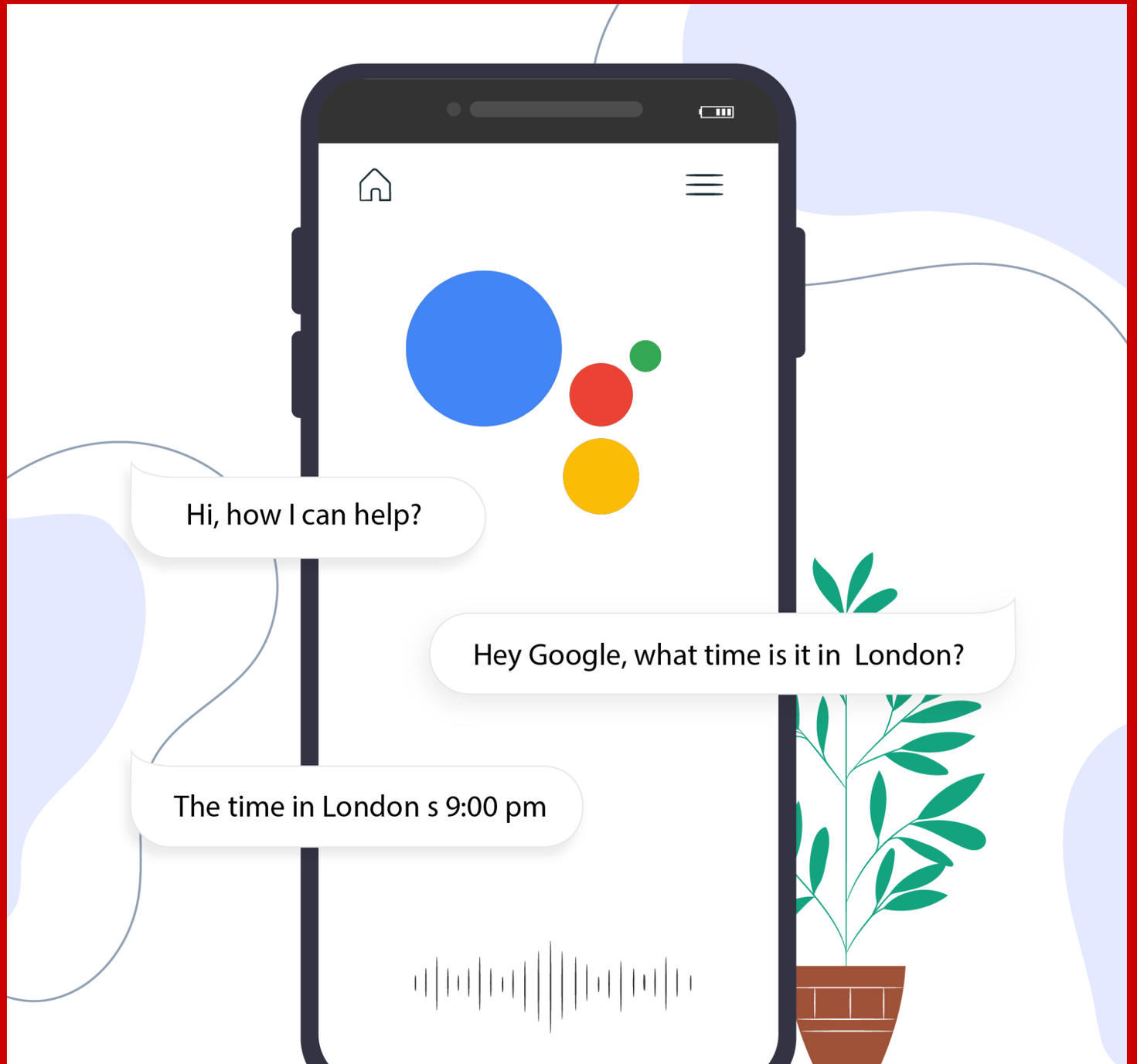
COMP30027 MACHINE LEARNING TUTORIAL

Workshop - 1

Real-World Applications of Machine Learning

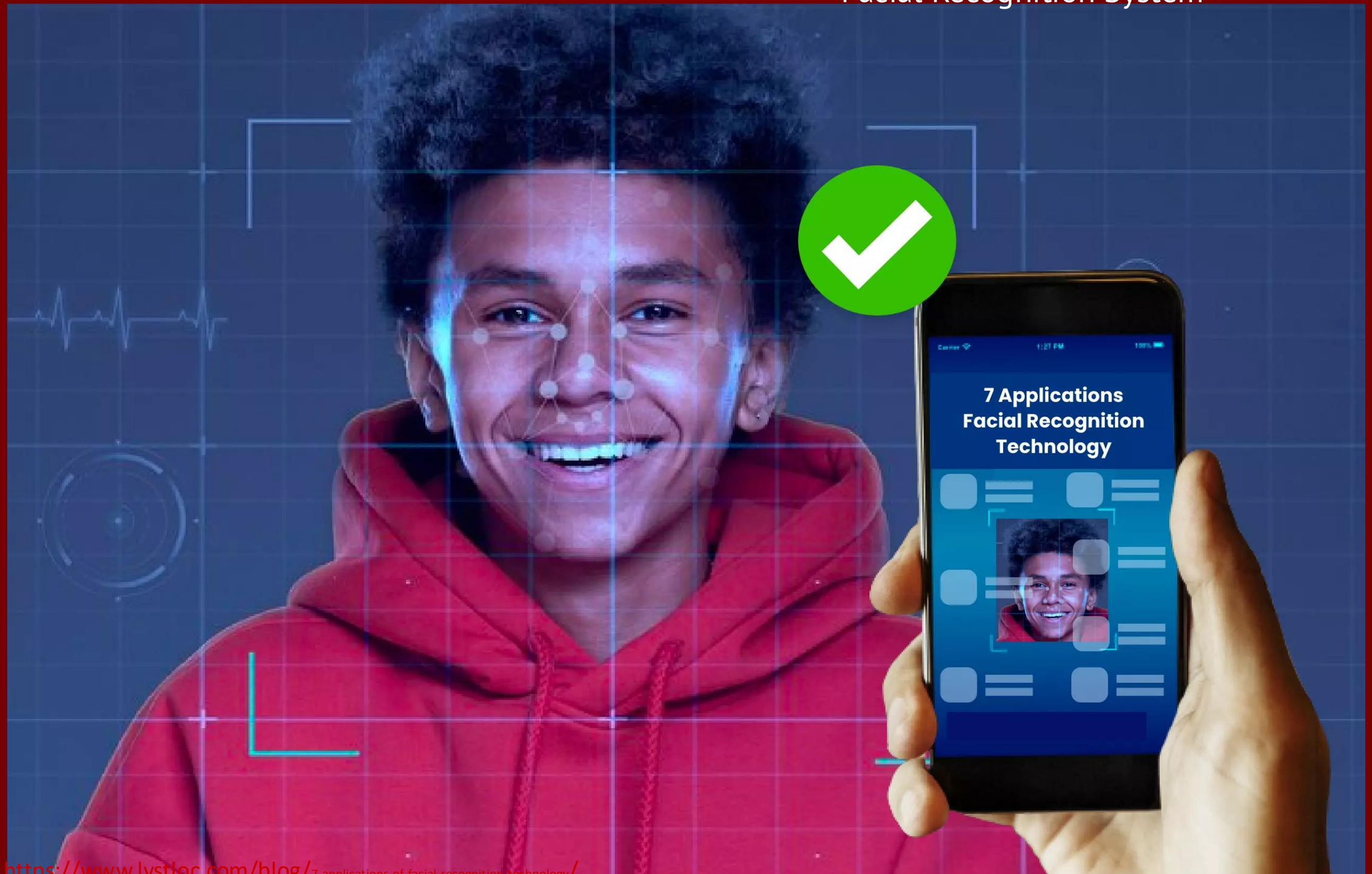
Voice Assistants



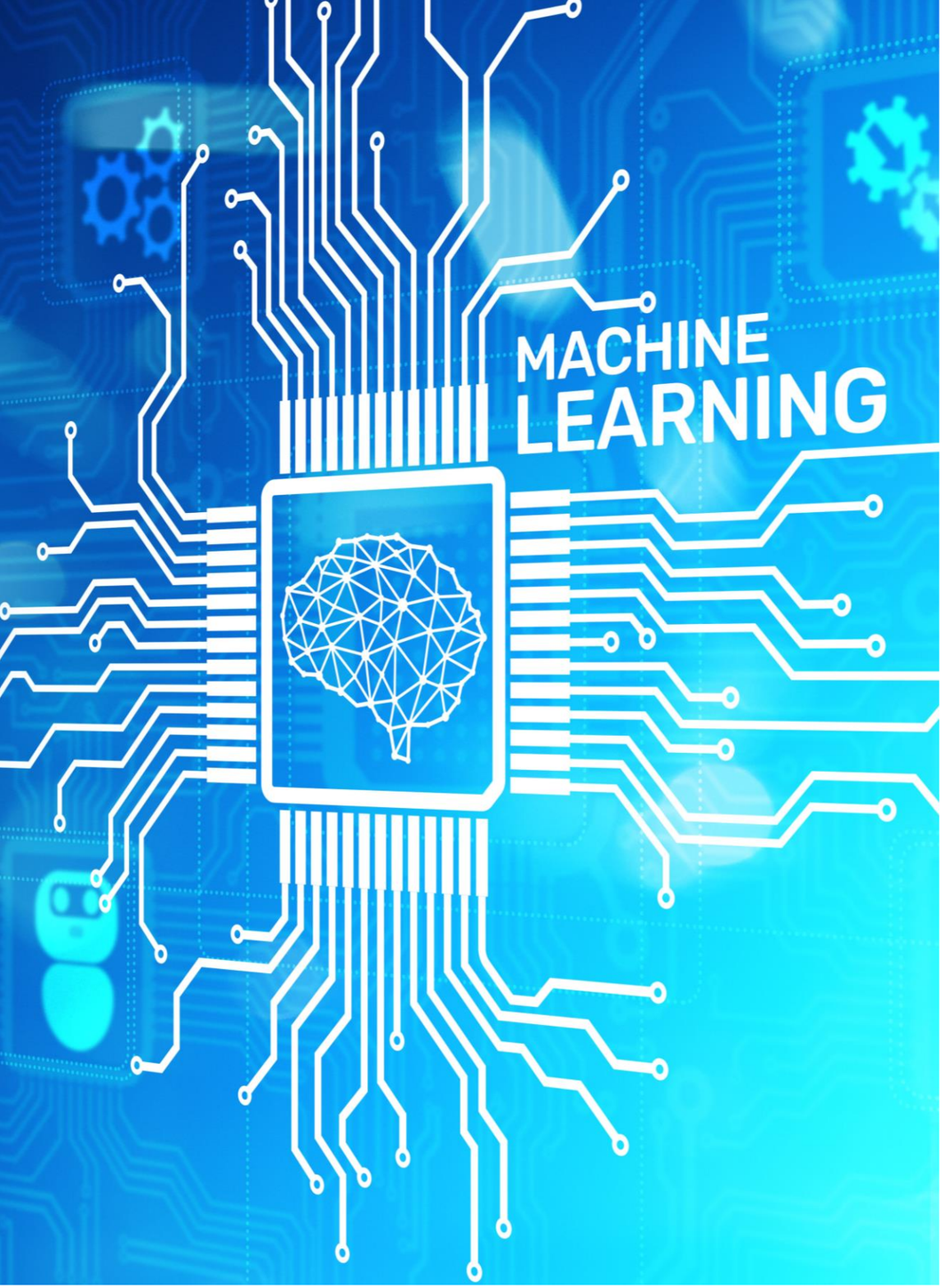
Movie Recommendation System

Self-Driving Cars

Facial Recognition System



https://www.lystloc.com/blog/7-applications-of-facial-recognition-technology/



Machine Learning

