


Associate Rule Learning

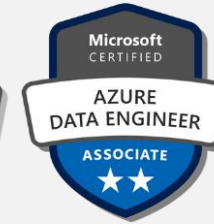


Presenter



 /arifmazumder

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



Mohammed Arif has more than twelve (13) years of working experience in Information Communication and Technology (ICT) industry. The highlights of his career are more than six (7) years of holding various senior management and/or C-Level and had five (5) 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.



Agenda

- What is Data Science
- Data
- Components of Data Science
- Machine Learning
- Types of Machine Learning
- Steps to do Machine Learning

Agenda

- Recommender System
- Recommender System - Solutions
- Association Rule
- Association Rule – Performance Measure
- Association Rule Generation – Apriori Algorithm
- Hands-on Problem Solving (Using Python & Azure ML)

Resource Link

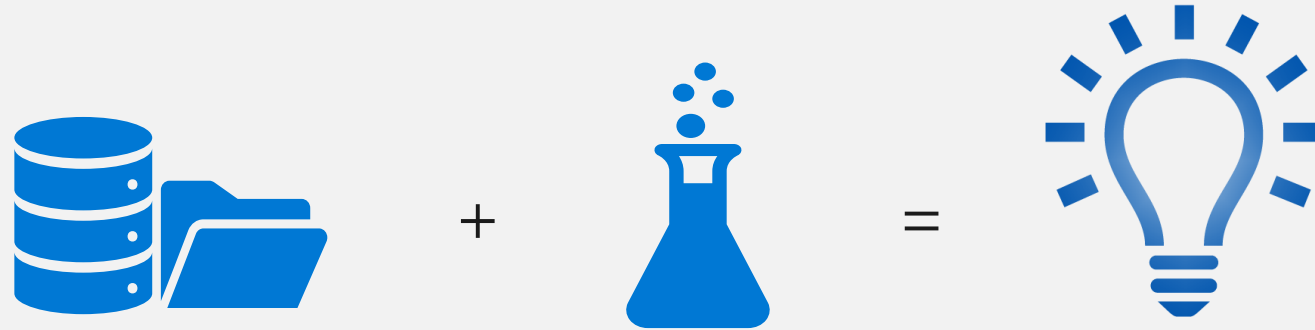
<http://arif.works/mltaylor/>

Resource Link

<http://arif.works/mltaylor/>

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



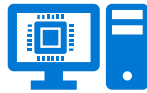
Velocity

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 $\stackrel{?}{=}$ Analytics

When we need to apply Machine Learning

Analysis



Past

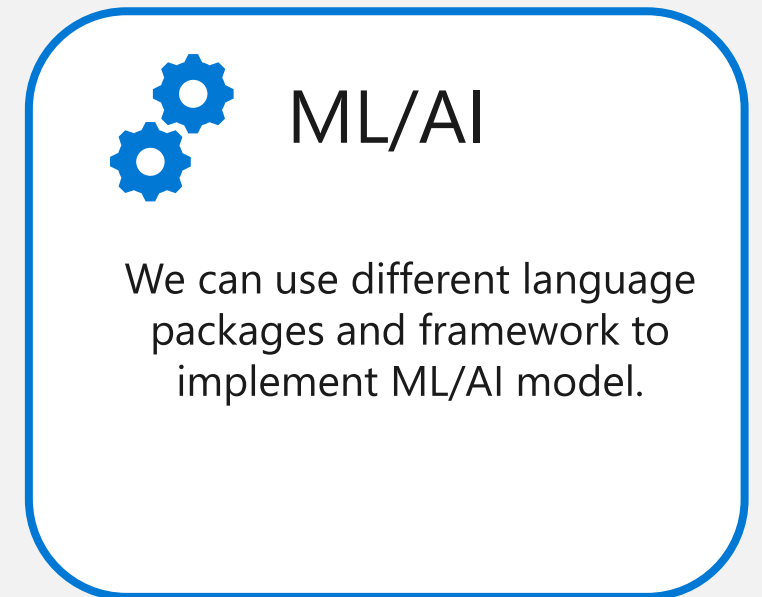
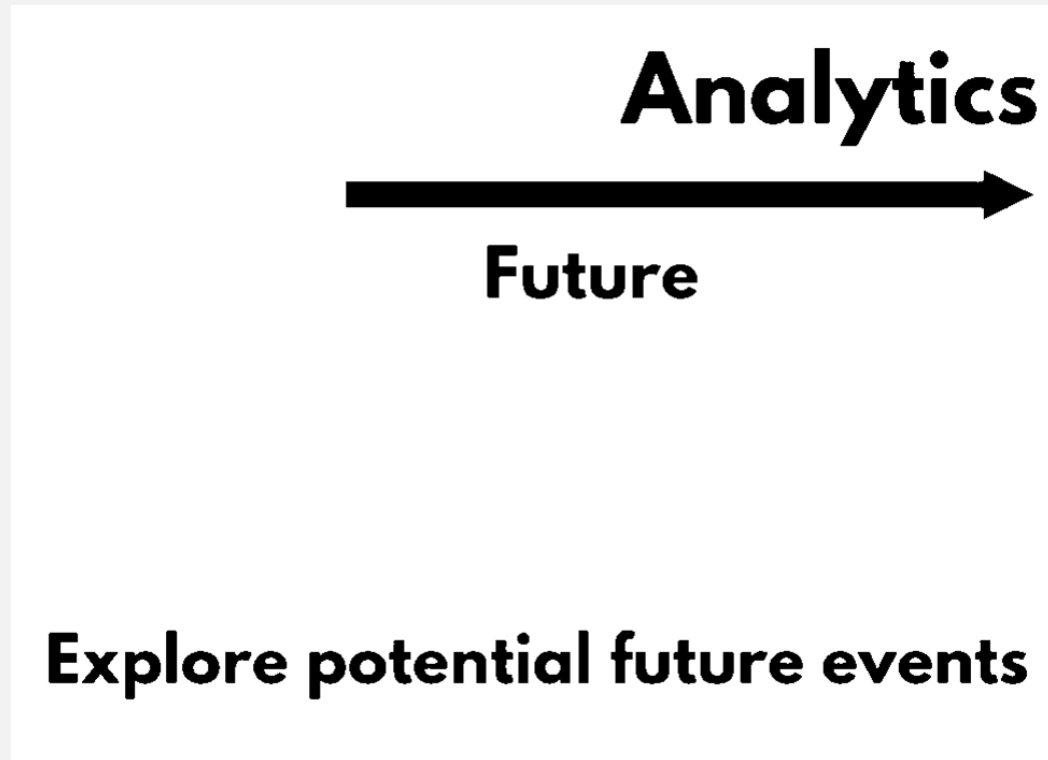
Explain
How? Why?



BI

We can use different tools to explain the previous trends like, PowerBI, Tableau, Qlikview etc.

When we need to apply Machine Learning



Business Analytics



Descriptive

Explains what happened.



Diagnostic

Explains why it happened.



Predictive

Forecasts what might happen.



Prescriptive

Recommends an action based on the forecast.

