Machine Learning with AzureML

Intro to Data Science and ML

Presenter

Mohammed Arif, PhD
Lead Data Scientist
Big Data | Machine Learning | AI

Mohammed Arif has more than seventeen (17) years of working experience in Information Communication and Technology (ICT) industry. The highlights of his career are more than eight (8) years of holding various senior management and/or C-Level and had six (6) years of international ICT consultancy exposure in various countries (APAC and Australia), specially on Big Data, Data Engineering, Machine Learning and AI arena.
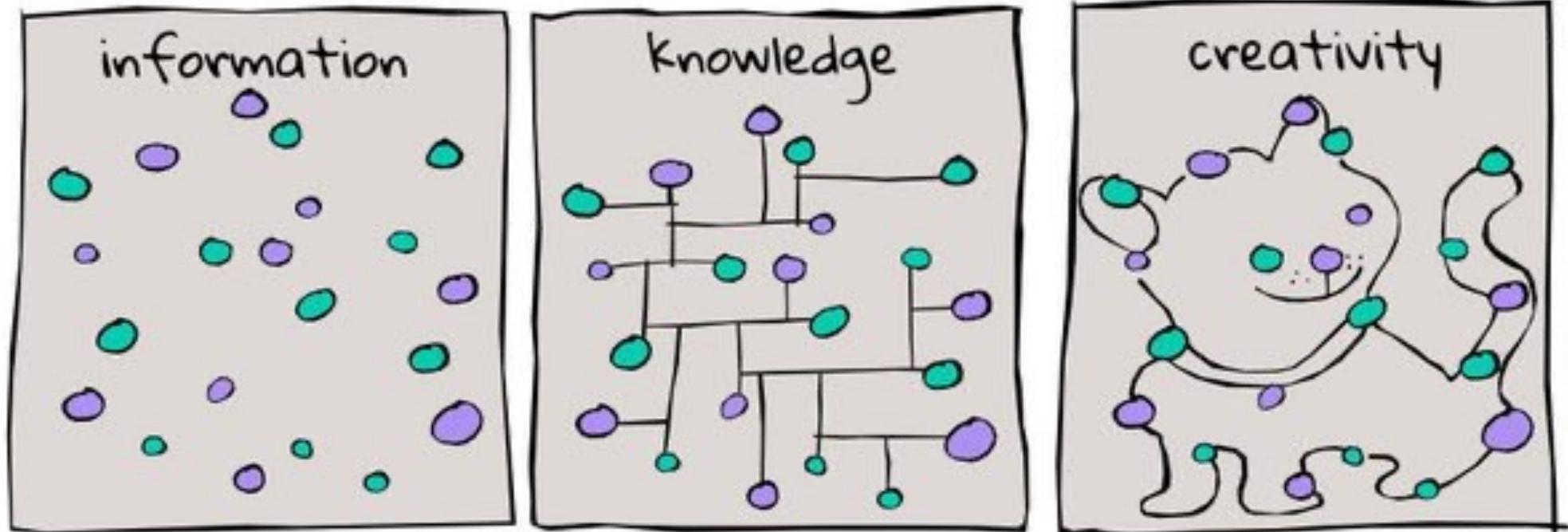
He is also Certified Trainer for Microsoft.

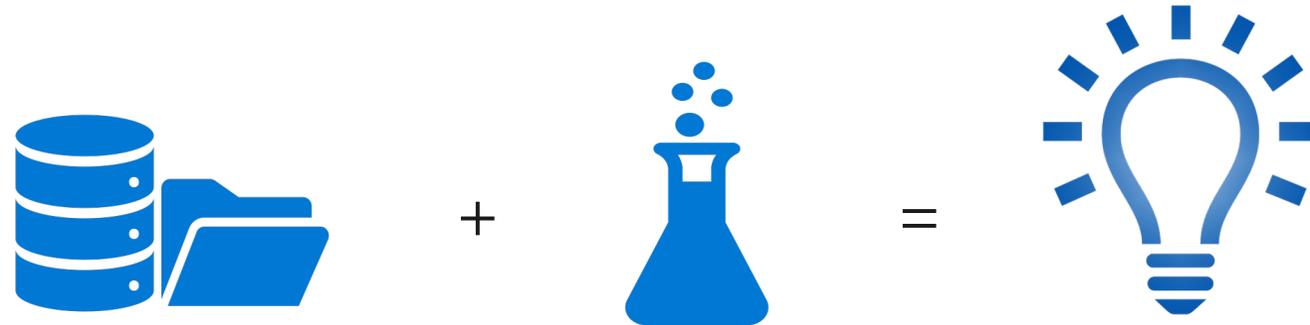**in** **/arifmazumder**

Agenda

- What is Data Science
- Data
- Components of Data Science
- Machine Learning
- Types of Machine Learning
- Steps to do Machine Learning
- Hands-on Problem Solving

Resources Links : http://arif.works/azureml

What is Data Science?