Automatic extraction of valid, novel, useful, and comprehensible knowledge (rules, regularities, patterns, constraints, models, …) from arbitrary sets of data

“Field of study that gives computers the ability to learn without being explicitly programmed” –Arthur Samuel

Types of Machine Learning

Supervised Learning

Algorithms trained on labeled data, learning to map inputs to known outputs.

Unsupervised Learning

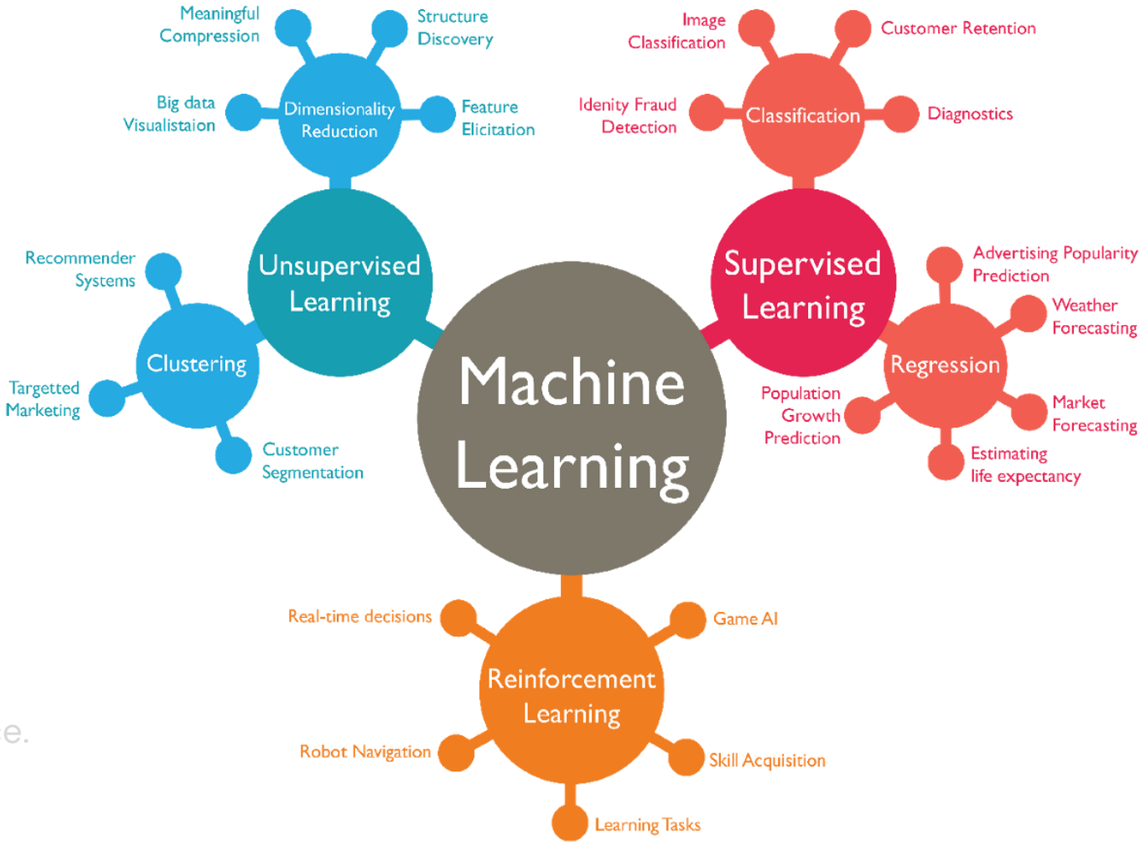
Discovers hidden patterns in unlabeled data without human guidance.