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 →

Types of Machine Learning

Supervised

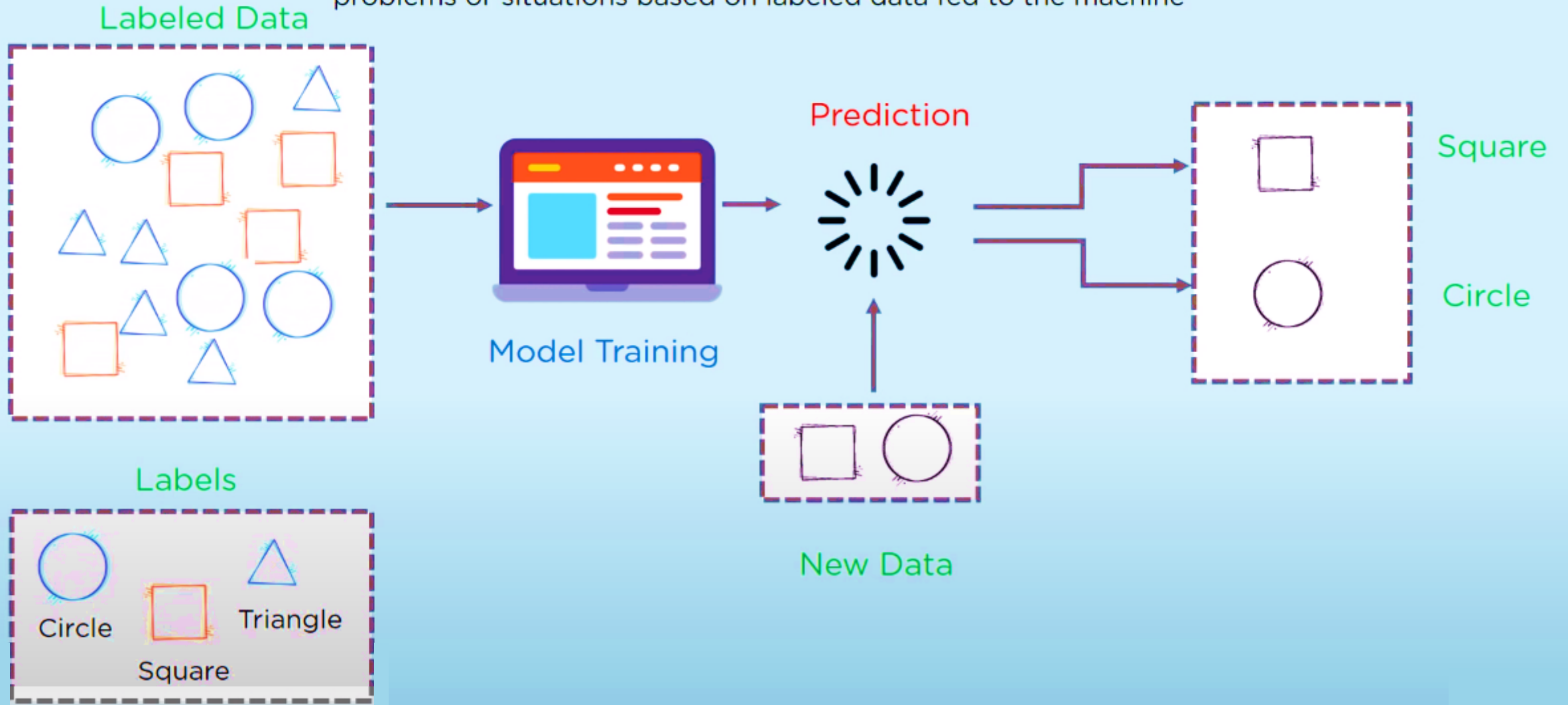
Reinforcement

Un-Supervised



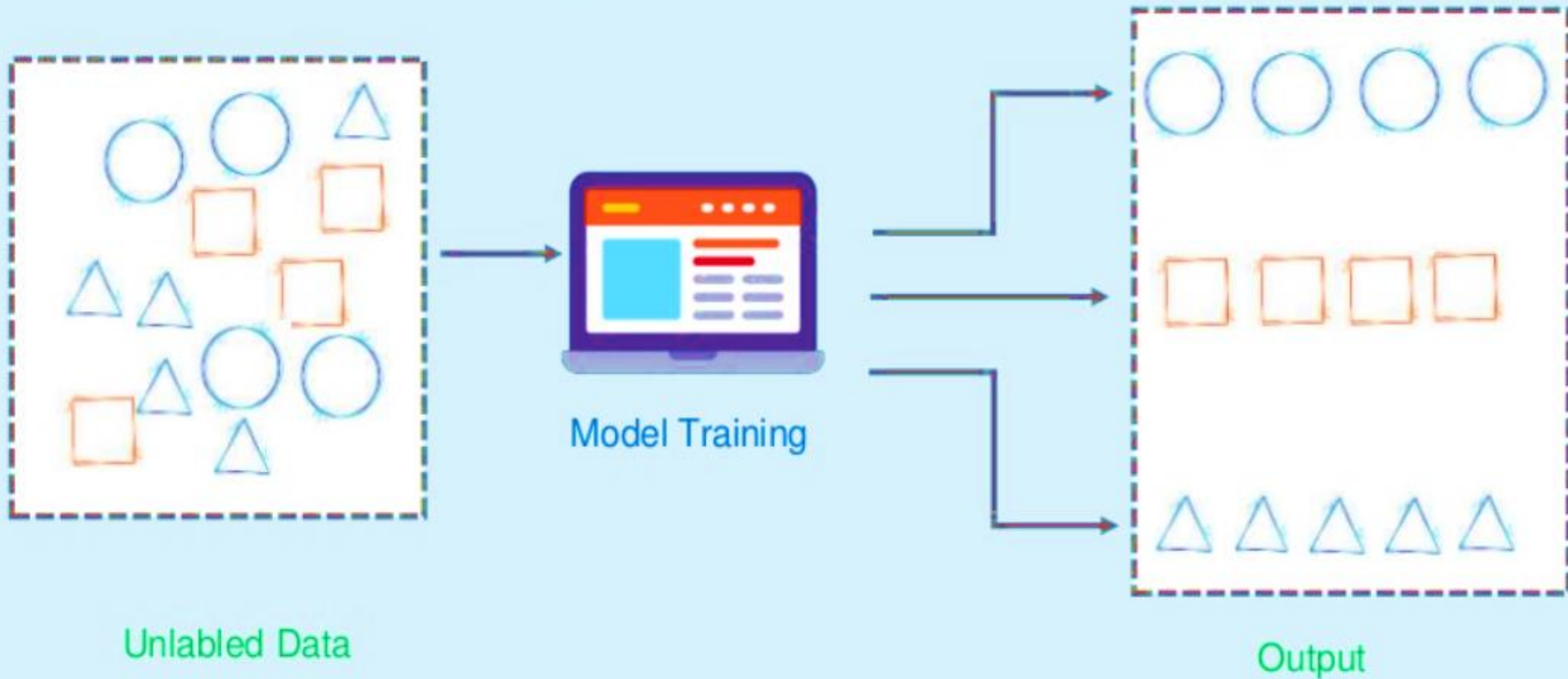
Supervised Learning

Supervised learning is a method used to enable machines to classify/ predict objects, problems or situations based on labeled data fed to the machine