Apply Scientific Methods to extract Knowledge from Data.
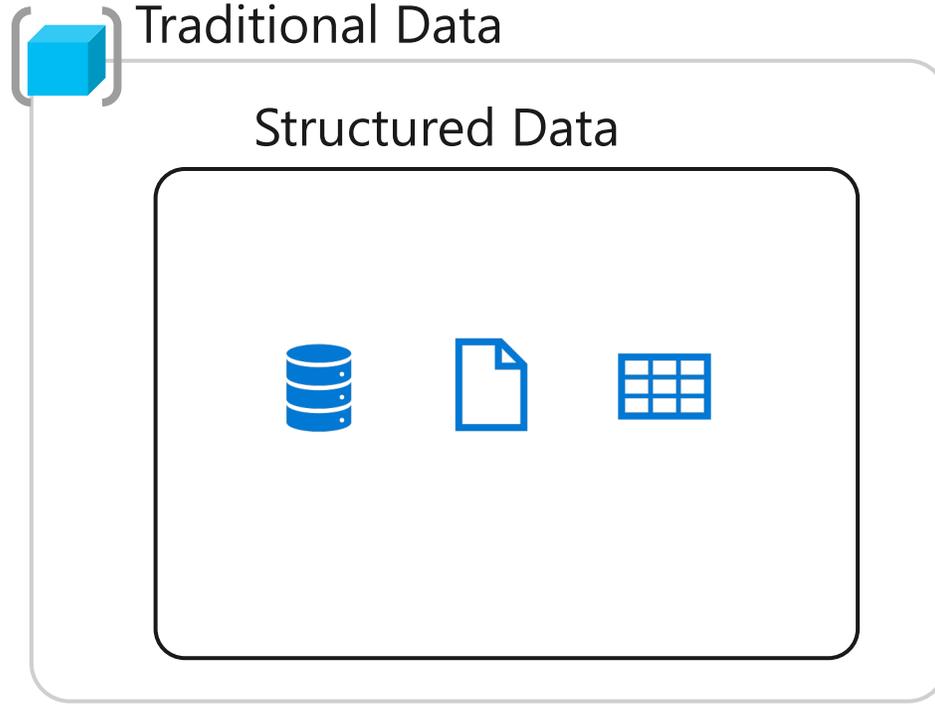
Data

Big Data

Both Structured and Unstructured Data

Traditional Data

Structured Data

Volume

Variety

Velocity

Scientific Methods

Statistics

Designed for inference about the relationships between variables

Machine Learning

Designed to make the most accurate predictions possible

Artificial Intelligence

Designed to mimic human behavior using ML and Deep Learning

Machine Learning

Machine (computer) tries to find the pattern (self-learn) from the data without being explicitly programmed.

When we need to apply
Machine Learning

Analysis ⁝= **Analytics**

When we need to apply Machine Learning

Analysis

←

Past

Explain

How?   Why?

BI

We can use different tools to explain the previous trends like, PowerBI, Tableau, Qlikview etc.
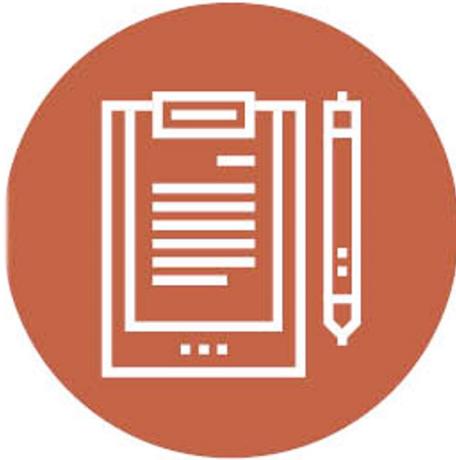
When we need to apply Machine Learning

Analytics

→

Future

Explore potential future events

ML/AI

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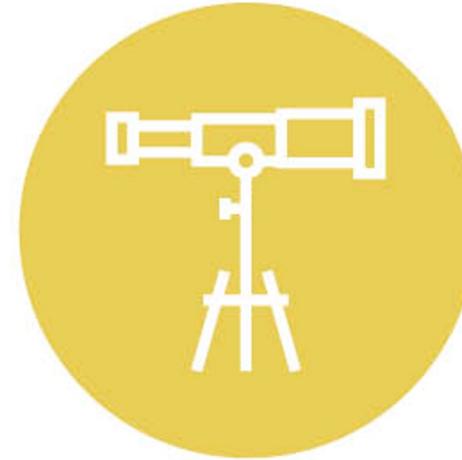
Business Analytics

Descriptive
What has happened?
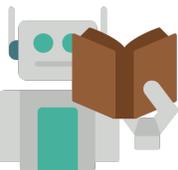
Diagnostic
Why did it happen?

