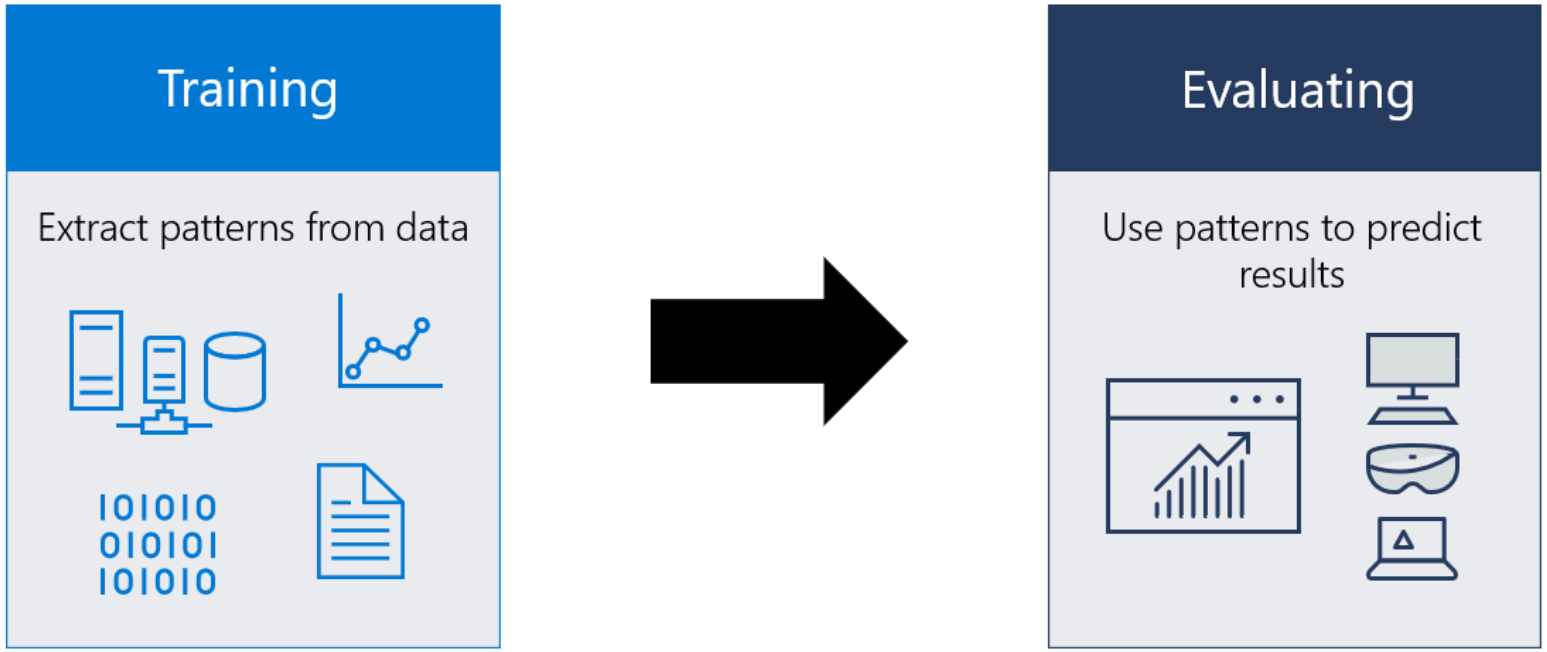
An e-commerce company is analyzing its sales data from the past year. The company's data team compiles information on product purchases, order dates, customer locations, and purchase amounts. By examining this data, the company aims to uncover trends and patterns in its sales performance. The focus is on understanding which products sold well during specific periods, identifying peak buying times, and discerning whether there are any geographical preferences among customers. This analysis will guide the company in making informed decisions about inventory management, marketing strategies, and potential expansions into new markets.