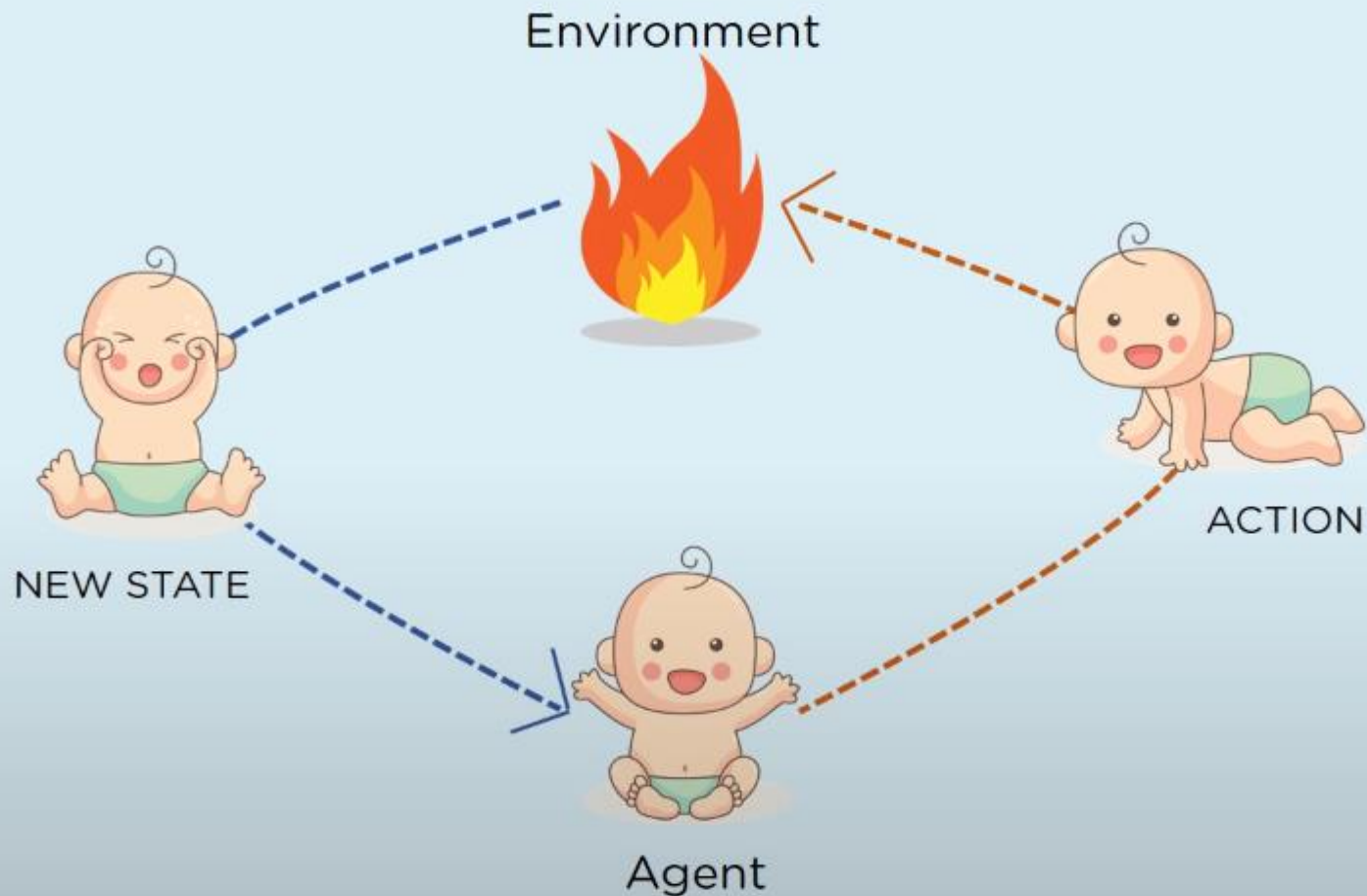
Unsupervised Learning

In Unsupervised learning, Machine Learning model finds the hidden pattern in an unlabeled data



Reinforcement Learning

Reinforcement learning is an important type of Machine Learning where an agent learns how to behave in an environment by performing actions and seeing the results



Types of Machine Learning

Classification

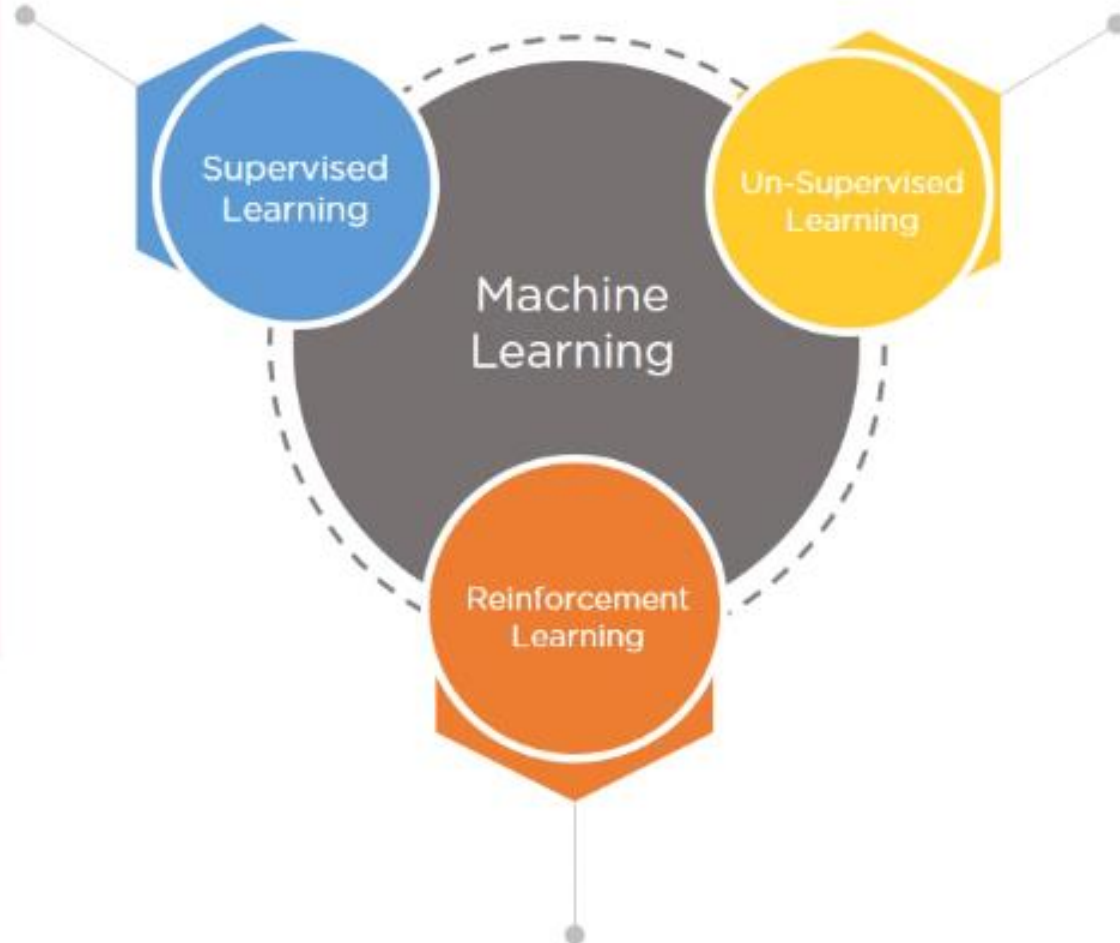
- Fraud Detection
- Email Spam Detection
- Image Classification

Categorical

Regression

- Weather Forecasting
- Risk Assessment
- Score Prediction

Numerical



Types of Machine Learning

Supervised learning, algorithms are trained using marked data, where the input and the output are known.

ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

labels

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

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

Types of Machine Learning

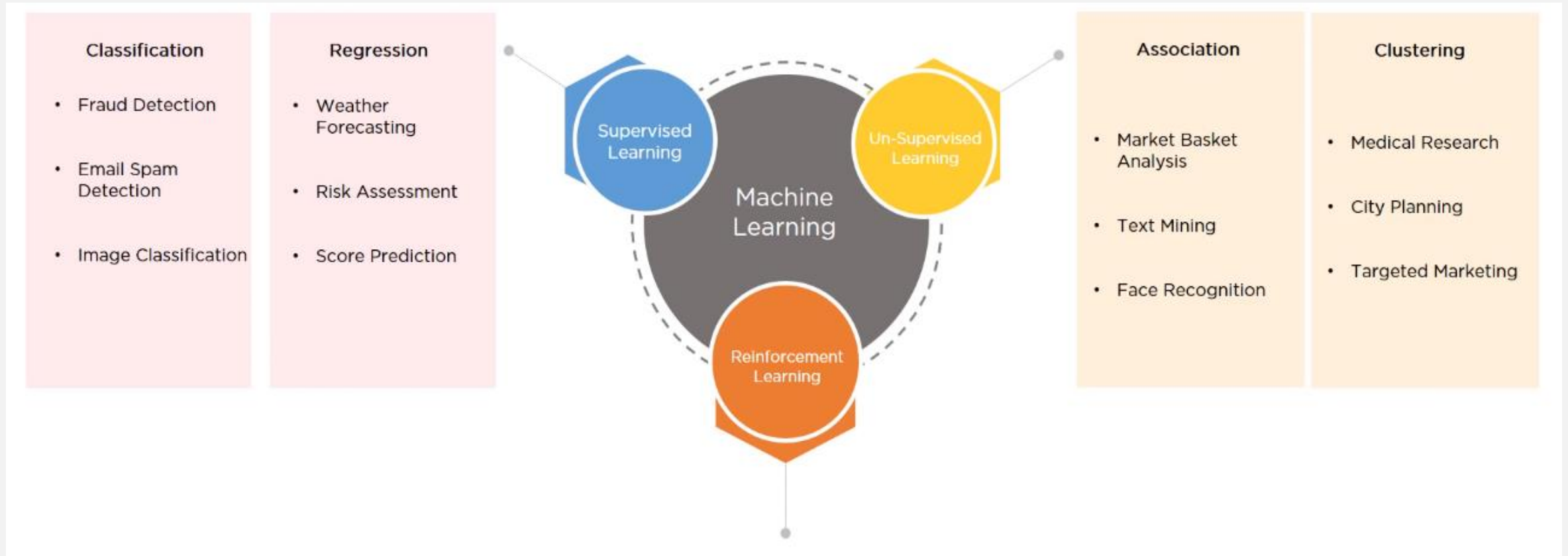
User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	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