Semi-Supervised Learning

Leverages both labeled and unlabeled data to improve learning efficiency.

Reinforcement Learning

Learns optimal actions through trial and error with a reward system.



Types of Machine Learning

Supervised Learning

Algorithms trained on labeled data, learning to map inputs to known outputs.

Unsupervised Learning

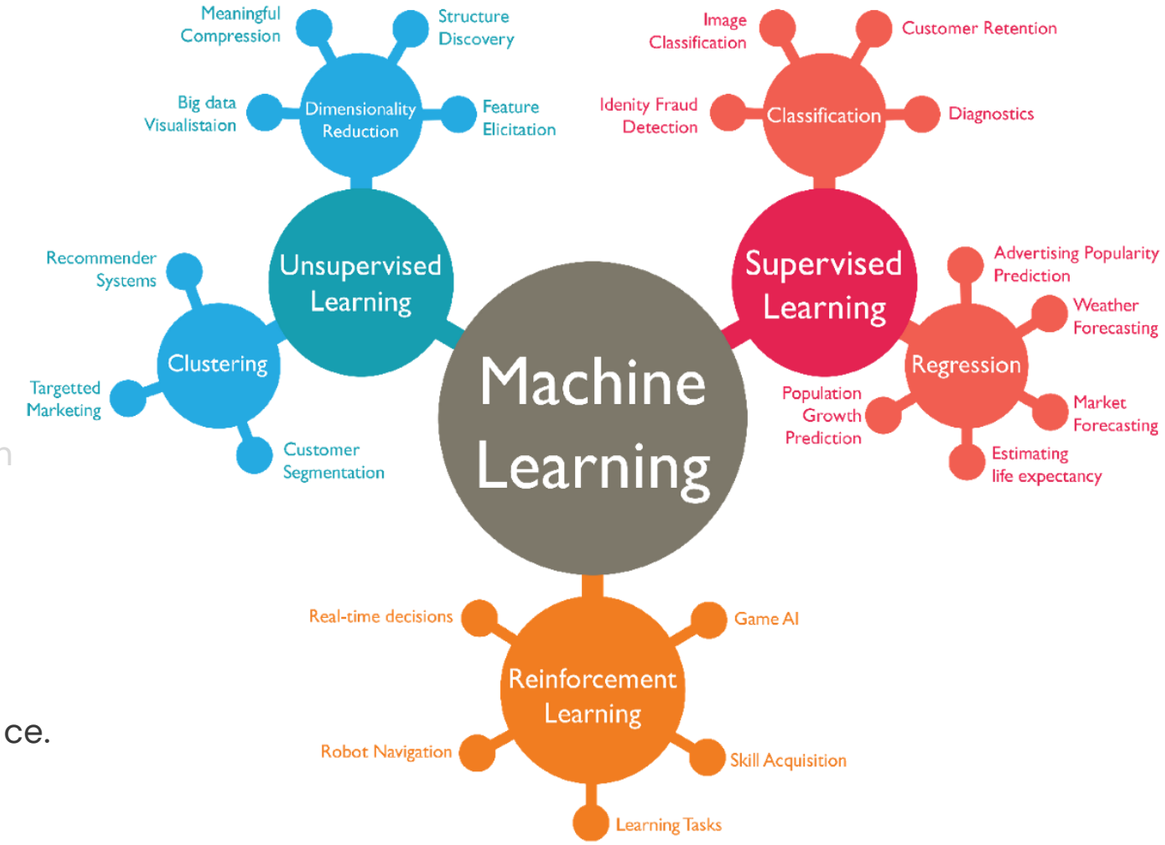
Discovers hidden patterns in unlabeled data without human guidance.

Semi-Supervised Learning

Leverages both labeled and unlabeled data to improve learning efficiency.

Reinforcement Learning

Learns optimal actions through trial and error with a reward system.



Types of Machine Learning

Supervised Learning

Algorithms trained on labeled data, learning to map inputs to known outputs.

Unsupervised Learning

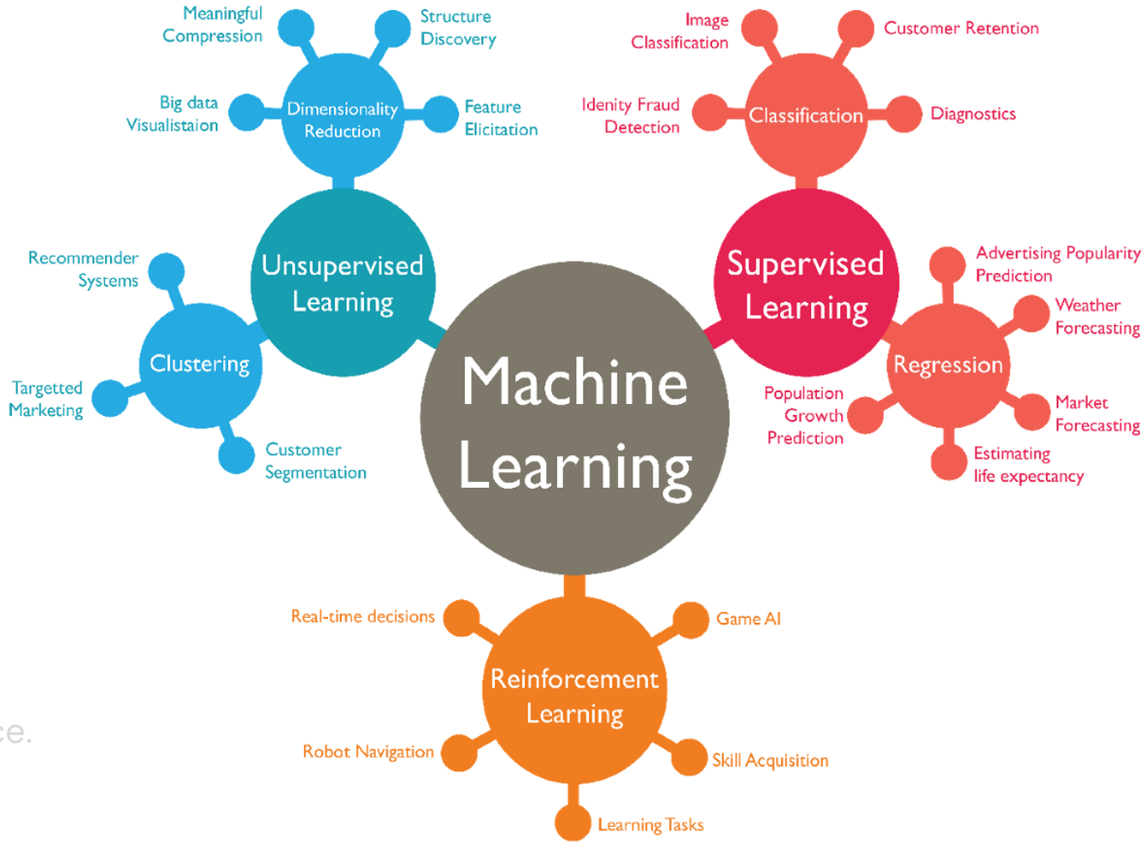
Discovers hidden patterns in unlabeled data without human guidance.

Semi-Supervised Learning

Leverages both labeled and unlabeled data to improve learning efficiency.

Reinforcement Learning

Learns optimal actions through trial and error with a reward system.



Types of Machine Learning

Supervised Learning

Algorithms trained on labeled data, learning to map inputs to known outputs.