Case Study

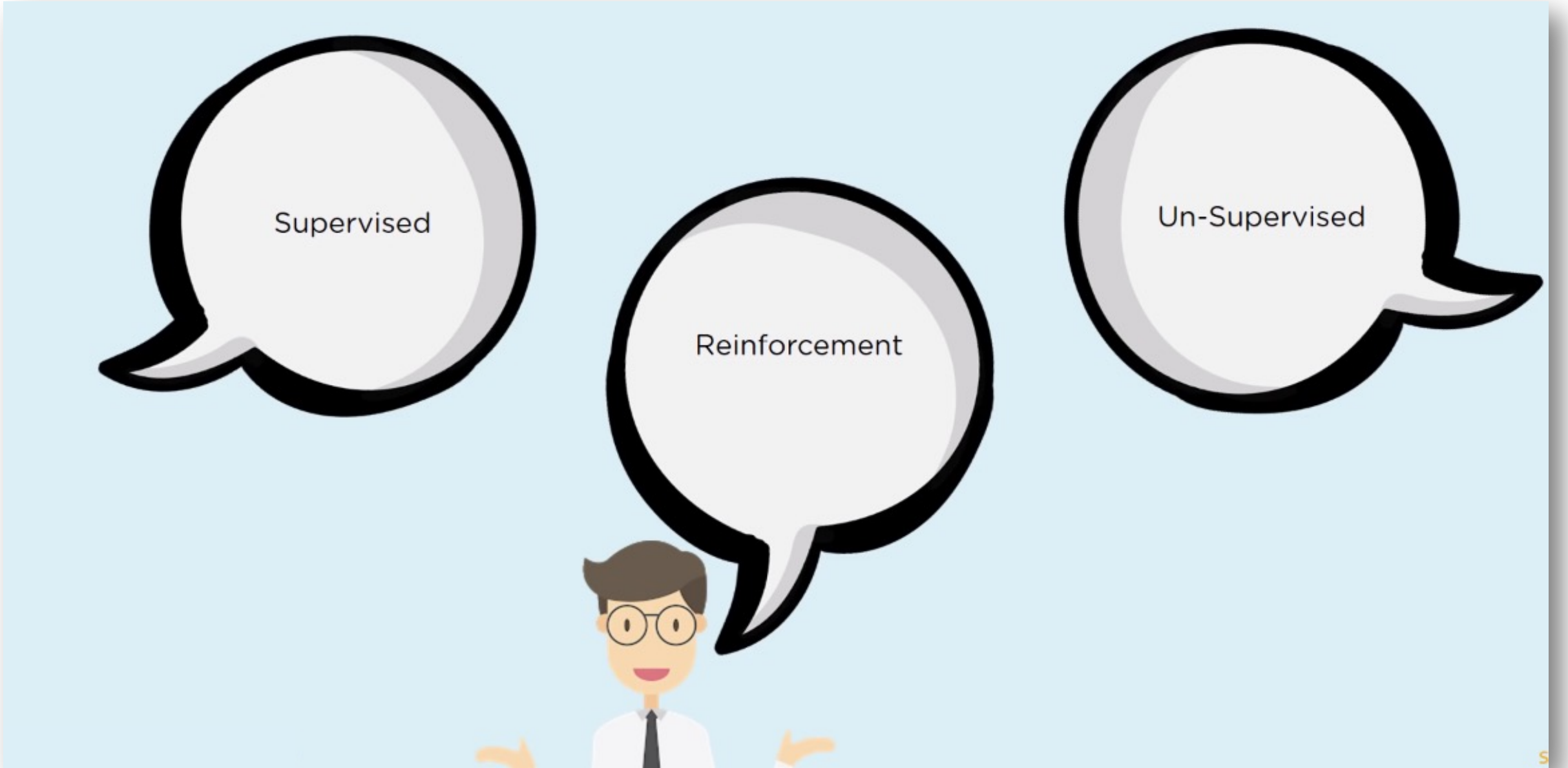
A healthcare provider notices an increase in patient readmissions for a specific chronic condition. They gather data on patient medical histories, treatment plans, medications prescribed, and post-discharge follow-up procedures. By analyzing this data, they aim to develop a system that recommends personalized treatment plans for patients with the identified chronic condition. The focus is on creating a solution that can predict potential complications based on historical patient data and suggest optimal interventions to prevent readmissions. This analysis will guide healthcare professionals in making more informed decisions about patient care, ultimately reducing readmission rates and improving overall patient outcomes.

Machine Learning

Machine (computer) tries to find the pattern (self-learn) from the data.



Types of Machine Learning



Churn Prediction

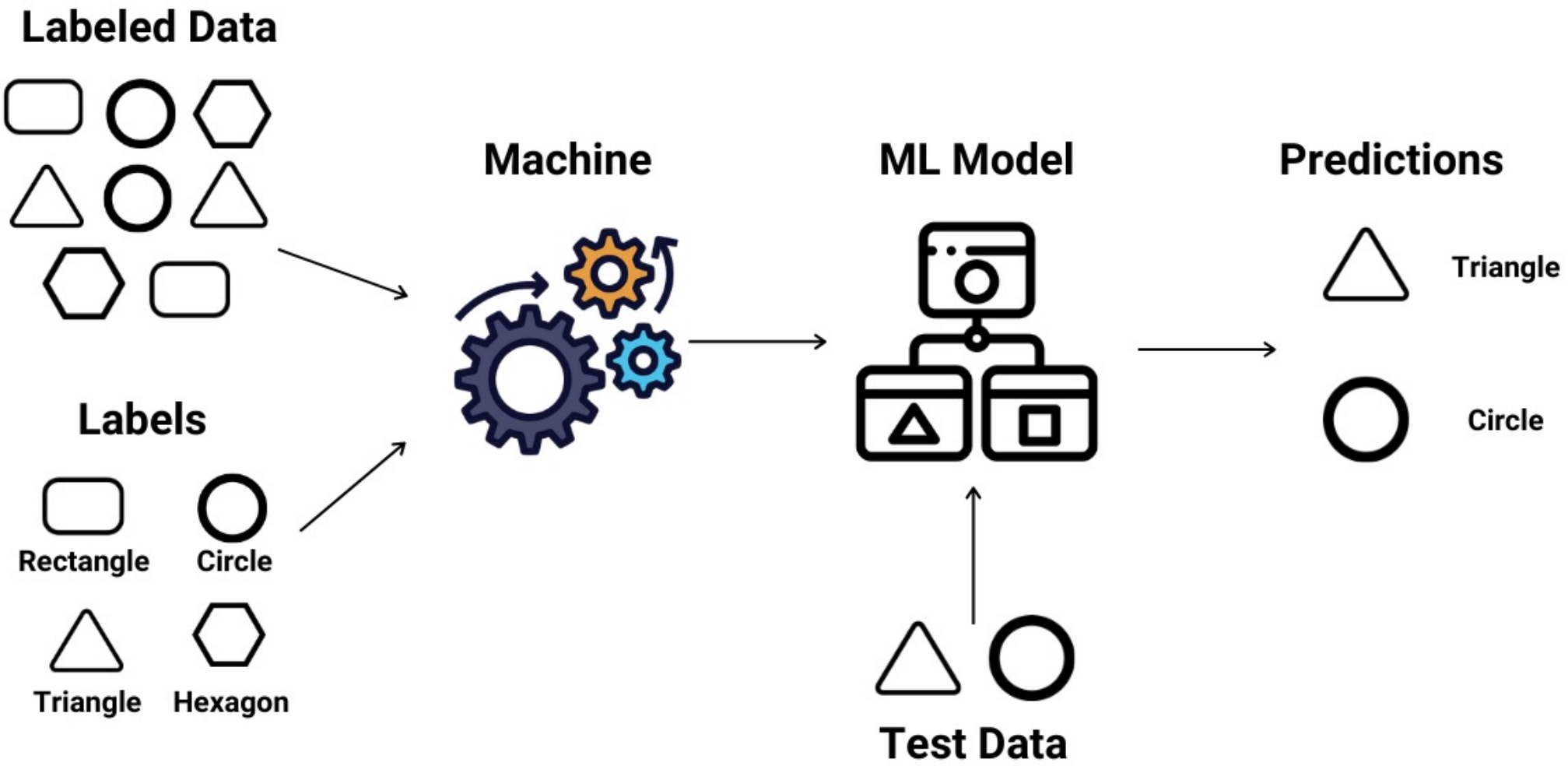
Customer ID	Age	Monthly Charges	Contract Length	Data Usage	Customer Service Calls	Churn (Target)
1	30	50	12	150 GB	2	No
2	45	75	24	300 GB	1	No
3	22	35	6	50 GB	3	Yes
4	55	60	12	200 GB	2	No
5	40	85	24	250 GB	1	Yes
...	...	...	...	...	...	...