Types of Machine Learning



Types of Machine Learning

Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Address	DebtIncomeRatio
1	41	2	6	19	0.124	1.073	NBA001	6.3
2	47	1	26	100	4.582	8.218	NBA021	12.8
3	33	2	10	57	6.111	5.802	NBA013	20.9
4	29	2	4	19	0.681	0.516	NBA009	6.3
5	47	1	31	253	9.308	8.908	NBA008	7.2
6	40	1	23	81	0.998	7.831	NBA016	10.9
7	38	2	4	56	0.442	0.454	NBA013	1.6
8	42	3	0	64	0.279	3.945	NBA009	6.6
9	26	1	5	18	0.575	2.215	NBA006	15.5
10	47	3	23	115	0.653	3.947	NBA011	4
11	44	3	8	88	0.285	5.083	NBA010	6.1
12	34	2	9	40	0.374	0.266	NBA003	1.6

unlabeled



Types of Machine Learning



Labeled Data



Direct feedback



Predict output

Supervised

vs

Unsupervised



Non-labeled data

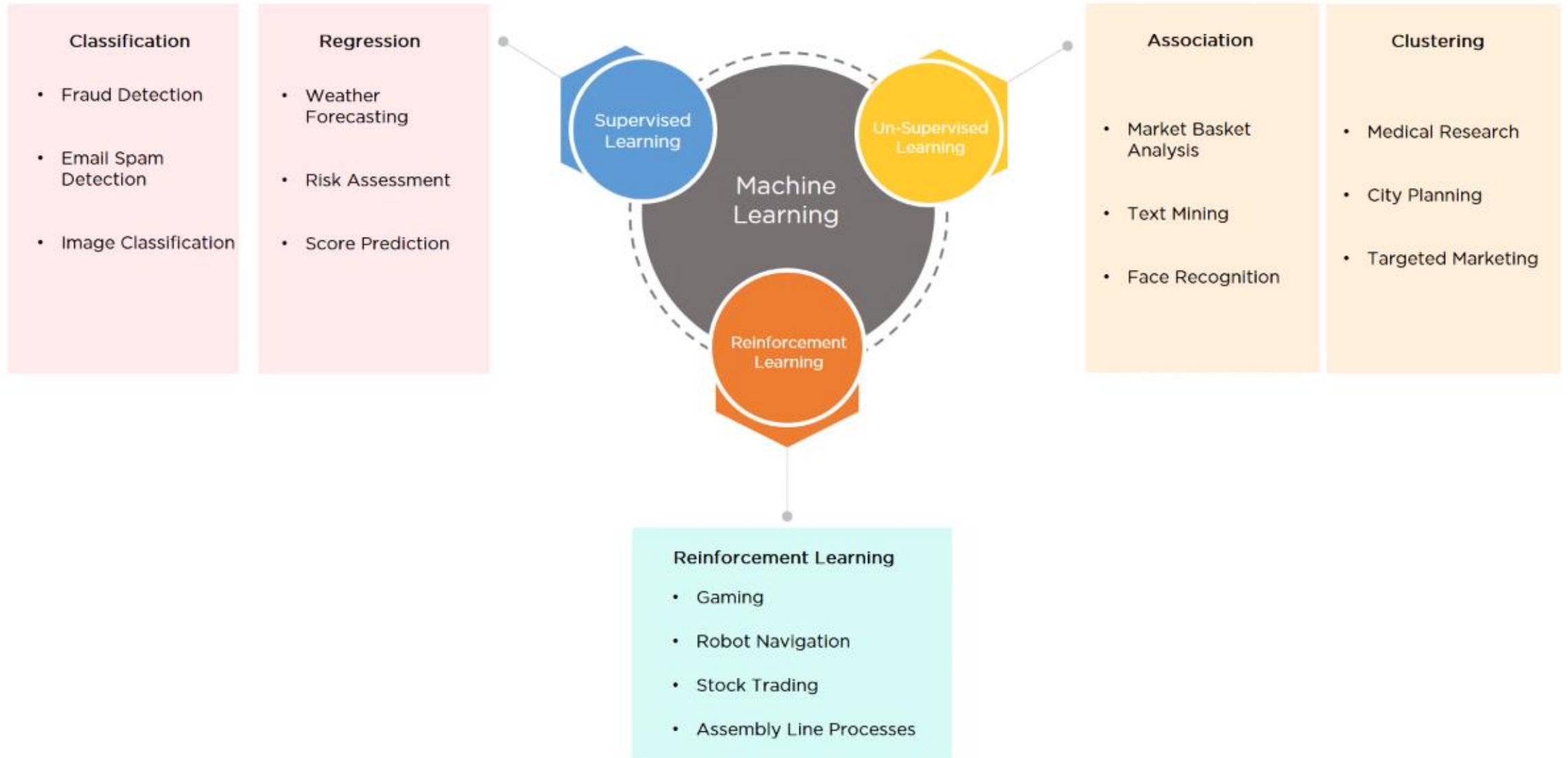


No feedback

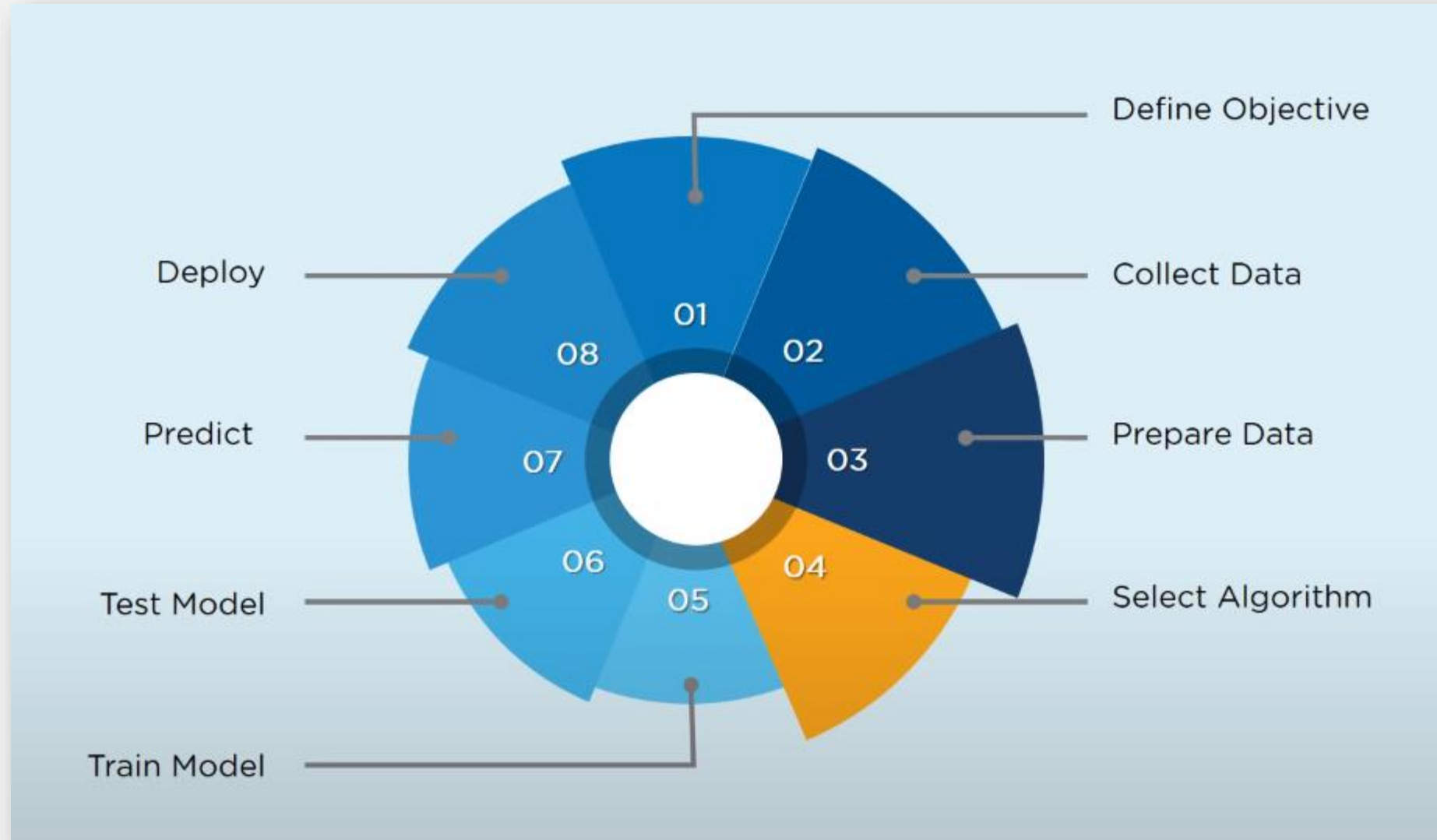


Find hidden
structure in data

Types of Machine Learning



Processing Steps for Machine Learning



Recommender System

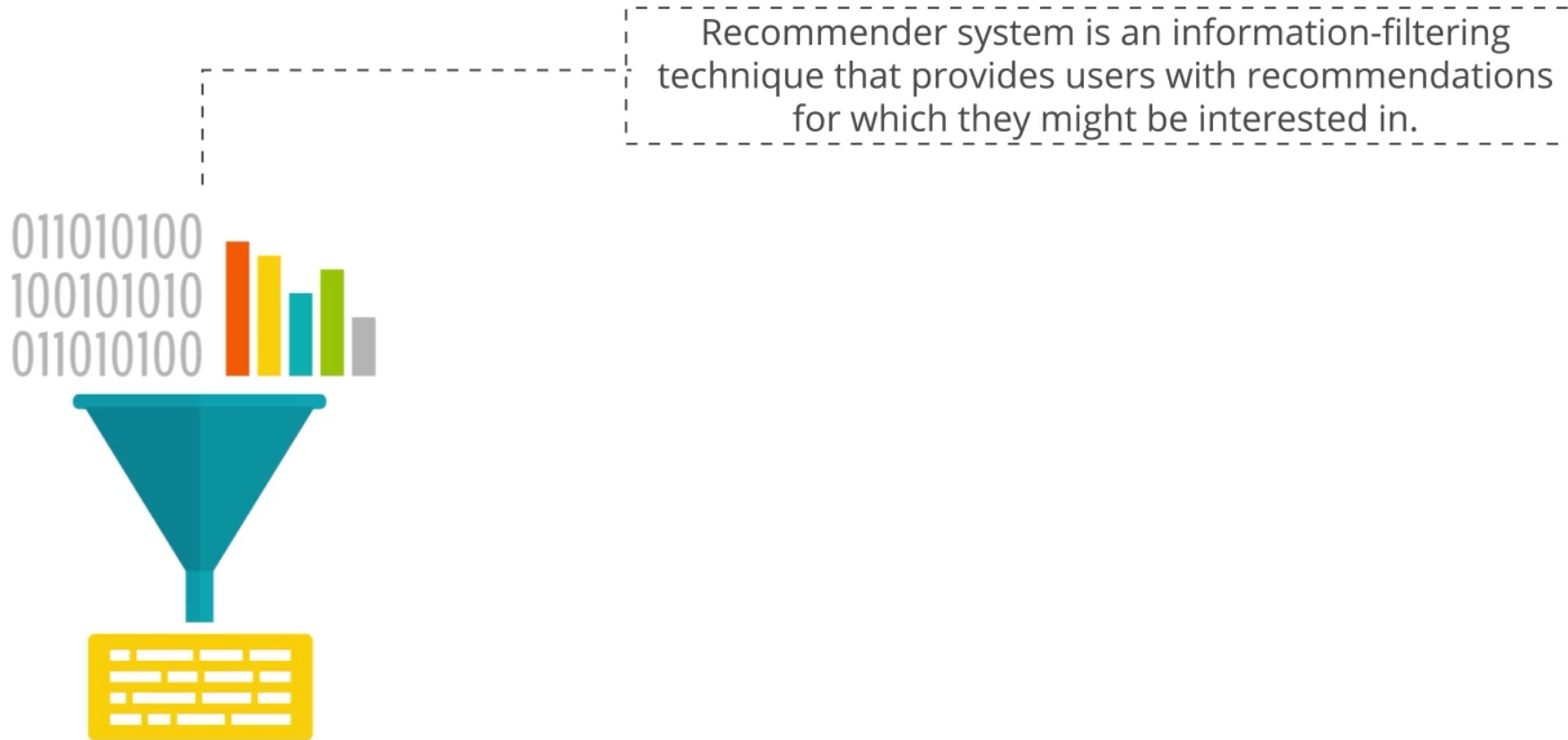


A recommender system predicts users' interests and recommends products that the users may be interested in.

Recommender System (Gather Data)



Recommender System



Recommender System



				
	✓	✗	✓	✓
		✓	✗	✗
	✓	✓	✗	
	✗		✓	
	✓	✓		✗

