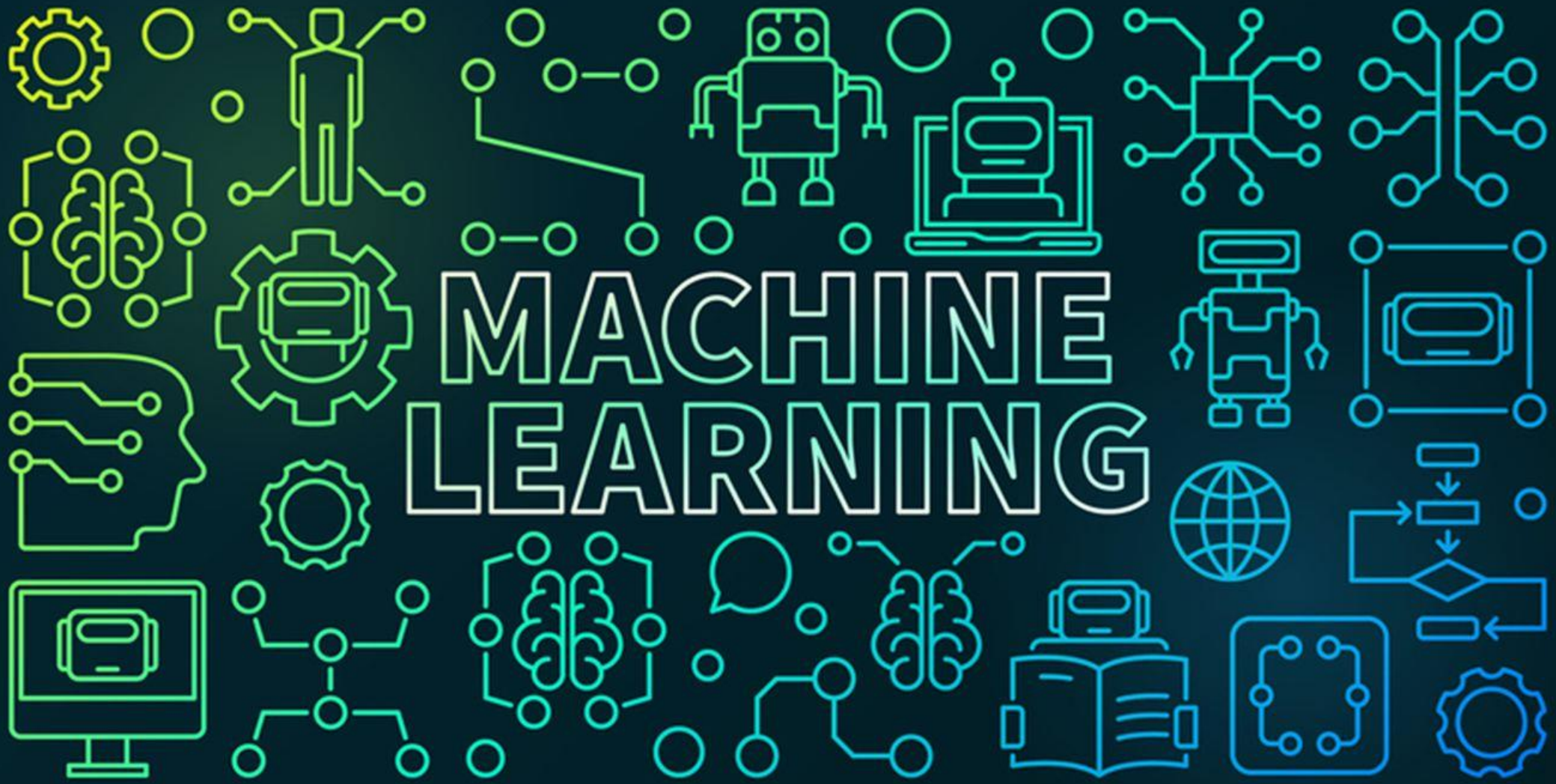
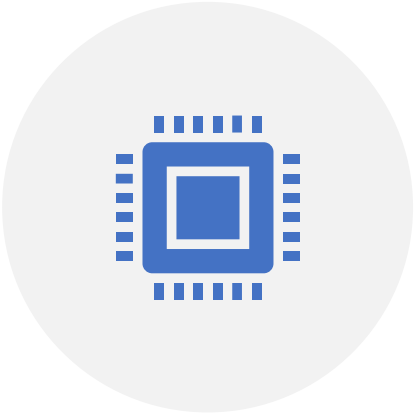