Predicting customer churn can help companies proactively address customer issues, improve service quality, and implement targeted retention strategies. Supervised machine learning algorithms can be employed to build predictive models that identify customers at risk of churning.

- Set of inputs ~ [Features] / [Independent Variables] / [X]

- Outputs ~ [Labels] / [Dependent Variables] / [Y]

Supervised Learning



Supervised Learning

Classification

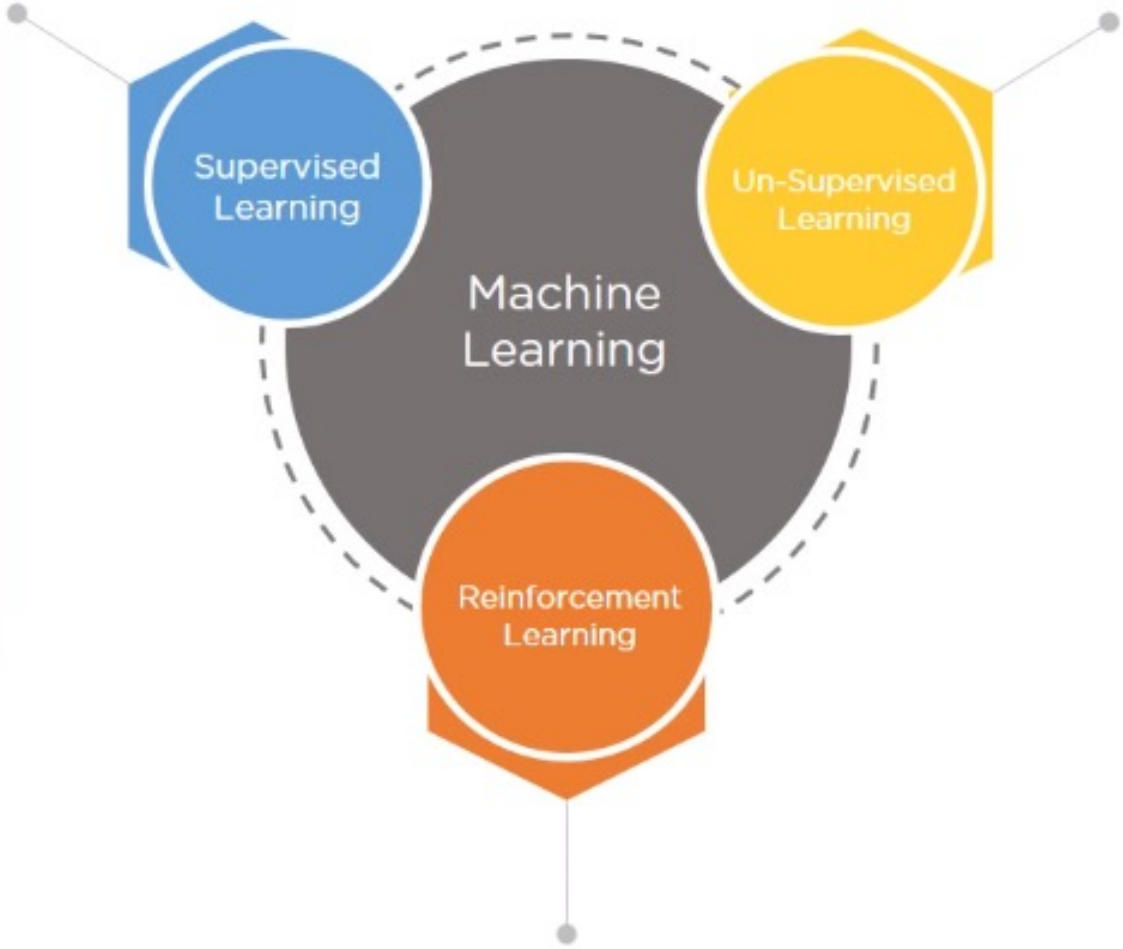
- Fraud Detection
- Email Spam Detection
- Image Classification