Recommender System : Solutions



Association Rule Mining

Association rule mining uses machine learning models to analyze data for [patterns](#) or [co-occurrence](#) in a database



Each transaction is considered to be a list of items.



Association rule finds all rules that correlate the presence of one set of items with that of another set of items.



It identifies frequent patterns.



It is most commonly used for market basket analysis.

Association Rule (How to create)



Searching data for
frequent if-then patterns



Using the criteria
support



Identifying the most
important relationships.

Association Rule (Performance Measure)



- Indicates how frequently the items appear in the data
- Provides fraction of transactions that contain X and Y
- Formula = No. of times item X occurred / Total number of transactions = $P(X \cup Y) = \frac{\sigma(x \cup y)}{N}$

Association Rule (Performance Measure)



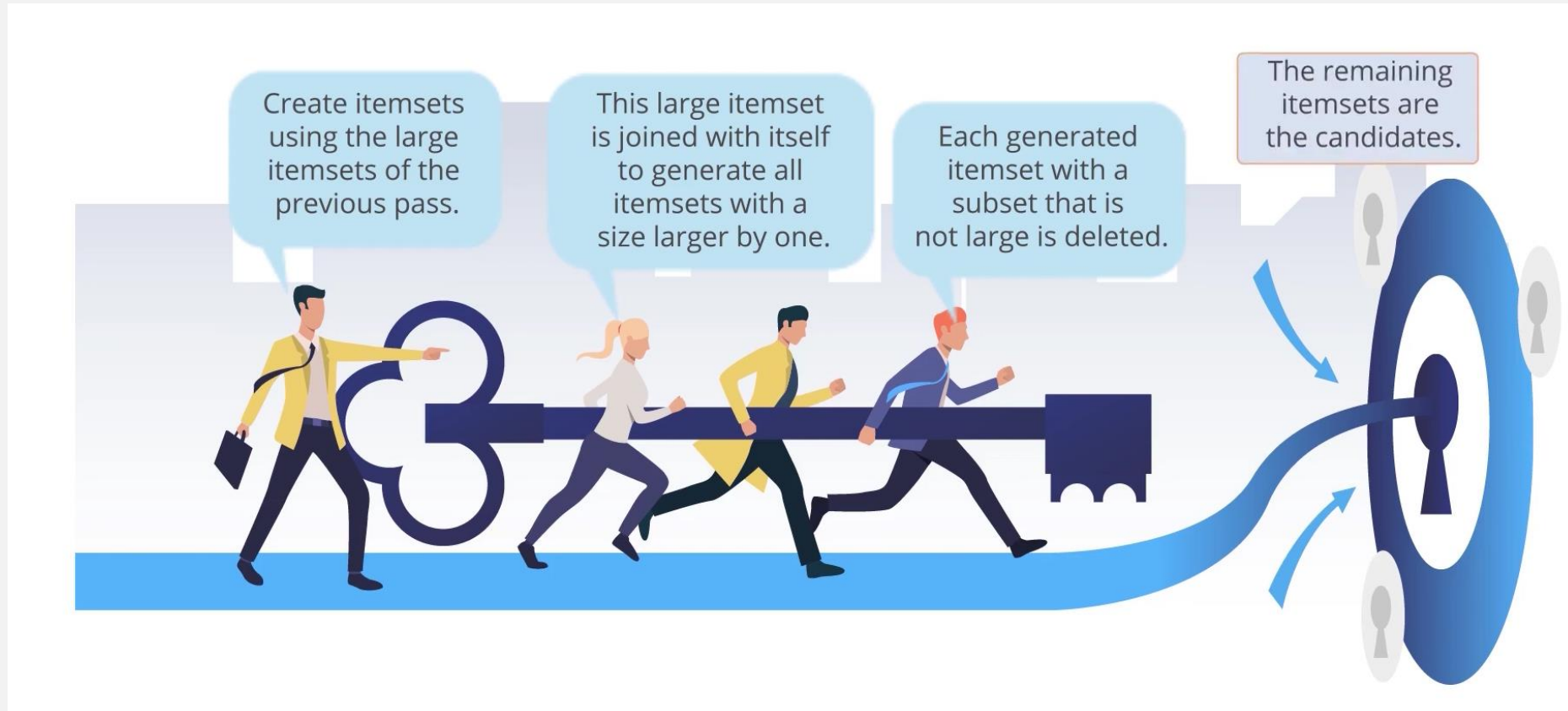
- Indicates the number of times the if-then statements are found true
- Indicates how often X and Y occur together, given the no. of times X occurs
- Formula = No. of times X and Y occurred / Total occurrence of X = $\Pr(Y | X) = \frac{\sigma(x \cup y)}{\sigma(x)}$