Unsupervised Learning

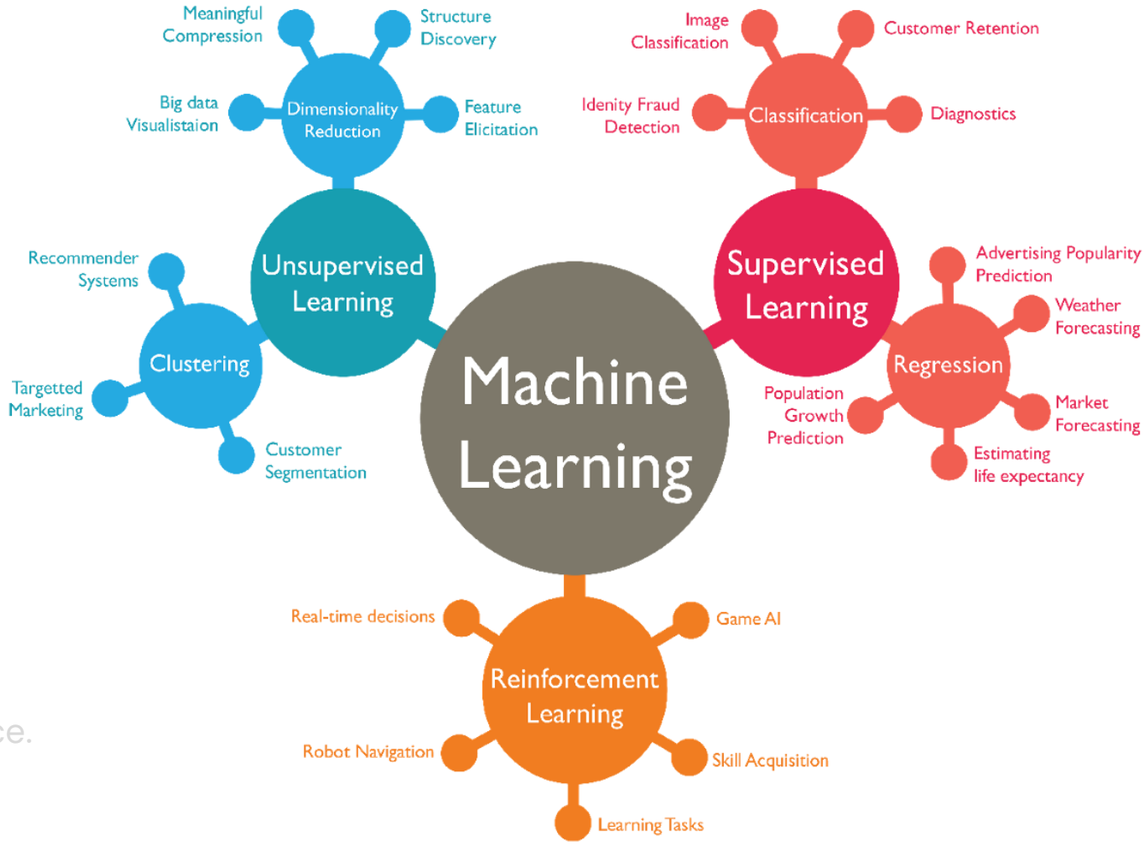
Discovers hidden patterns in unlabeled data without human guidance.

Semi-Supervised Learning

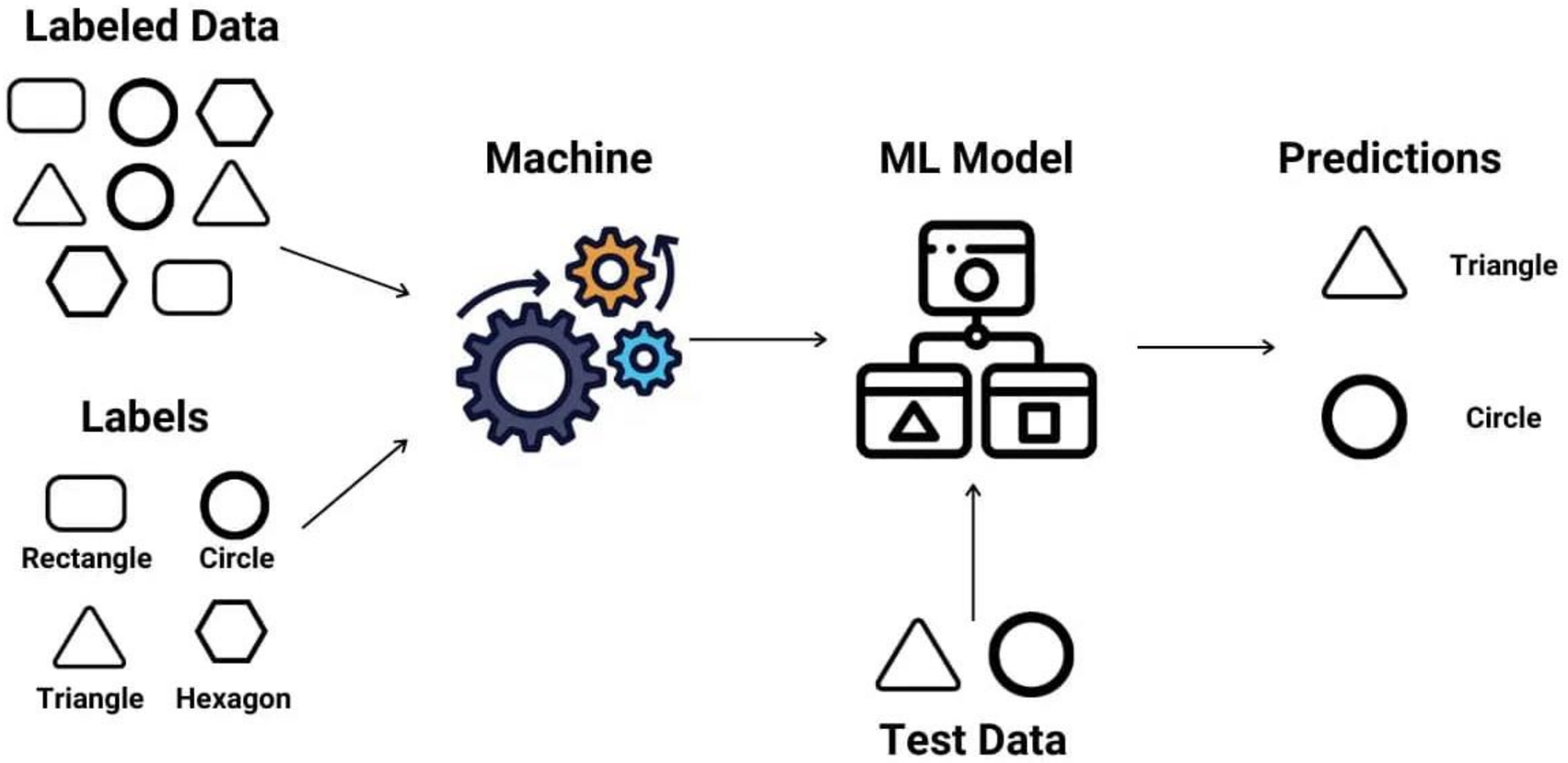
Leverages both labeled and unlabeled data to improve learning efficiency.

Reinforcement Learning

Learns optimal actions through trial and error with a reward system.