Categorical

Regression

- Weather Forecasting
- Risk Assessment
- Score Prediction

Numerical



Supervised Learning

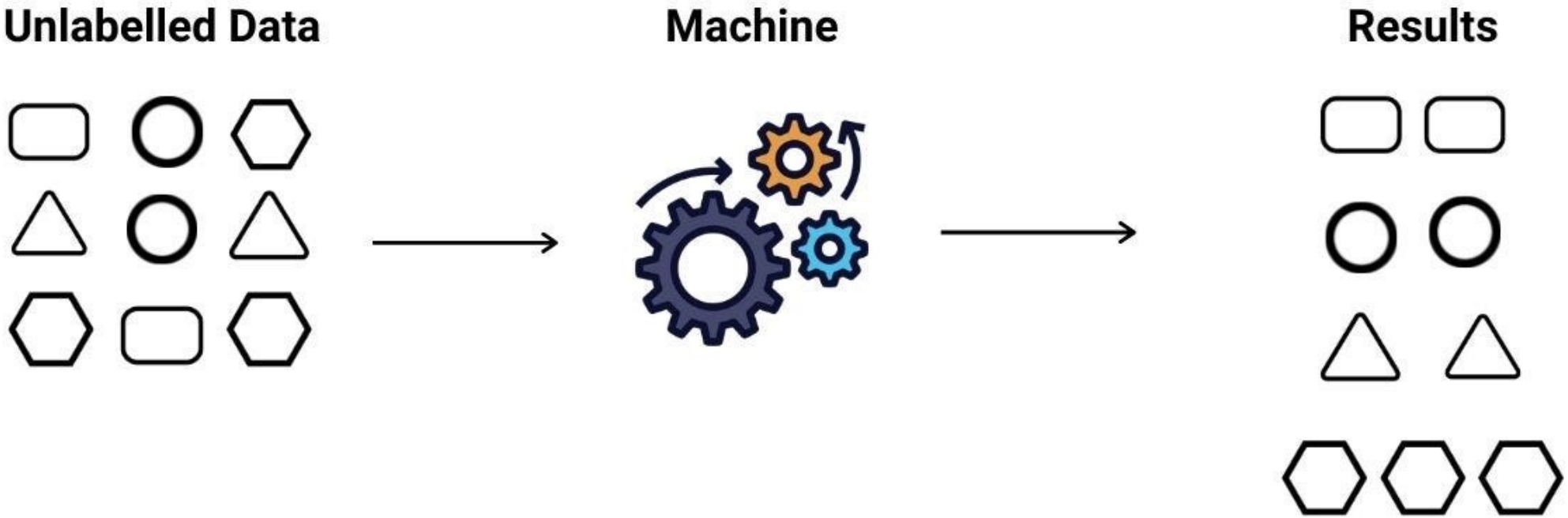
User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Figure A: CLASSIFICATION

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Figure B: REGRESSION

Unsupervised Learning



Customer Segmentation

Customer	Call Minutes	Data Usage	Text Messages	Tenure	Service Interactions	Avg. Monthly Bill
1	250	2.5	100	12	2	$50
2	100	1.0	20	24	1	$30
3	800	5.0	300	6	3	$70
...	...	...	...	...	...	...

Unsupervised machine learning techniques can be applied to segment customers based on their usage patterns, preferences, and behavior. One common approach is using clustering algorithms to group customers with similar characteristics together.

Cluster 1 (High Usage):

- Customers with high call minutes, data usage, and text messages
- Relatively short tenure
- Moderate to high service interactions
- Higher average monthly bill

Cluster 2 (Medium Usage):

- Customers with moderate call minutes, data usage, and text messages
- Medium tenure
- Moderate service interactions
- Moderate average monthly bill

Cluster 3 (Low Usage):

- Customers with low call minutes, data usage, and text messages
- Longer tenure
- Low service interactions
- Lower average monthly bill