Predictive
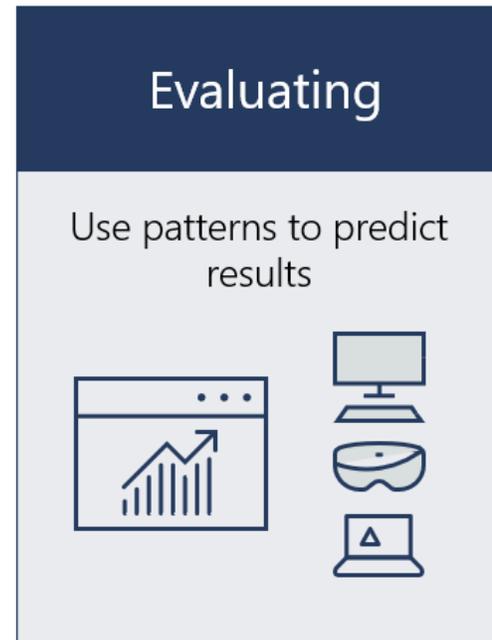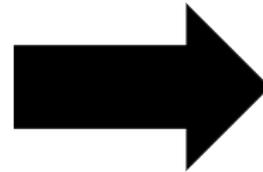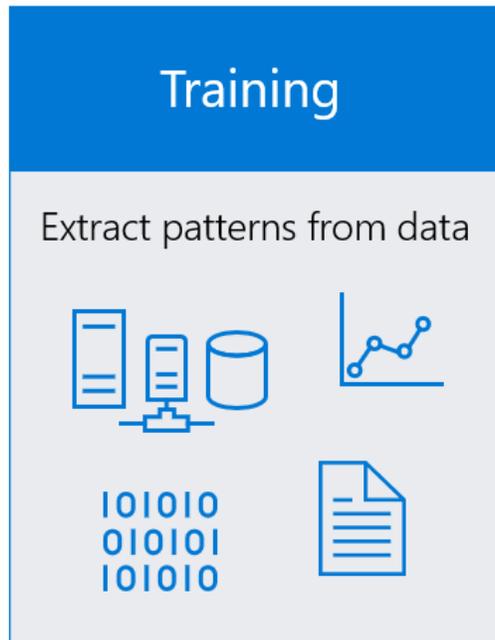What will happen next?

Prescriptive
What should I do?

Looking back | **Looking forward**

Machine Learning

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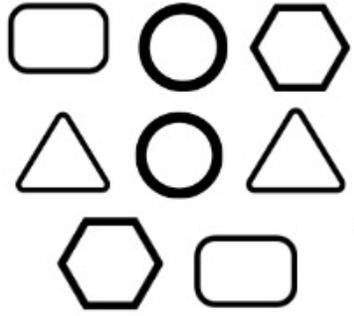
Types of Machine Learning

Supervised Learning
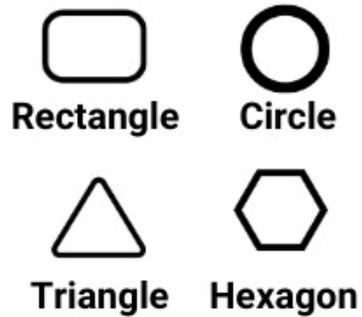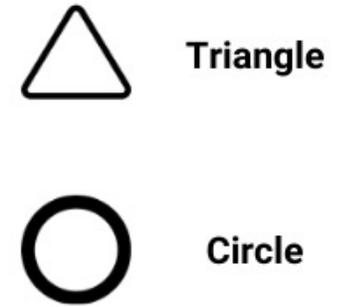
Churn Prediction