MACHINE
LEARNING

About me

NAME: PRANAV GHATIGAR

STUDYING AT DAYANANDA SAGAR UNIVERSITY



COMPUTER SCIENCE AND ENGINEERING ’24

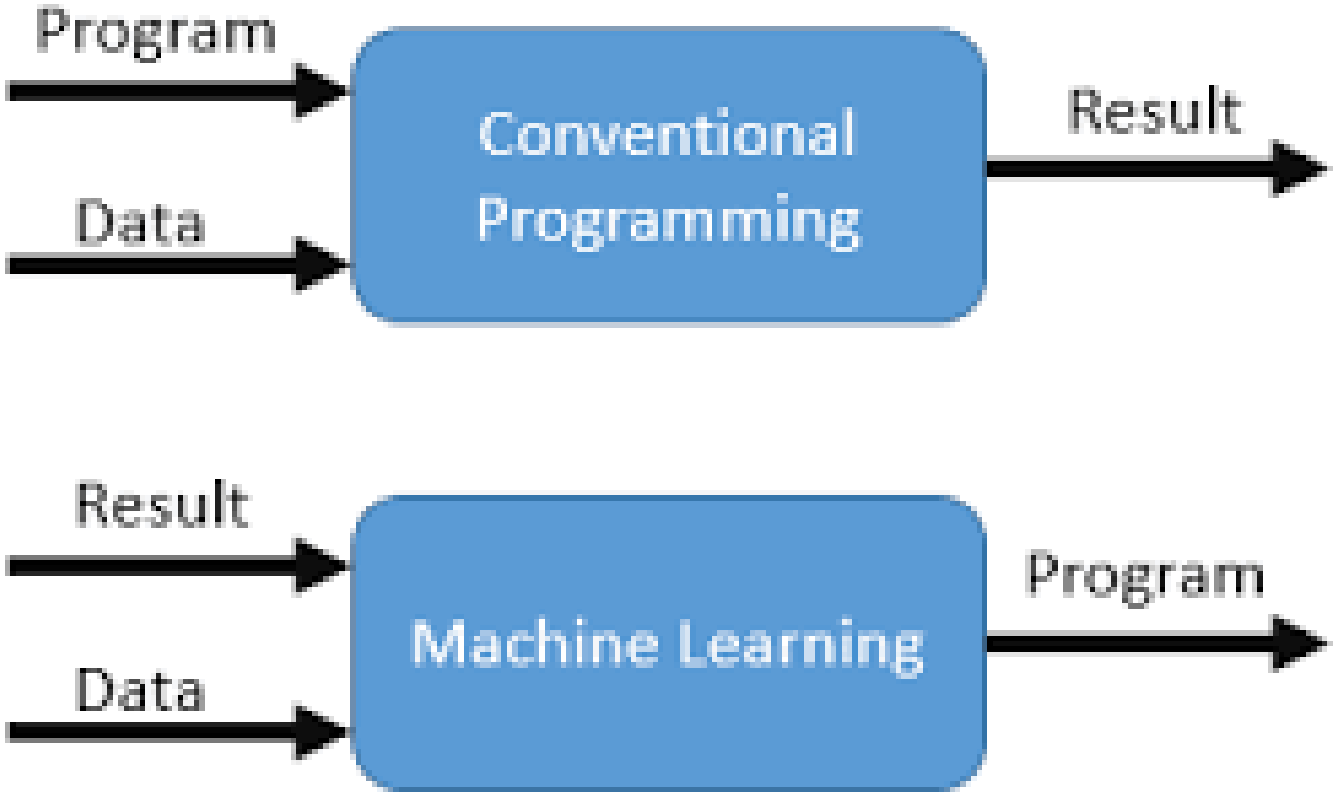
Content

- Introduction to Machine learning
- Supervised, unsupervised and Reinforcement Learning
- Data warehouse and data mining
- Implementation of the following Algorithms
  - Linear Regression
  - Logistic Regression
  - Decision Tree
  - Random Forest
  - KNN
  - SVM
- Applications Of Machine Learning