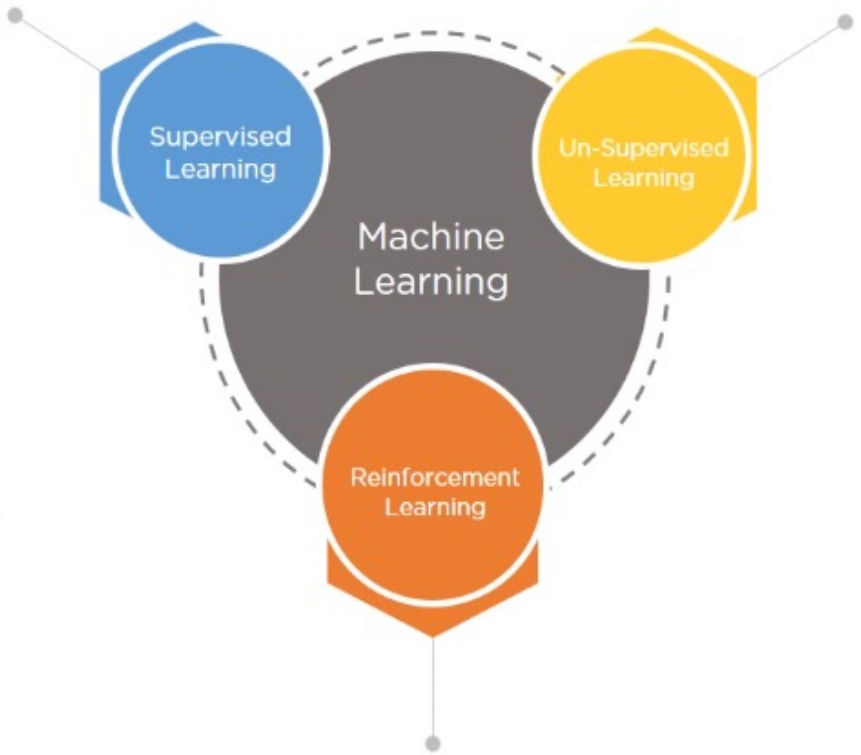
Unsupervised Learning

Classification

- Fraud Detection
- Email Spam Detection
- Image Classification

Regression

- Weather Forecasting
- Risk Assessment
- Score Prediction



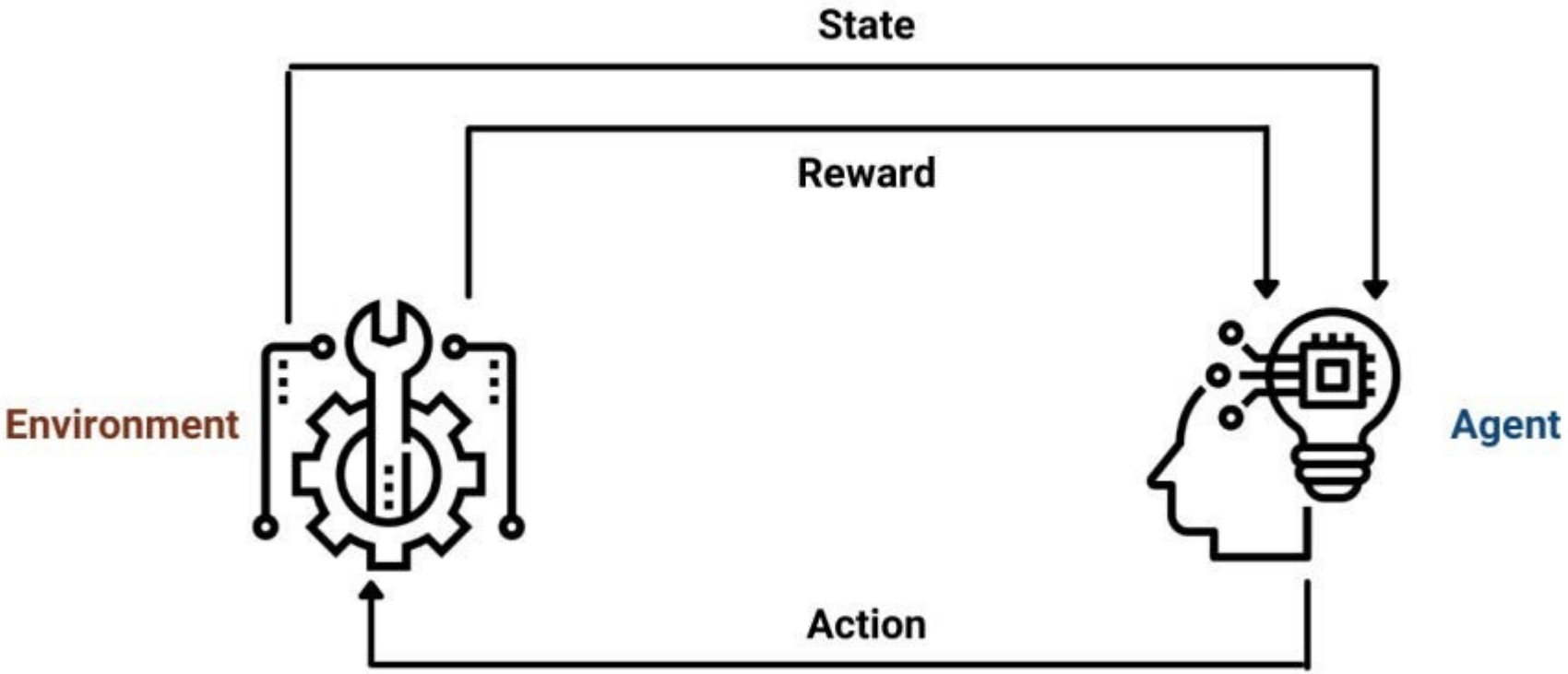
Association

- Market Basket Analysis
- Text Mining
- Face Recognition

Clustering

- Medical Research
- City Planning
- Targeted Marketing

Reinforcement Learning



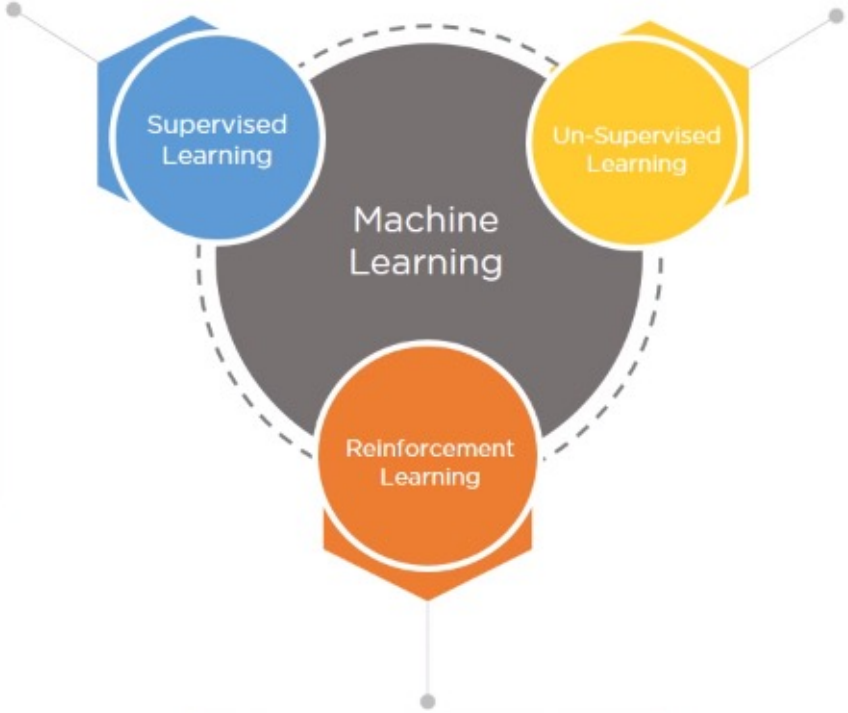
Reinforcement Learning

Classification

- Fraud Detection