Supervised Learning Algorithms

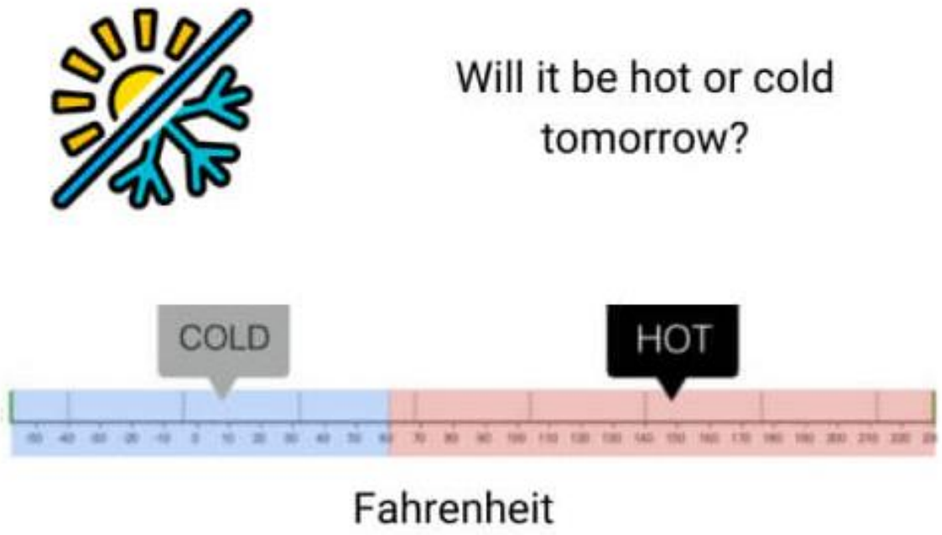


https://www.pipedrive.com/en/blog/types-of-machine-learning

Supervised Learning Algorithms

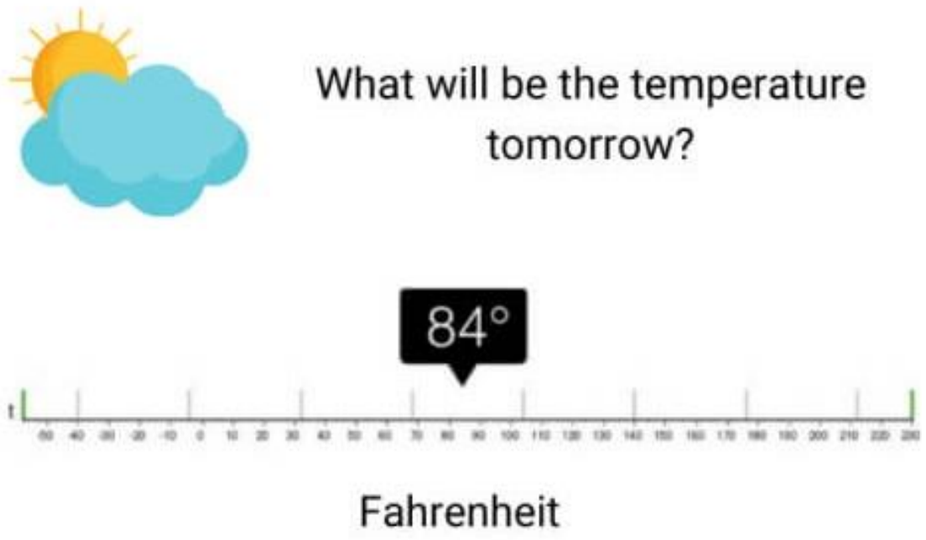
Classification

Categorizes data into predefined classes.



Regression

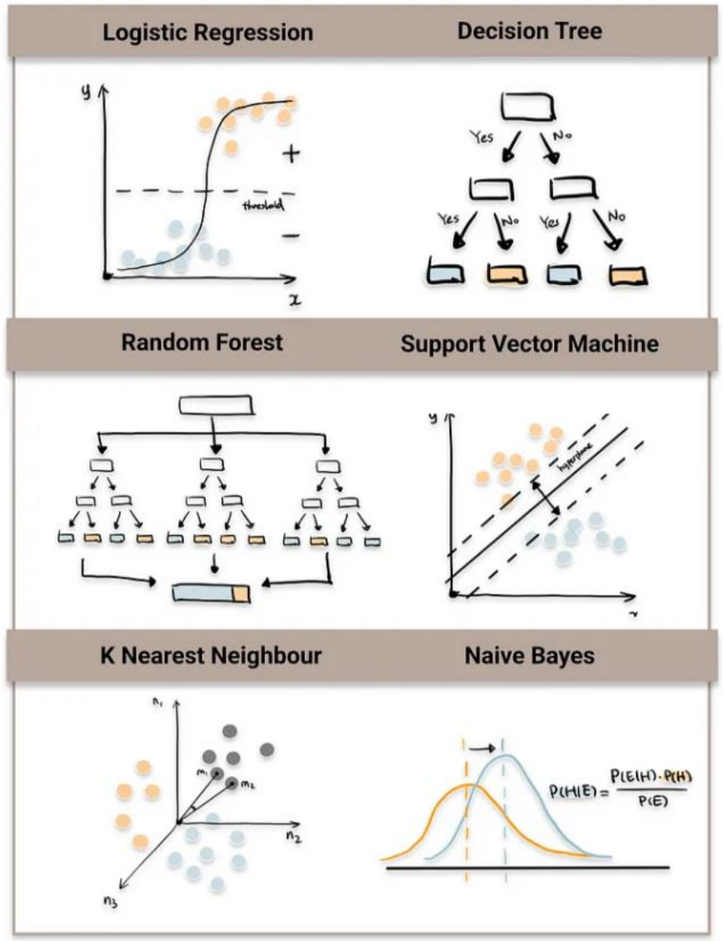
Predicts continuous numerical values.



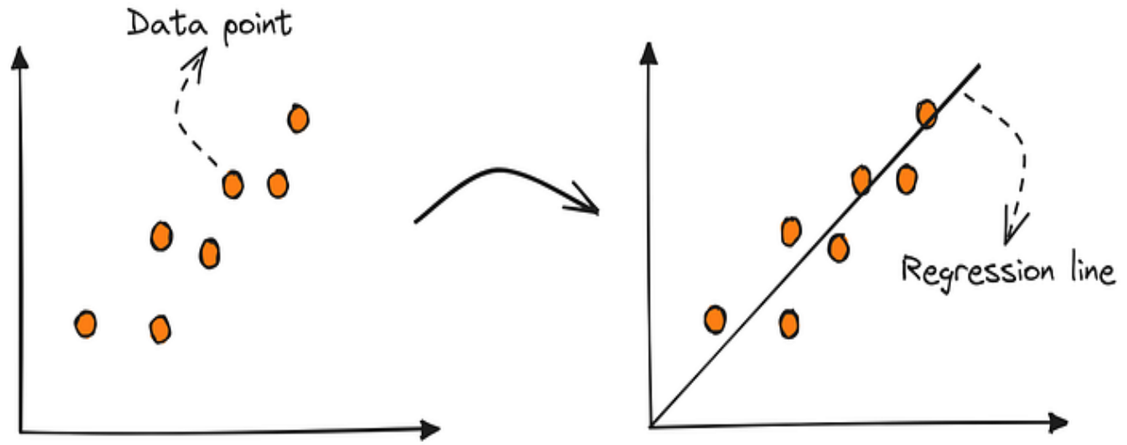
https://www.superannotate.com/blog/supervised-learning-and-other-machine-learning-tasks

Supervised Learning Algorithms

Classification



Regression



Linear Regression

https://pub.towardsai.net/machine-learning-the-first-step-is-to-understand-simple-linear-regression-30d

https://towardsdatascience.com/top-machine-learning-algorithms-for-classification-2197870ff501/