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

Classification

- Fraud Detection

- Email Spam Detection

- Image Classification

Categorical

Regression

- Weather Forecasting

- Risk Assessment

- Score Prediction

Numerical

Supervised Learning

Un-Supervised Learning
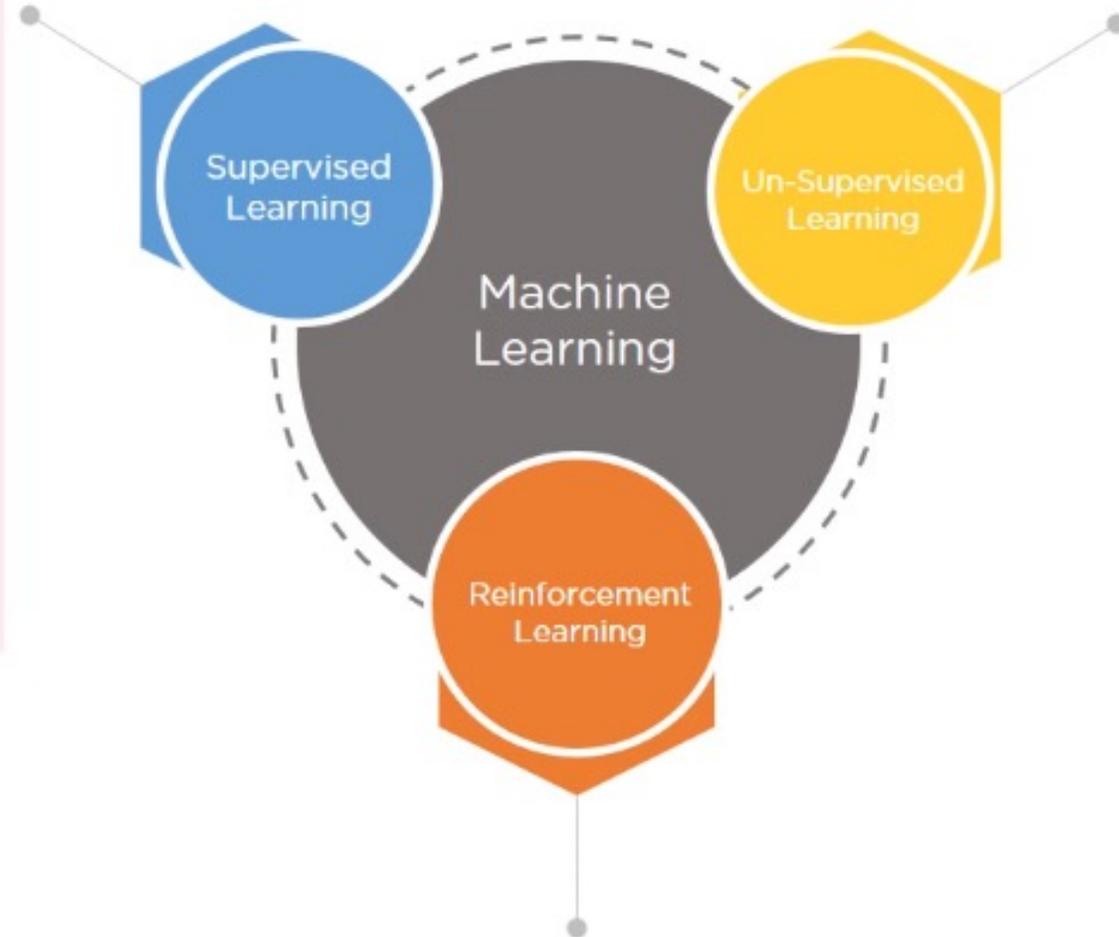
Machine Learning

Reinforcement Learning

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
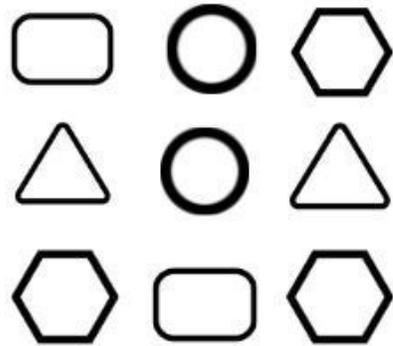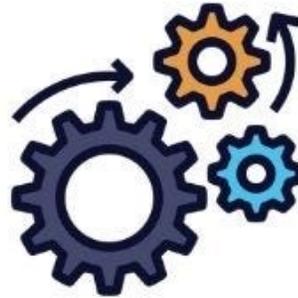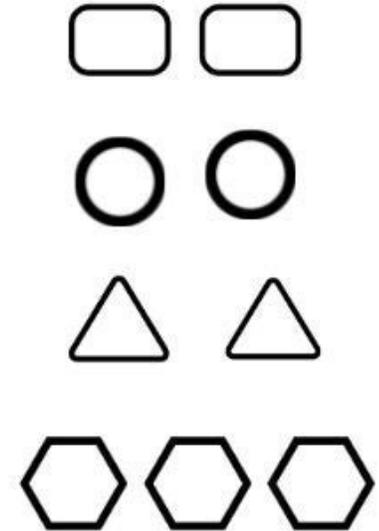
Customer Segmentation

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
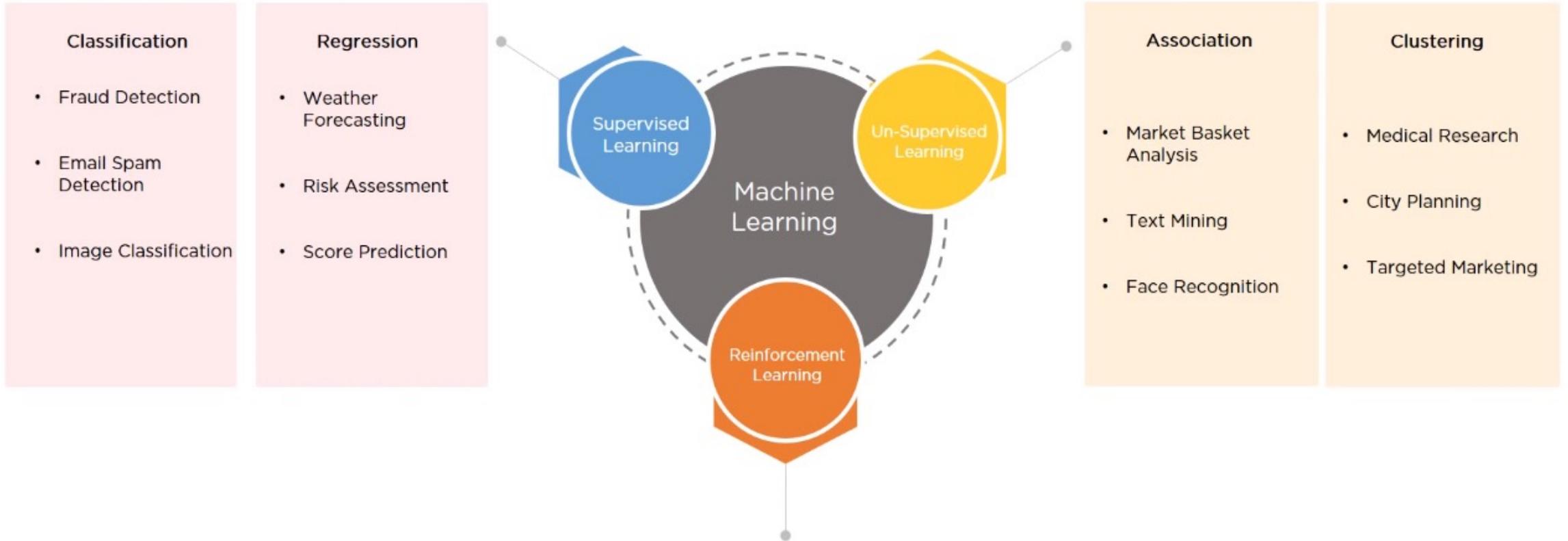- ❑ Higher average monthly bill