- Email Spam Detection
- Image Classification

Regression

- Weather Forecasting
- Risk Assessment
- Score Prediction



Association

- Market Basket Analysis
- Text Mining
- Face Recognition

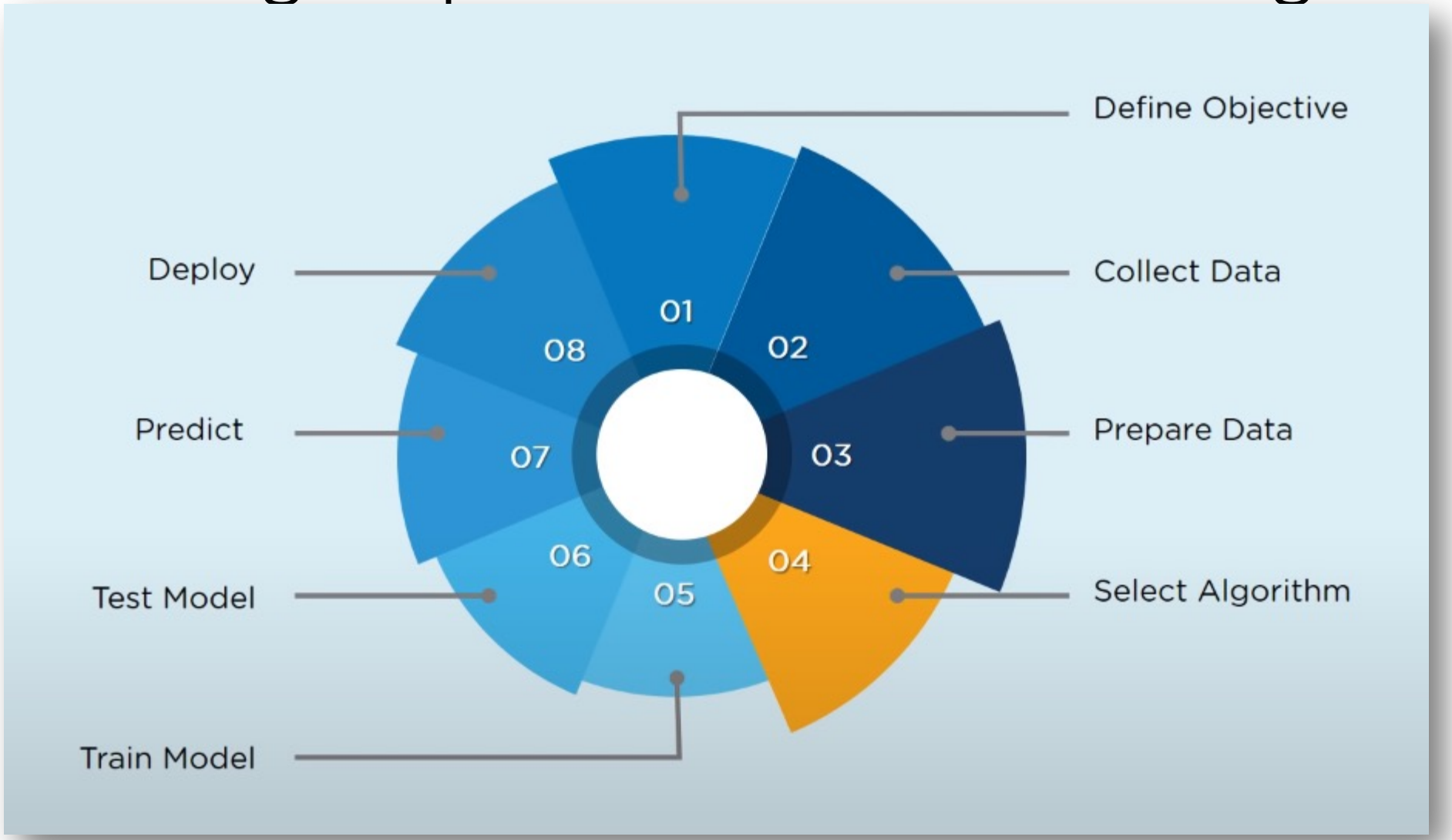
Clustering

- Medical Research
- City Planning
- Targeted Marketing

Reinforcement Learning

- Gaming
- Robot Navigation
- Stock Trading
- Assembly Line Processes

Processing Steps for Machine Learning

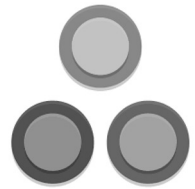


Major ML supported Languages

Python / R / Java / Scala / Spark / Julia / **No Code**