Real-World Applications of Supervised Learning

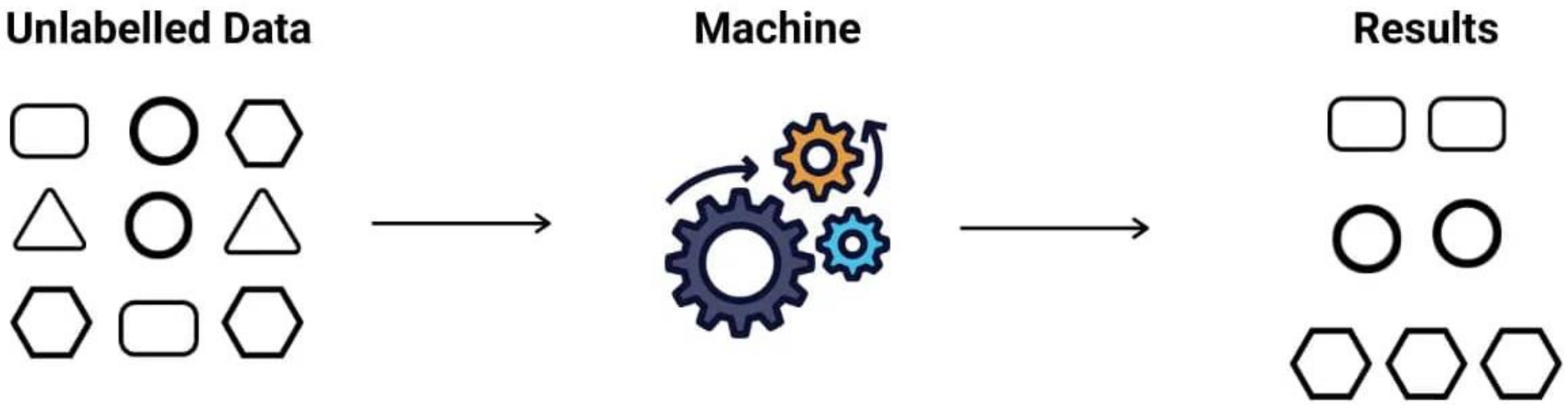
Email Spam Filtering

ML algorithms analyze message content and sender patterns to identify unwanted emails.

Medical Diagnostics

ML helps detect diseases in medical images with remarkable accuracy.

Unsupervised Learning Algorithms



https://www.pipedrive.com/en/blog/types-of-machine-learning

Unsupervised Learning Algorithms

Clustering

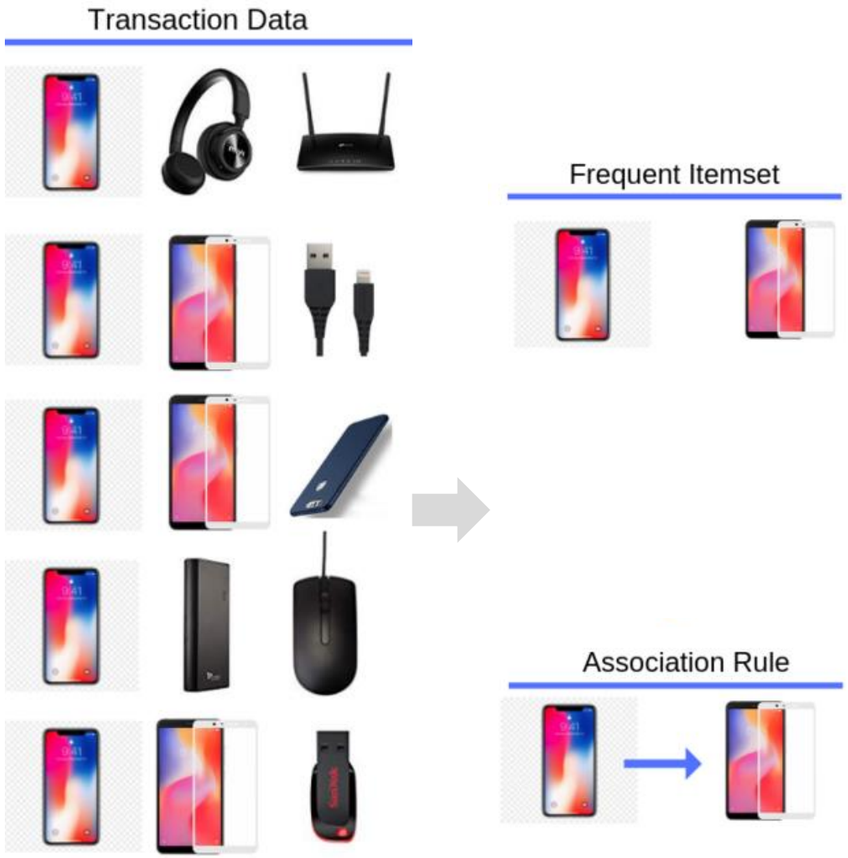
Groups similar data points together. K-means divides data into K clusters based on similarity. Businesses use this for customer segmentation.

Dimensionality Reduction

Reduces the number of variables while preserving important information. Principal Component Analysis (PCA) transforms data to fewer dimensions.

Anomaly Detection

Identify rare or unusual patterns in data that do not conform to expected behavior. It is commonly applied in fields like fraud detection.



Association Rule Learning in Machine Learning

Association learning algorithms discover relationships between variables in large datasets. They identify patterns that might otherwise remain hidden.

The classic example is market basket analysis, where retailers learn that customers who buy phone often purchase screen card simultaneously.

https://www.r-bloggers.com/2019/05/practical-introduction-to-market-basket-analysis-asociation-rules-2/

Key Machine Learning Terminology

Instance

A single example in your dataset. Each row in a table represents one instance, like an individual customer record.

Attribute

A feature or characteristic of an instance. For a house, attributes might include square footage, location, and age.

Concept

The target parameter to be learned. It defines the relationship between inputs and the desired output.

Training

Collection of example data used to teach the model patterns and relationships.

Testing

Separate data used to evaluate model performance on unseen examples.

Nominal

Variable can take multiple values which are discrete types or categories

Ordinal

Variable has discrete values, and they have a natural order

Continuous

Variable is real-valued with a defined zero point
and no explicit bound

Example: weather dataset

Outlook	Temperature	Humidity	Windy	Play?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
…	…	…	…	…

Example: weather dataset

Outlook	Temperature	Humidity	Windy	Play?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
…	…	…	…	…

Instance

Example: weather dataset

Outlook	Temperature	Humidity	Windy	Play?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
…	…	…	…	…