Cluster 2 (Medium Usage):

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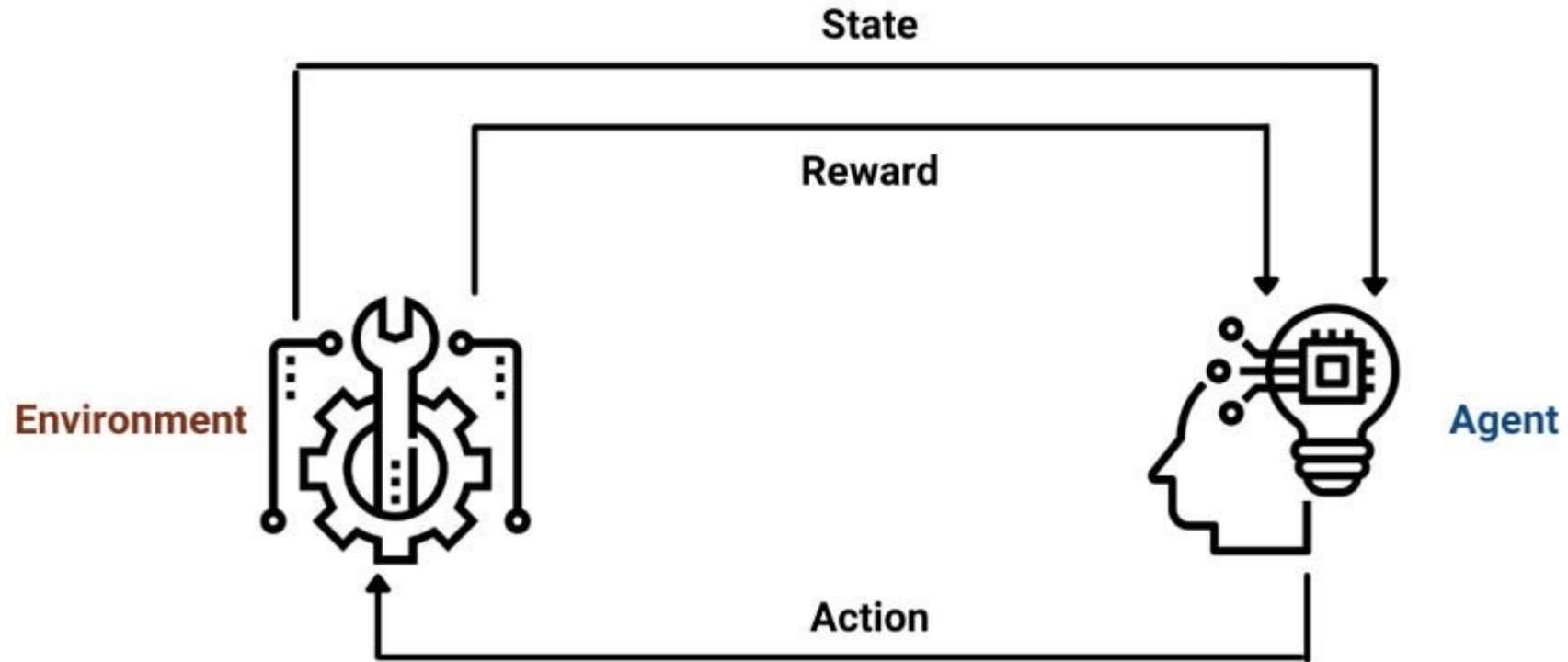- ❑ Moderate average monthly bill

Cluster 3 (Low Usage):