These language provide all necessary ML packages





Hands-on

Supervised Machine Learning

Desktop Based Data Mining : **Orange**

Performance Metrics - Regression

- **BAD MODEL**

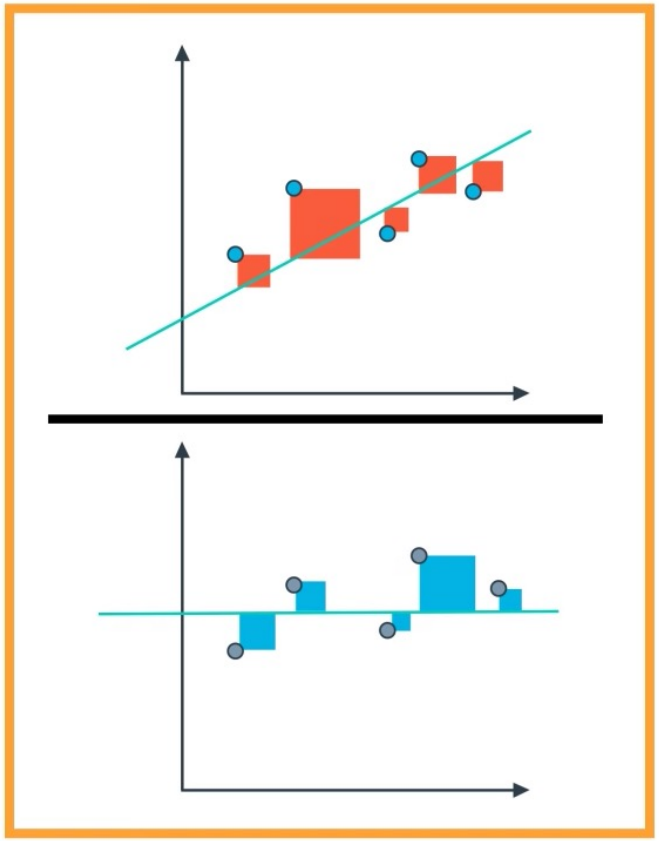
  The errors should be similar.
  R2 score should be close to 0.

- **GOOD MODEL**

  The mean squared error for the linear regression model should be a lot smaller than the mean squared error for the simple model.

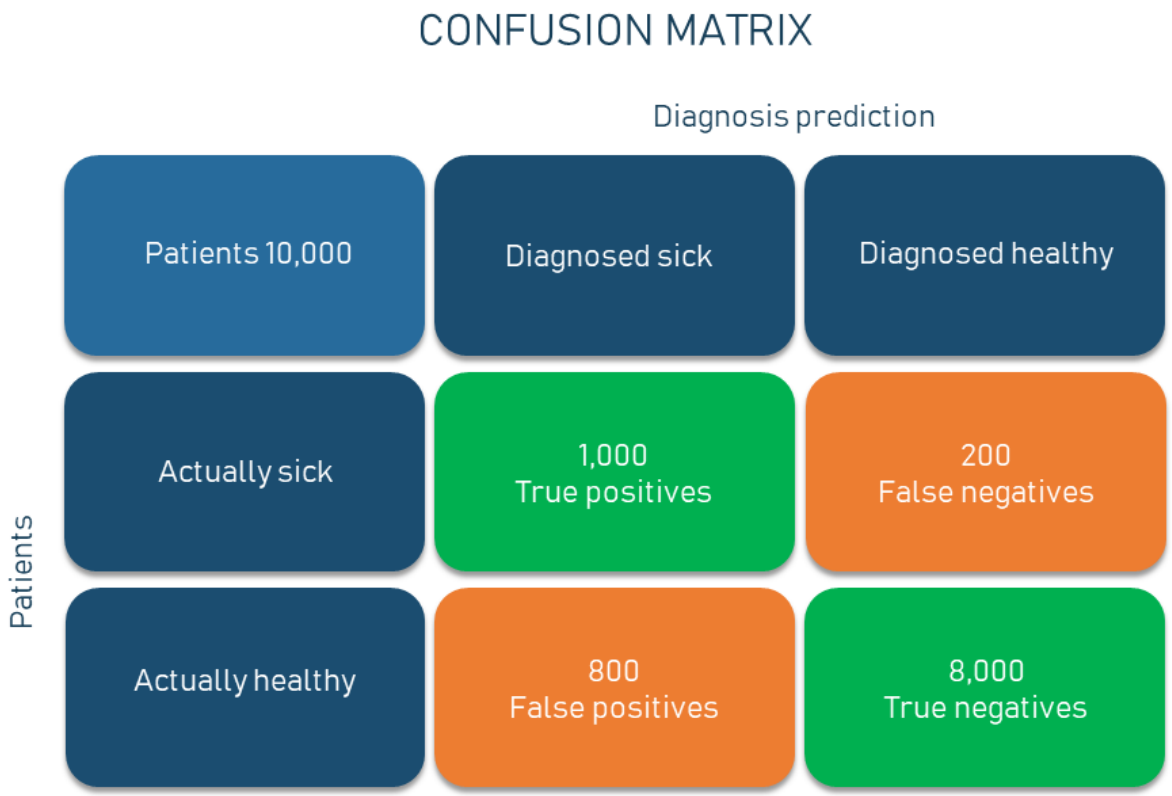
  R2 score should be close to 1.