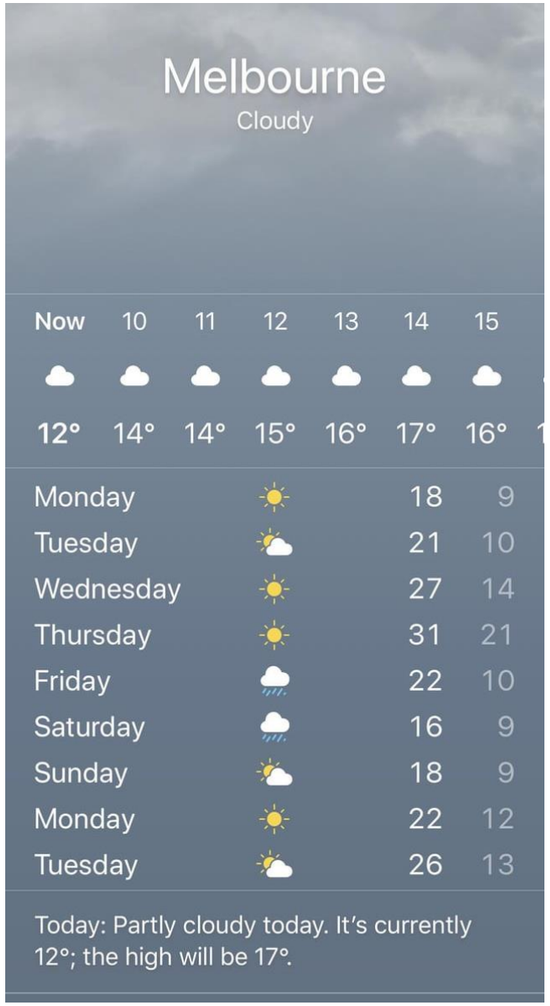
Attribute

Example: weather dataset

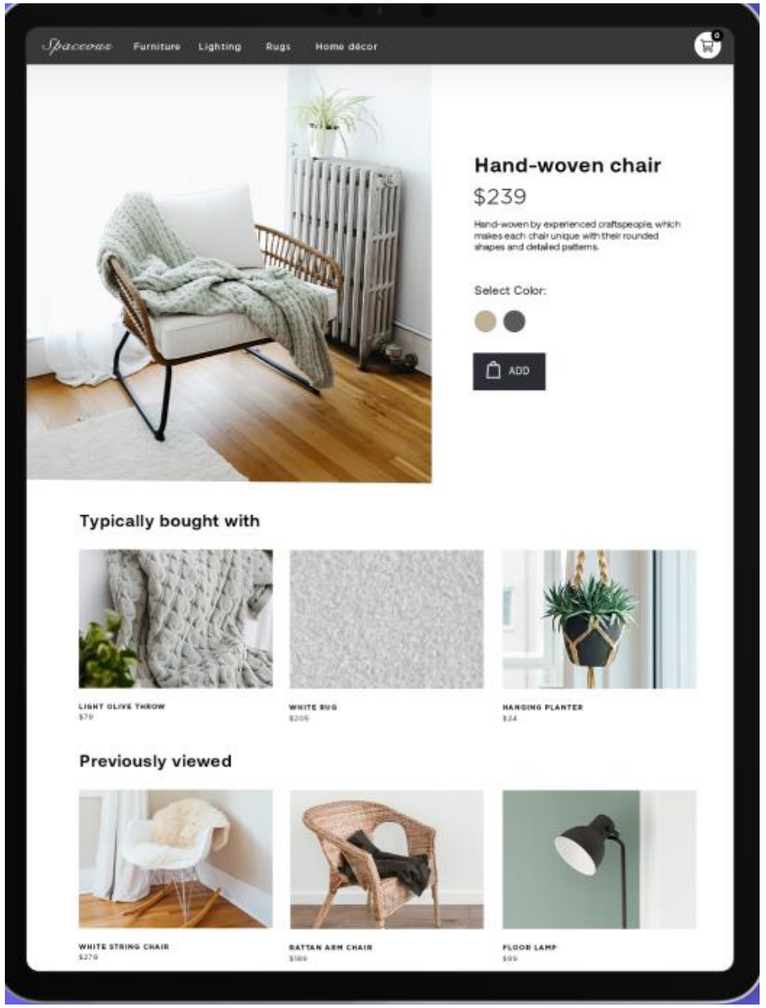
Outlook	Temperature	Humidity	Windy	Play?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
…	…	…	…	…

Concept

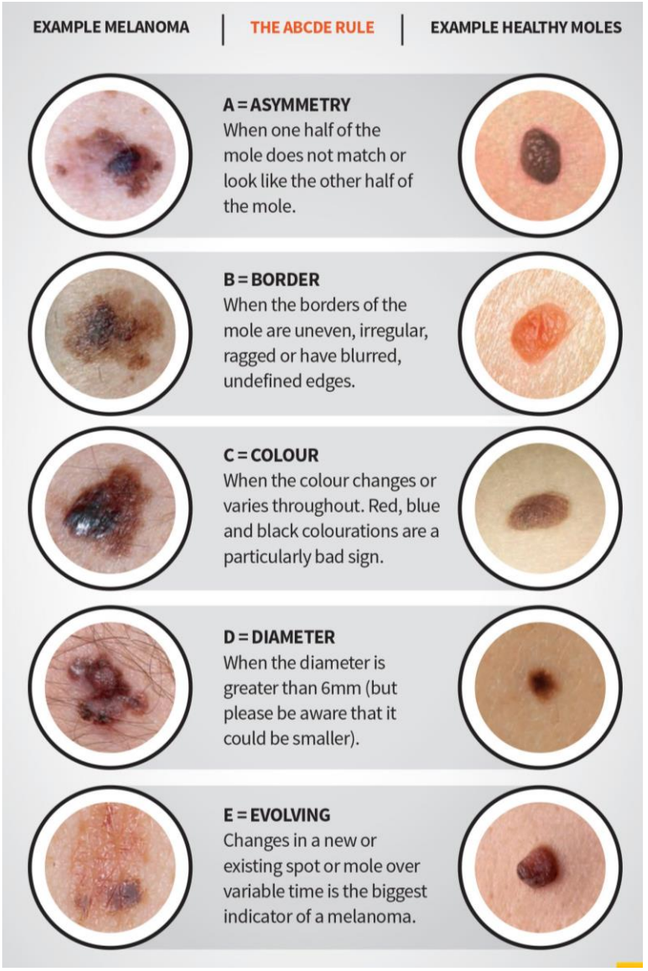
Real-World Applications of Machine Learning



Weather Prediction



Product Recommendation



Skin Cancer

Concepts, Instances, and Attributes in Real-World Problems

Problem	Concept	Instances	Attributes
Skin Cancer Screening	Presence of cancer or Type of cancer, i.e. (benign/malignant)	Patient (patient records) (Skin lesions or skin images)	Texture, shape, color, size, asymmetry, age, gender, family history, UV exposure, biopsy data
Weather Prediction	Temperature or Amount of rain or The UV index or any other weather feature	Day	Temperature, humidity, wind speed, pressure, precipitation, cloud cover, time
Product Recommendation	Interested or not	Customer product pairing Customer	name, age, address, gender, shopping log, credit card information, loyalty card information

Understanding these four elements is essential for effective machine learning implementation. The problem defines our goal, instances represent our data points, attributes provide the features for pattern identification, and the concept defines the o/p to be predicted.

Machine Learning Models

Problem	Likely ML Approach	Suitable Models
Skin Cancer Screening	Binary Classification	CNNs (for images), Logistic regression, SVM, Random Forests
Weather Prediction	Regression (temperature: 27℃) Classification (“Sunny”, “Rainy”)	Linear Regression, Random Forests, Logistic Regression
Product Recommendation	Clustering Association Rule Mining	K-means Clustering

Generalization in Machine Learning

Problem	Generalization Difficulty	Reasoning
Skin Cancer Screening	Difficult	Age biases
Weather Prediction	Easy	Add geographic information (longitude, altitude, distance from ocean)
Product Recommendation	Difficult	Country based holiday purchasing

Generalization is the ultimate goal of machine learning. A model that generalizes well can make accurate predictions on data it hasn't seen before.