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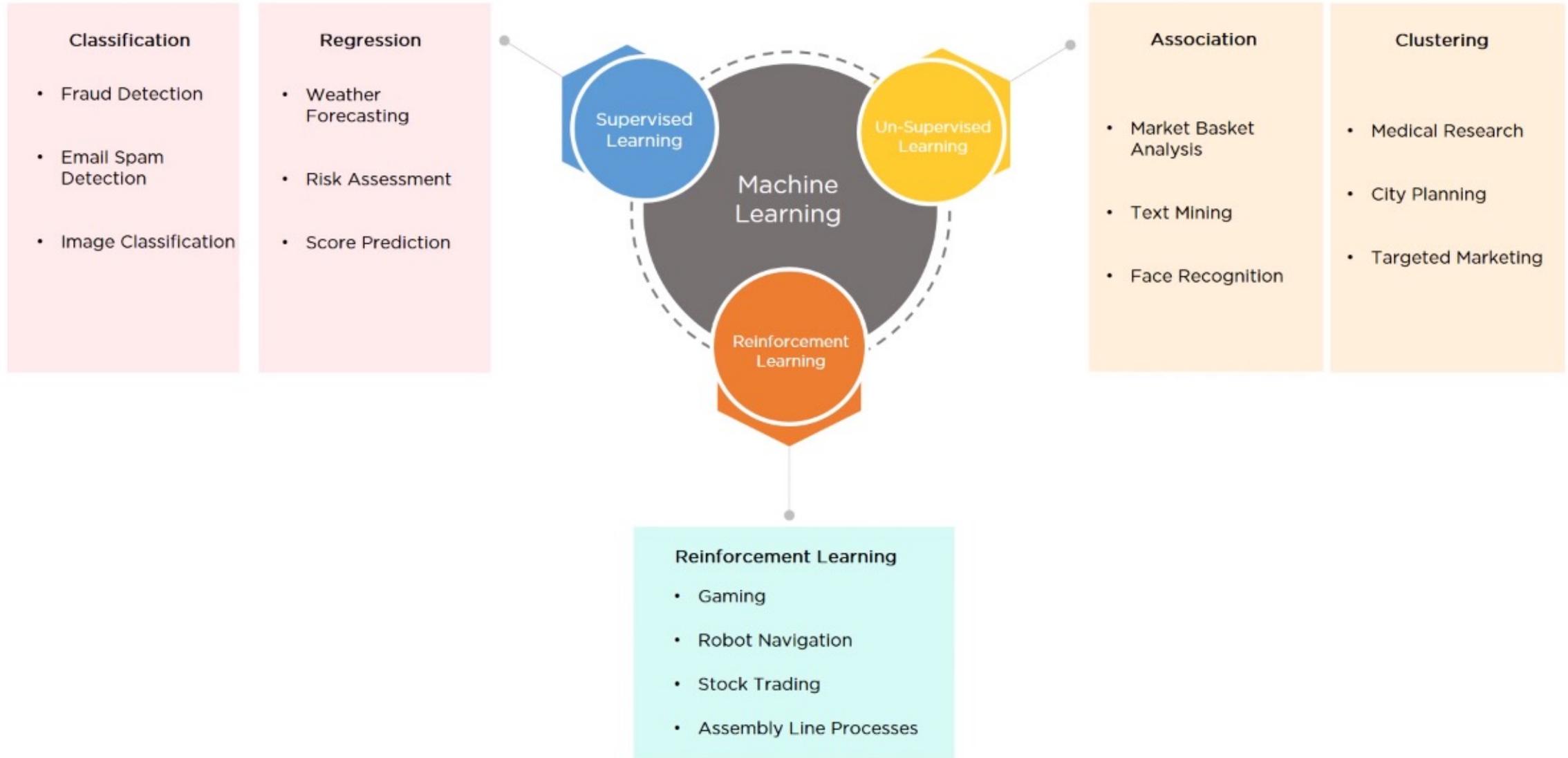- ❑ Lower average monthly bill