Association Rule (Performance Measure)



- Compare the actual confidence with the expected confidence
- Indicates the strength of a rule over the random co-occurrence of X and Y
- Formula = No. of times item X and Y occurred / Total occurrence of X multiplied by Total occurrence of Y =
$$\frac{\sigma(x \cup y)}{\sigma(x) \times \sigma(y)}$$

Association Rule (Apriori Algorithm)



Association Rule (Apriori Algorithm)



Uses frequent itemsets to generate association rules



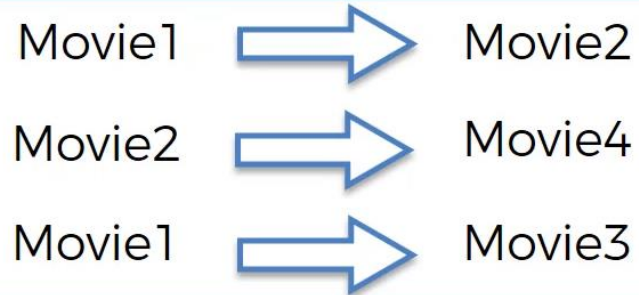
Support value of frequent itemsets is greater than the threshold value

The algorithm reduces the number of candidates being considered by only exploring the itemsets whose support count is greater than the minimum support count.

ARL

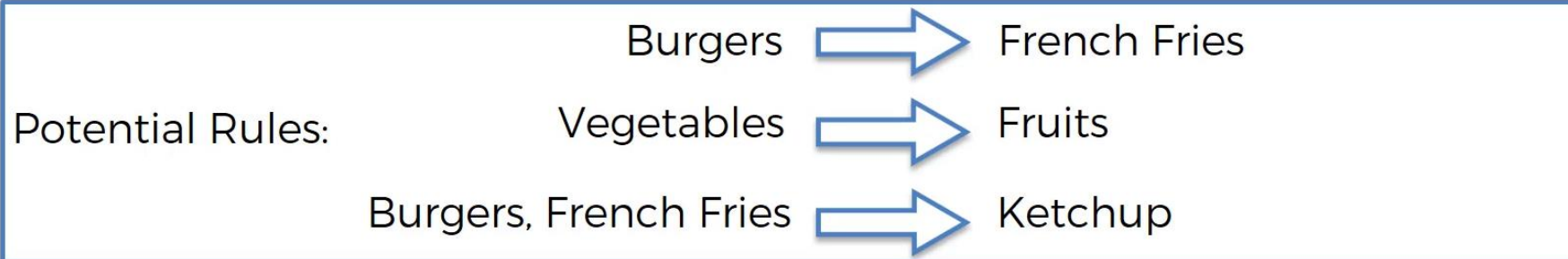
User ID	Movies liked
46578	Movie1, Movie2, Movie3, Movie4
98989	Movie1, Movie2
71527	Movie1, Movie2, Movie4
78981	Movie1, Movie2
89192	Movie2, Movie4
61557	Movie1, Movie3

Potential Rules:



ARL

Transaction ID	Products purchased
46578	Burgers, French Fries, Vegetables
98989	Burgers, French Fries, Ketchup
71527	Vegetables, Fruits
78981	Pasta, Fruits, Butter, Vegetables
89192	Burgers, Pasta, French Fries
61557	Fruits, Orange Juice, Vegetables
87923	Burgers, French Fries, Ketchup, Mayo



How Apriori Algorithm Works - Support

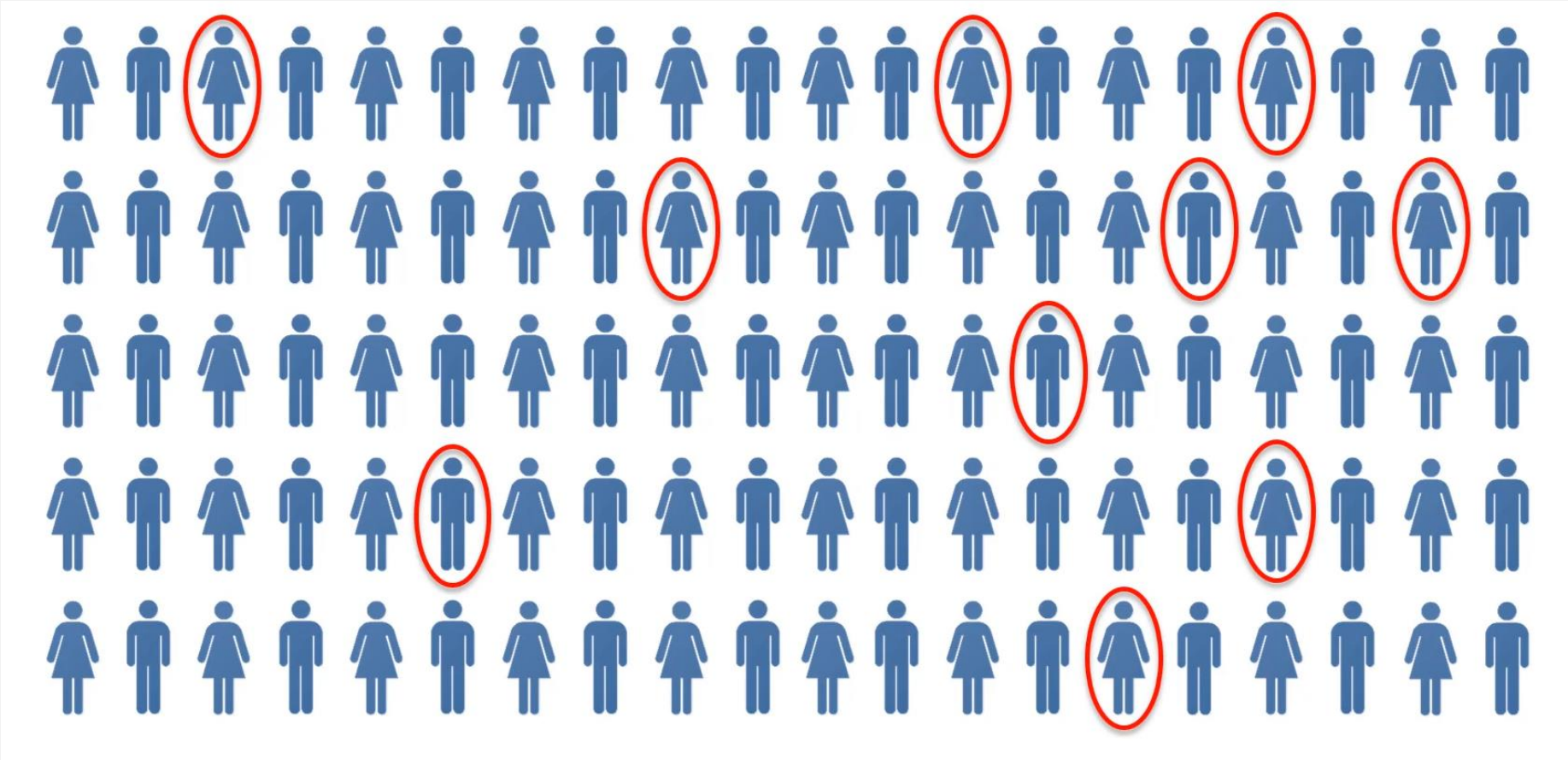
Movie Recommendation: $\text{support}(\mathbf{M}) = \frac{\# \text{ user watchlists containing } \mathbf{M}}{\# \text{ user watchlists}}$

Market Basket Optimisation: $\text{support}(\mathbf{I}) = \frac{\# \text{ transactions containing } \mathbf{I}}{\# \text{ transactions}}$