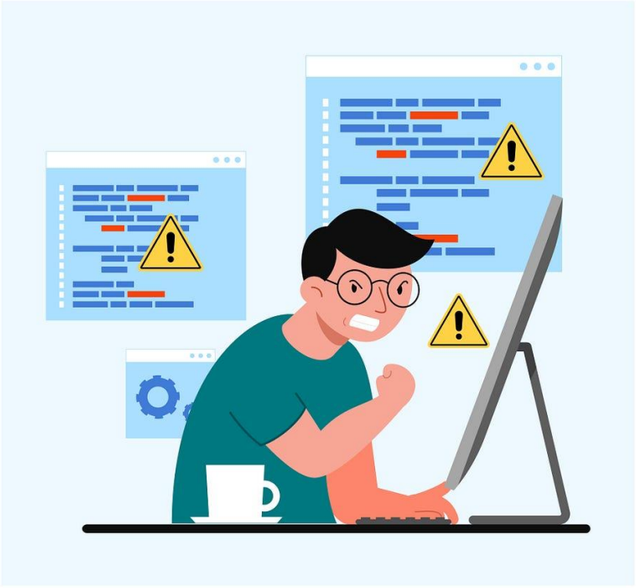
Common Assumptions in Machine Learning

1. Concept is actually related to the attributes

2. Feature Independence

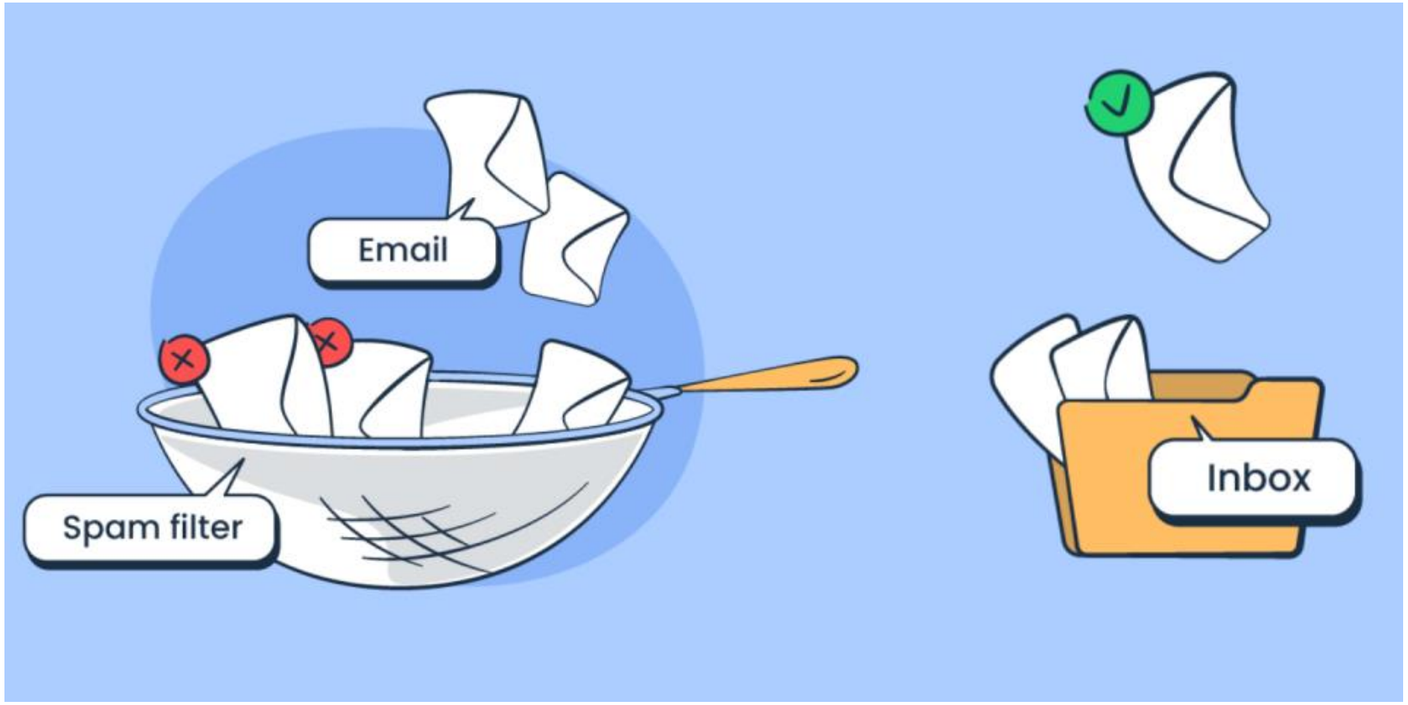
   Variables don't influence each other. Correlation between features can impact model performance.

What about Fraud Detection?



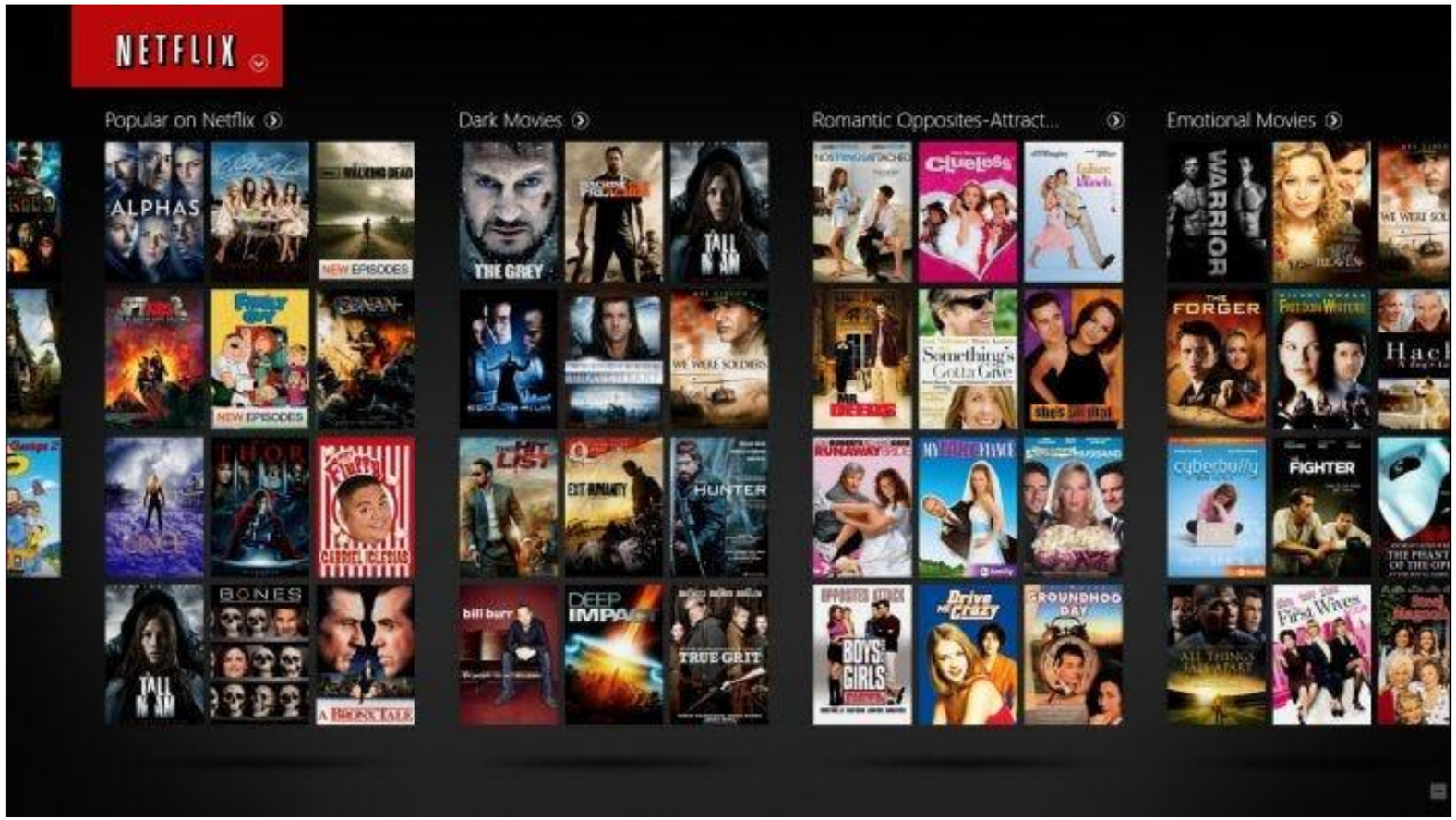
Instance → **Attributes** → **Concept** → **Likely ML Approach**

What about Email Spam Filtering?



Instance | **Attributes** | **Concept** | **Likely ML Approach**

What about Movie Recommendation System?



Instance → **Attributes** → **Concept** → **Likely ML Approach**

https://www.theodysseyonline.com/10-netflix-movies-watch-stressed


THANK
YOU