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

Supervised Learning

Un-Supervised Learning

Machine Learning

Reinforcement Learning

Reinforcement Learning

Reinforcement Learning

Classification
- Fraud Detection
- Email Spam Detection
- Image Classification

Regression
- Weather Forecasting
- Risk Assessment
- Score Prediction

Supervised Learning

Un-Supervised Learning

Machine Learning

Reinforcement Learning

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
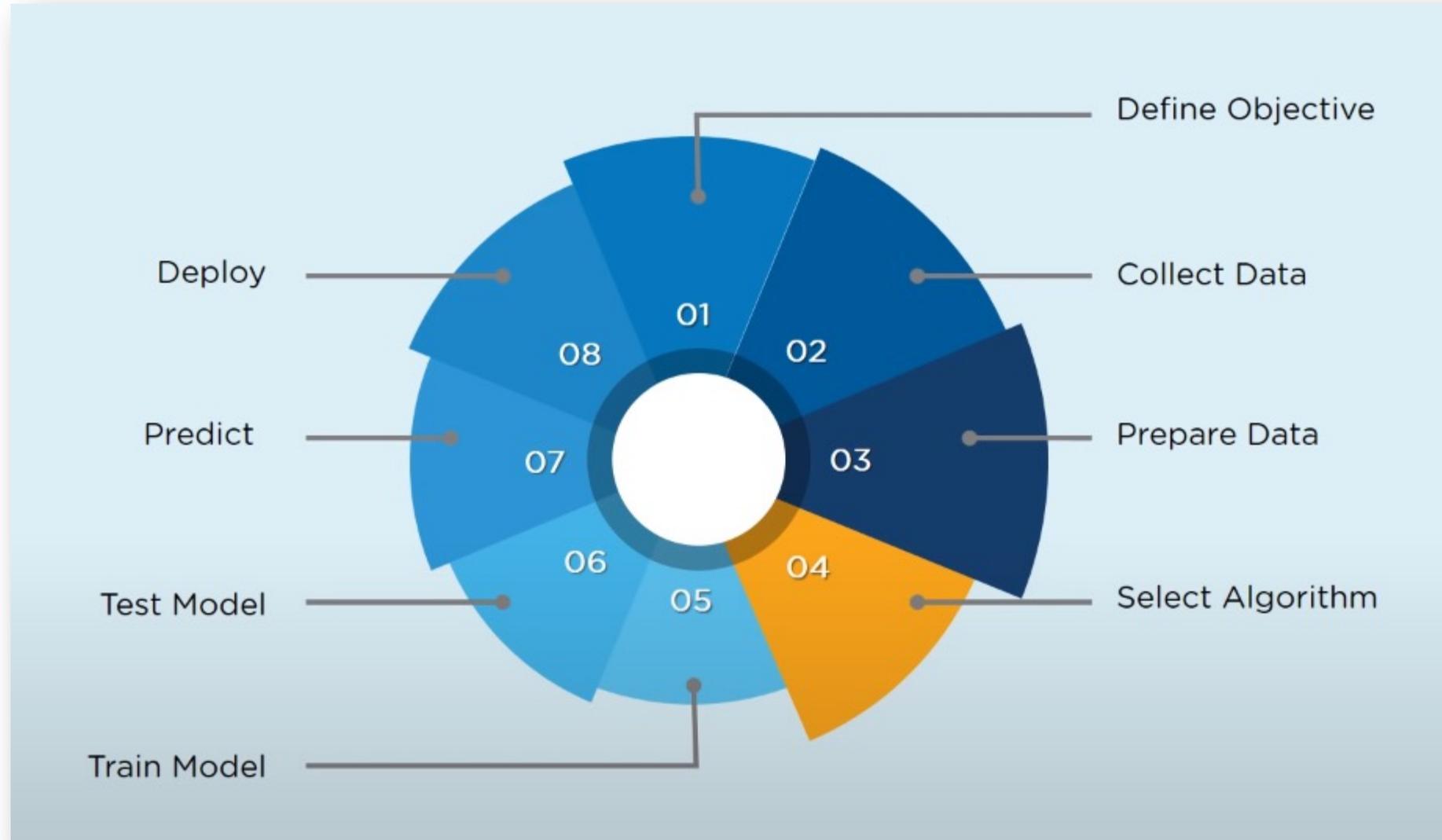
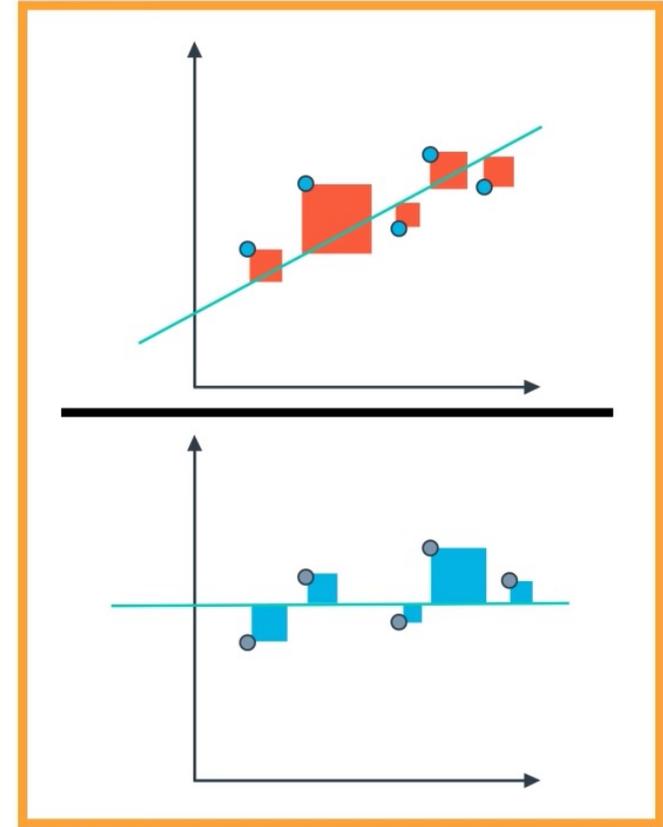# Performance Metrics - Regression

○ **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.

$$R2 = 1 - \frac{}{}$$

R2 (R-Square)score can be interpreted as a coefficient of determination where value propagate from

0 to 1, where 1 is the best.