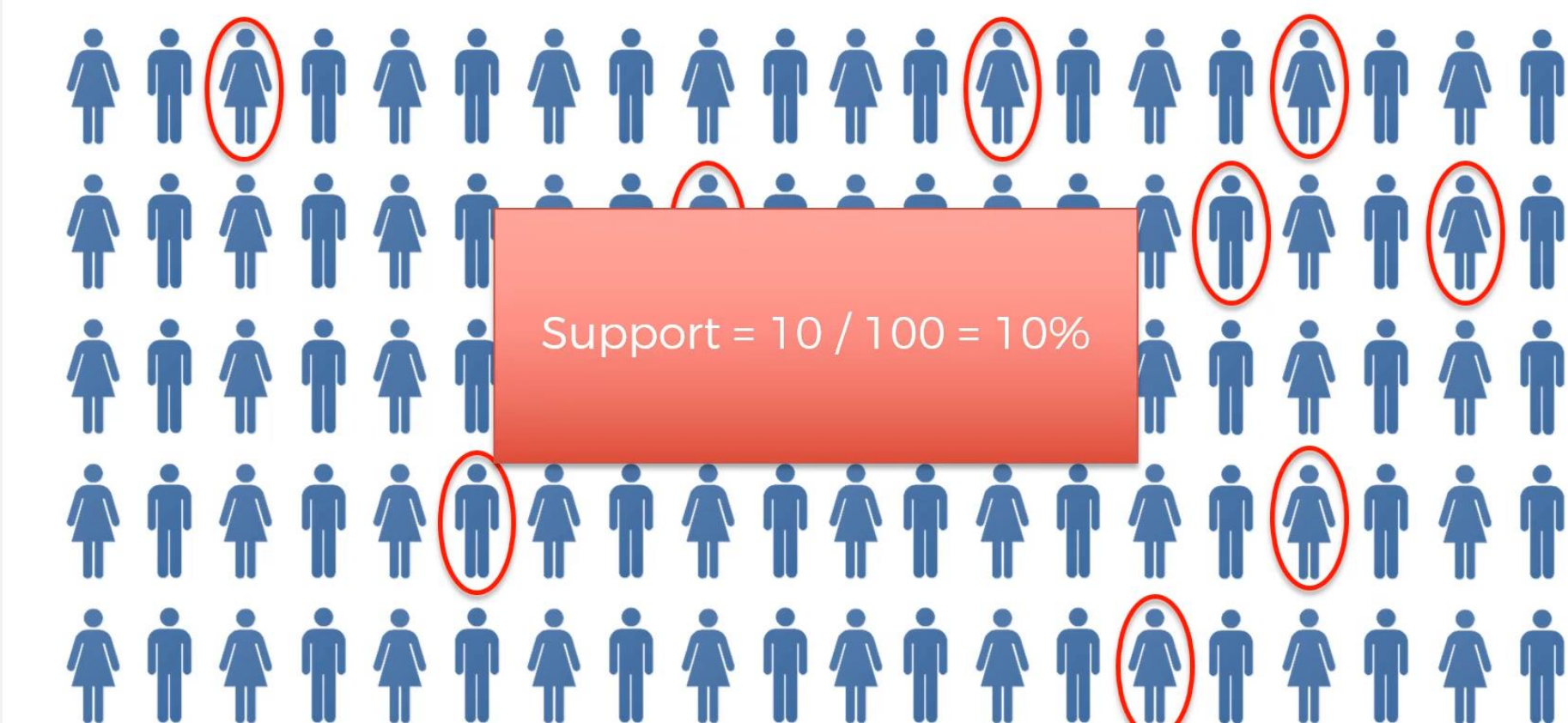
How Apriori Algorithm Works - Support



How Apriori Algorithm Works - Support



How Apriori Algorithm Works - Support



How Apriori Algorithm Works - Confidence

Movie Recommendation: $\text{confidence}(\mathbf{M_1} \rightarrow \mathbf{M_2}) = \frac{\# \text{ user watchlists containing } \mathbf{M_1} \text{ and } \mathbf{M_2}}{\# \text{ user watchlists containing } \mathbf{M_1}}$

Market Basket Optimisation: $\text{confidence}(\mathbf{l_1} \rightarrow \mathbf{l_2}) = \frac{\# \text{ transactions containing } \mathbf{l_1} \text{ and } \mathbf{l_2}}{\# \text{ transactions containing } \mathbf{l_1}}$

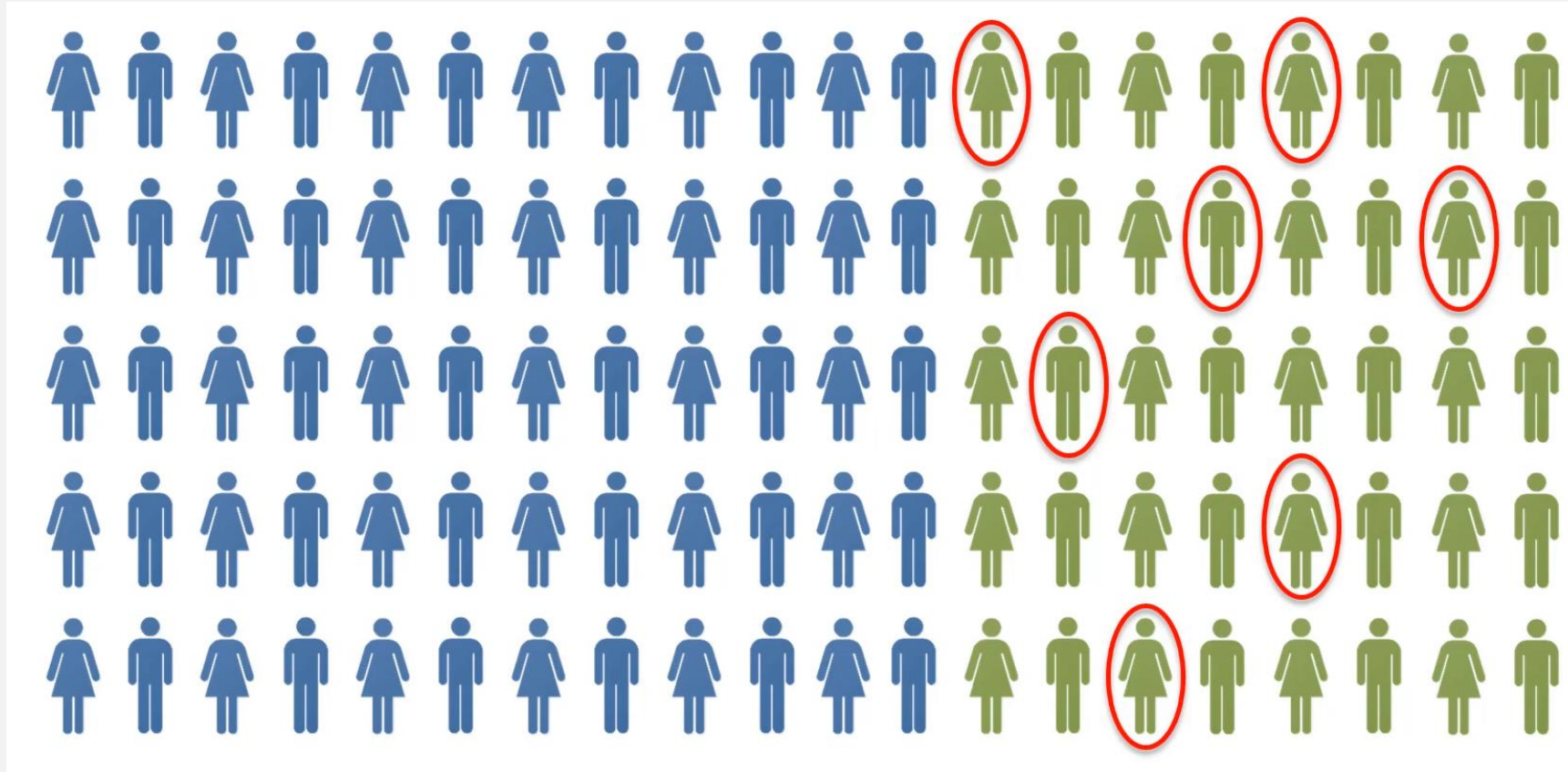
How Apriori Algorithm Works - Confidence

Green Color people Seen Movie [Interstellar](#)



How Apriori Algorithm Works - Confidence

Green Color people Seen Movie [Interstellar](#). Out of 40 , seven (7) people also see [Ex Machina](#)



How Apriori Algorithm Works - Confidence

Green Color people Seen Movie [Interstellar](#). Out of 40 , seven (7) people also see [Ex Machina](#)