$$R2 = 1 - \frac{\text{MSE(model)}}{\text{MSE(baseline)}}$$

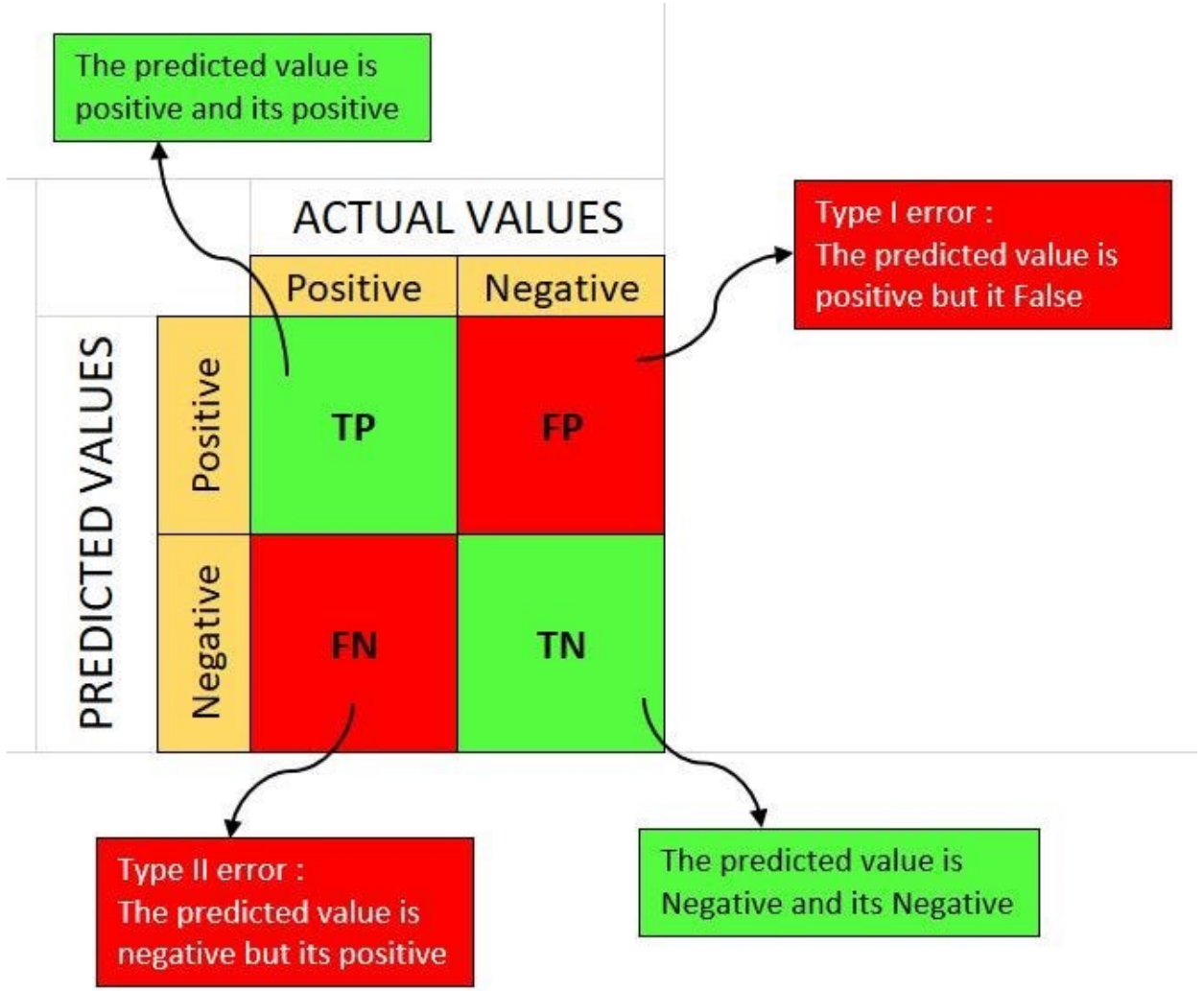


R2 (R-Square)score can be interpreted as a coefficient of determination where value propagate from 0 to 1, where 1 is the best.

Performance Metrics - Classification



Performance Metrics - Classification



Performance Metrics - Classification

		Actual class (ground truth)		
Total (n)=100		Dog (Positive)	Not a Dog (Negative)	
Predicted class	Dog (Positive)	15 (TP)	20 (FP, Type I Error)	**Precision** =TP/(TP+FP) =0.42
	Not a Dog (Negative)	5 (FN, Type II Error)	60 (TN)	
	Accuracy =(TP+TN)/Total =0.75	**Sensitivity,Recall, TPR** =TP/(TP+FN) = 0.75	**FPR** = FP/(FP+TN) =0.25	**F1 Score** =2*(Precision*Recall) / (Precision+Recall) =0.53
	Error Rate =(FP+FN)/Total =0.25	**Miss Rate, FNR** =FN/(TP+FN) =0.25	**Specificity, TNR** = TN/(FP+TN) =0.75	

Performance Metrics - Classification

Accuracy: This is like counting how many balls you put in the right baskets out of all the balls. It tells us how good the computer is at getting things right. If it's good, the accuracy number will be high, like if you got most of the balls in the right baskets.

Precision: Imagine you're very careful and only put a ball in a basket when you're really sure it's the right one. Precision is like that – it measures how many of the balls you put in the basket were actually supposed to go there. So, if you only put red balls in the red basket and some of them are there, your precision is good.

Recall: Sometimes, you might miss putting some balls in the right baskets. Recall helps us know if you missed any. It's like counting how many balls you managed to put in the right baskets out of all the balls that should've gone in those baskets. If you missed very few, your recall is high.

F1 Score: This one is like a combination of precision and recall. It helps us know how balanced you are – did you put the right balls in the right baskets and not miss too many? F1 score gives you a number that shows how well you balanced both of those things.

Performance Metrics - Classification

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1 score can be interpreted as a measure of overall model performance from 0 to 1, where 1 is the best.