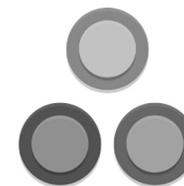# Performance Metrics - Classification

CONFUSION MATRIX

Performance Metrics - Classification

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

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

Major ML supported Languages

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

These language provide all necessary ML packages

Hands-on

Supervised Machine Learning

IDE   : No Code Machine Learning with **Azure Machine Learning Studio (Classic)**
Desktop Based Data Mining   : **Orange**

Hands-on

Step 1 : Please go to this site https://studio.azureml.net/
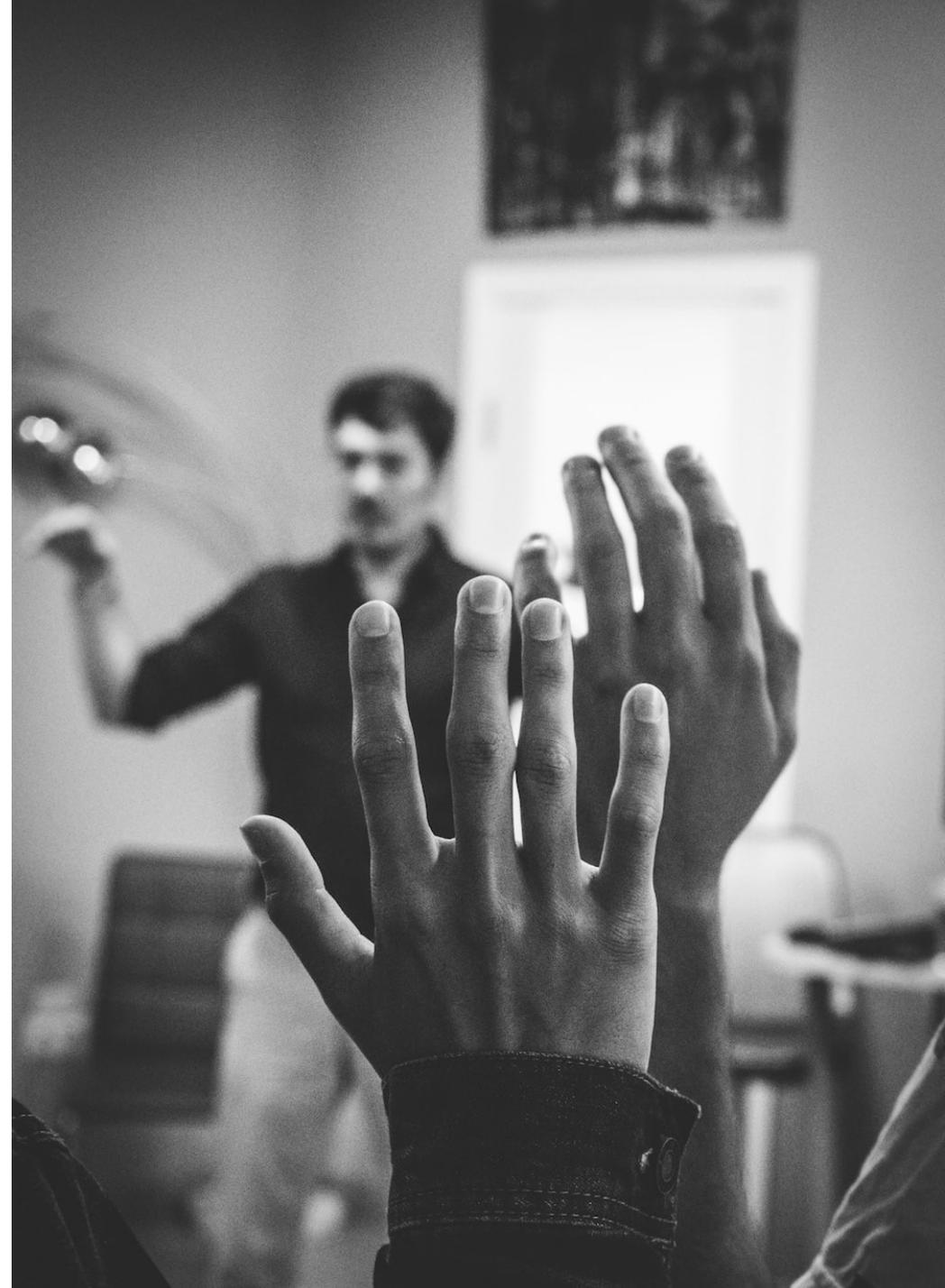
Step 2 : Use any Microsoft Account to Register and Login

Step 3 : Let's do some Prediction

Please go to this page for all resources : http://arif.works/azureml

Algorithms

1 Ask Questions to your Data !

2 Understand your Numbers !

3 **Visualize** your Data!

4 Data **Storytelling**
Connect Dots!

5 Data **Modelling**
Mingle your Data!

6 Predictive Analytics – with ML