How Apriori Algorithm Works - Lift

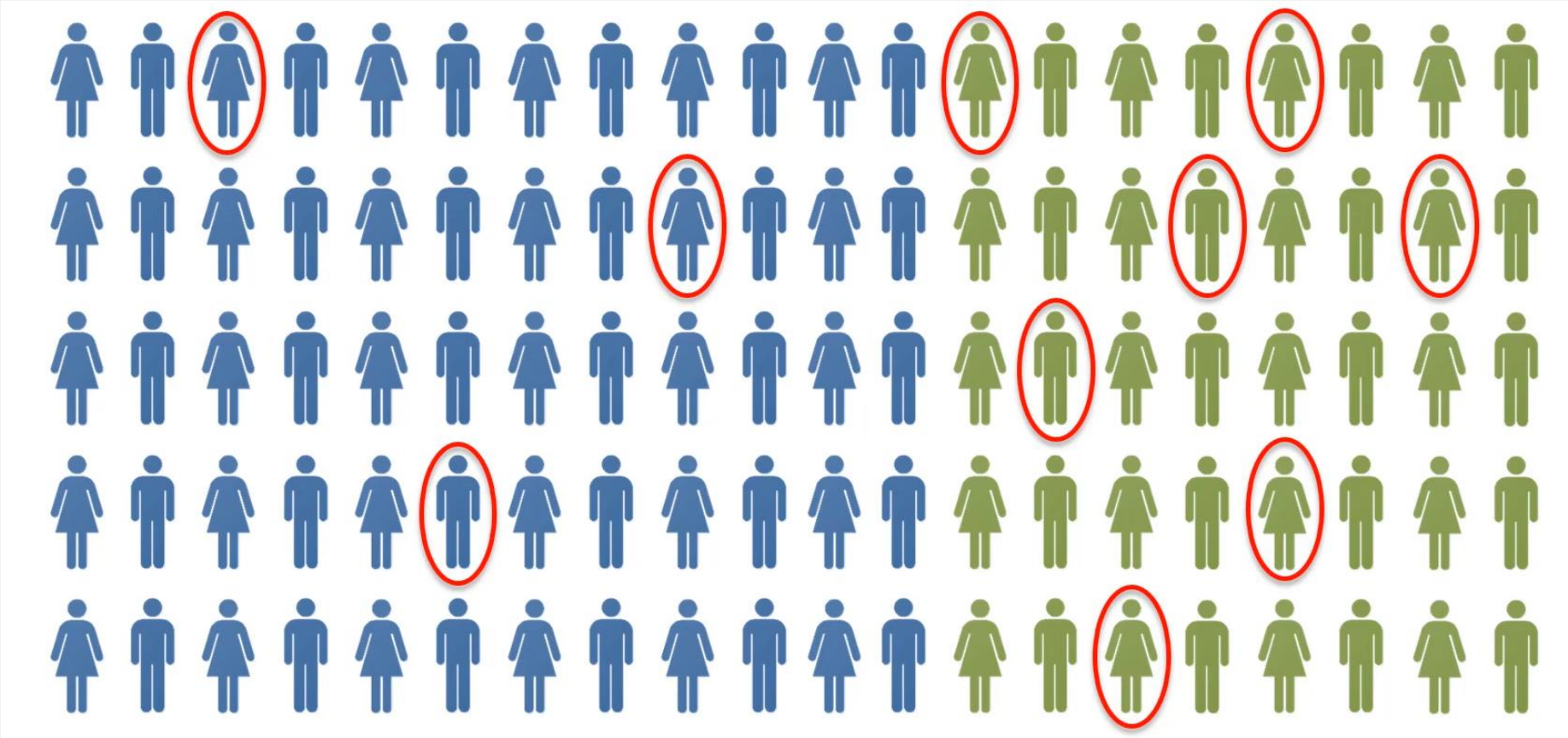
Movie Recommendation:

$$\text{lift}(\mathbf{M_1} \rightarrow \mathbf{M_2}) = \frac{\text{confidence}(\mathbf{M_1} \rightarrow \mathbf{M_2})}{\text{support}(\mathbf{M_2})}$$

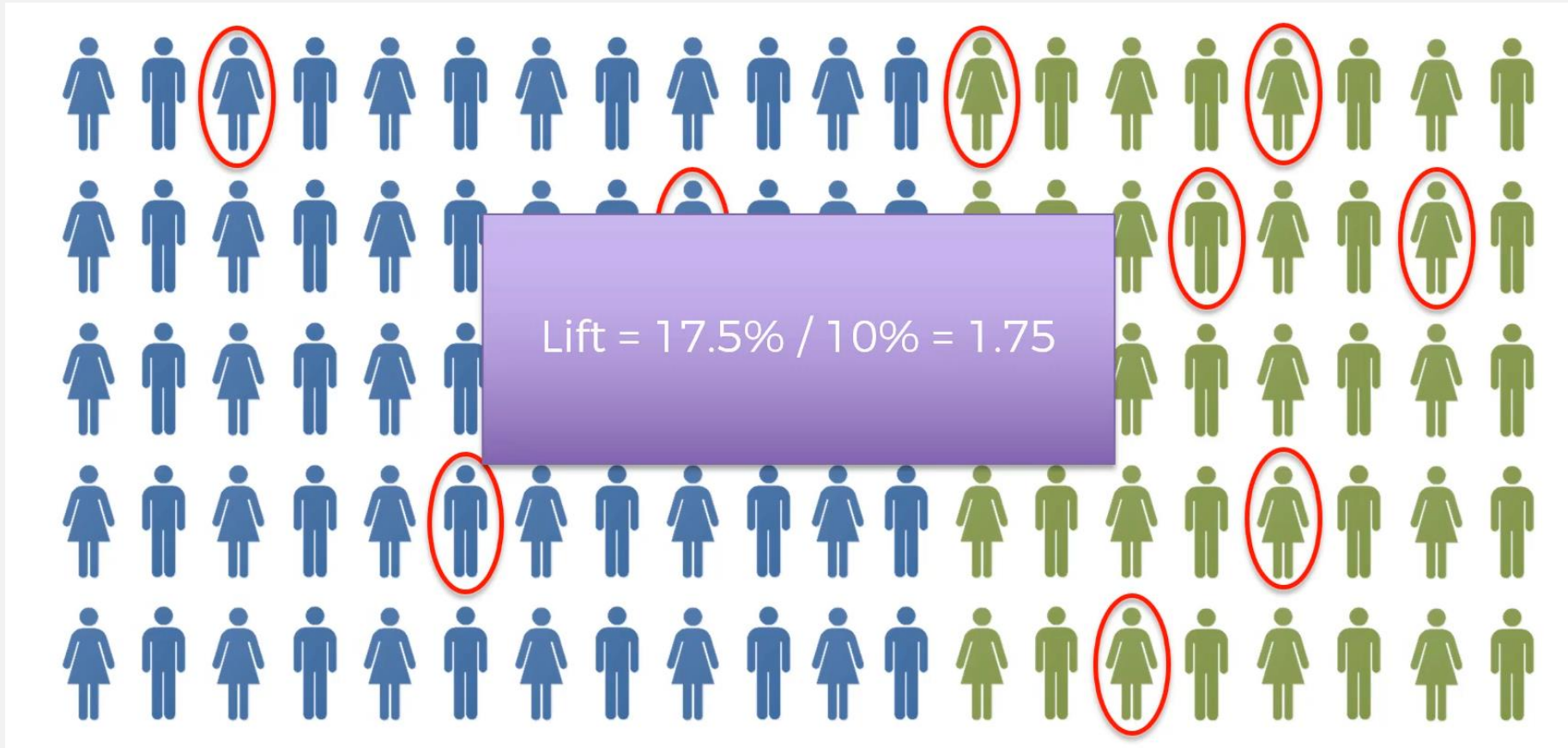
Market Basket Optimisation:

$$\text{lift}(\mathbf{l_1} \rightarrow \mathbf{l_2}) = \frac{\text{confidence}(\mathbf{l_1} \rightarrow \mathbf{l_2})}{\text{support}(\mathbf{l_2})}$$

How Apriori Algorithm Works - Lift



How Apriori Algorithm Works - Lift



How Apriori Algorithm Works

Step 1: Set a minimum support and confidence



Step 2: Take all the subsets in transactions having higher support than minimum support



Step 3: Take all the rules of these subsets having higher confidence than minimum confidence

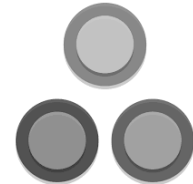


Step 4: Sort the rules by decreasing lift

Major ML supported Languages

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

These language provide all necessary ML packages



Hands-on