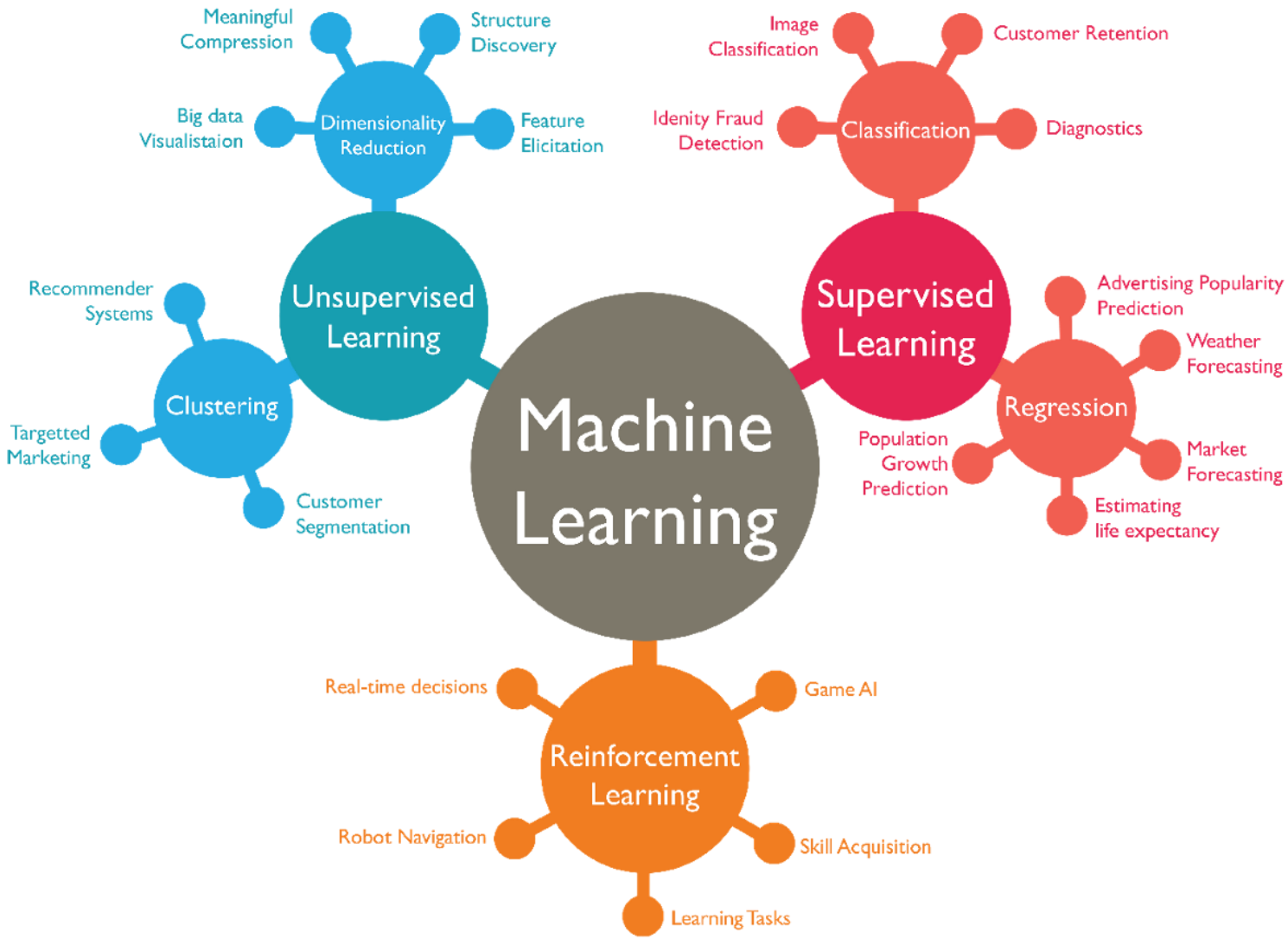
What is Machine Learning

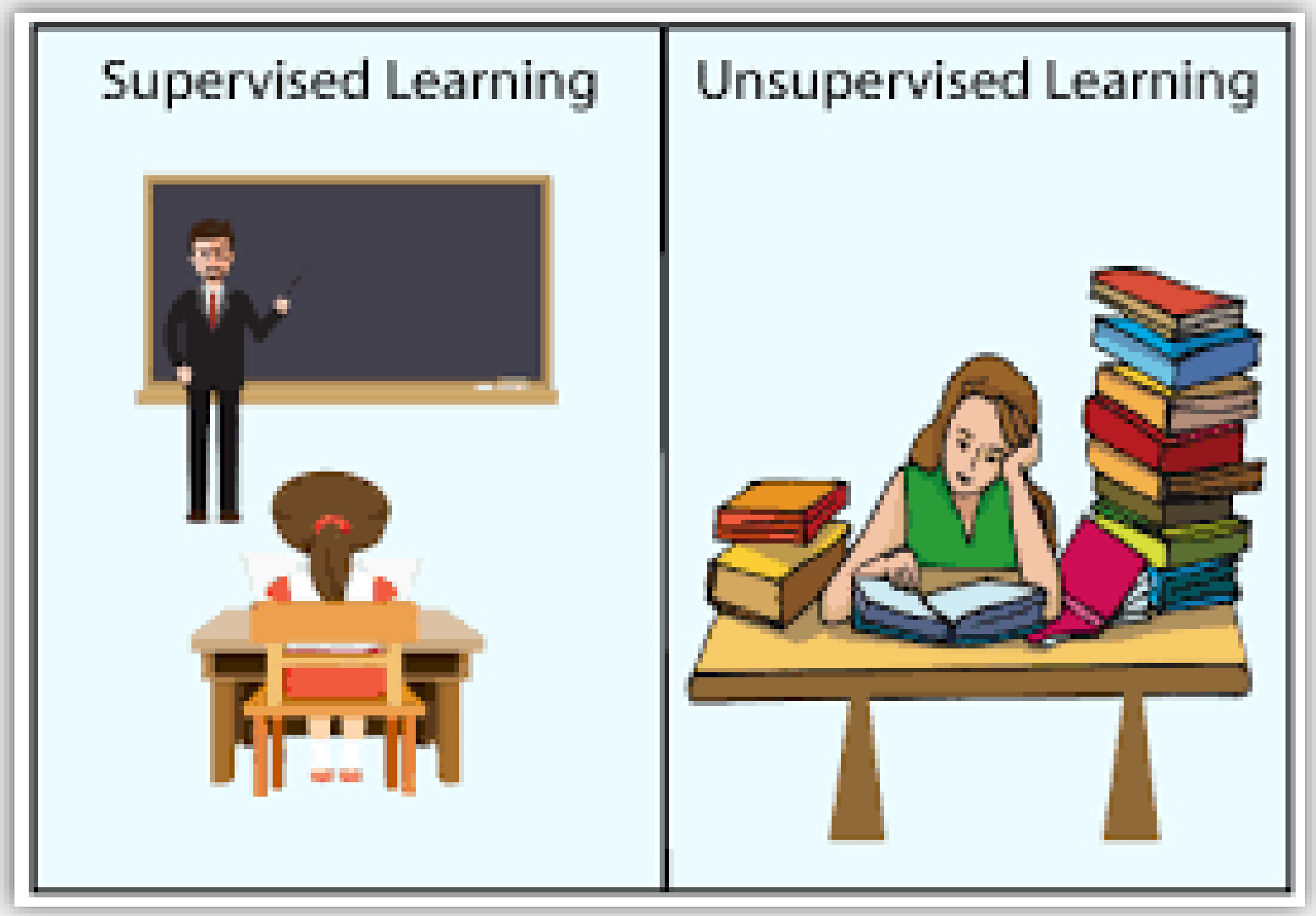
- It is the study of computer algorithms that can improve automatically through experience and by the use of data.
- Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to learn without explicit programming.

VS Traditional Programming

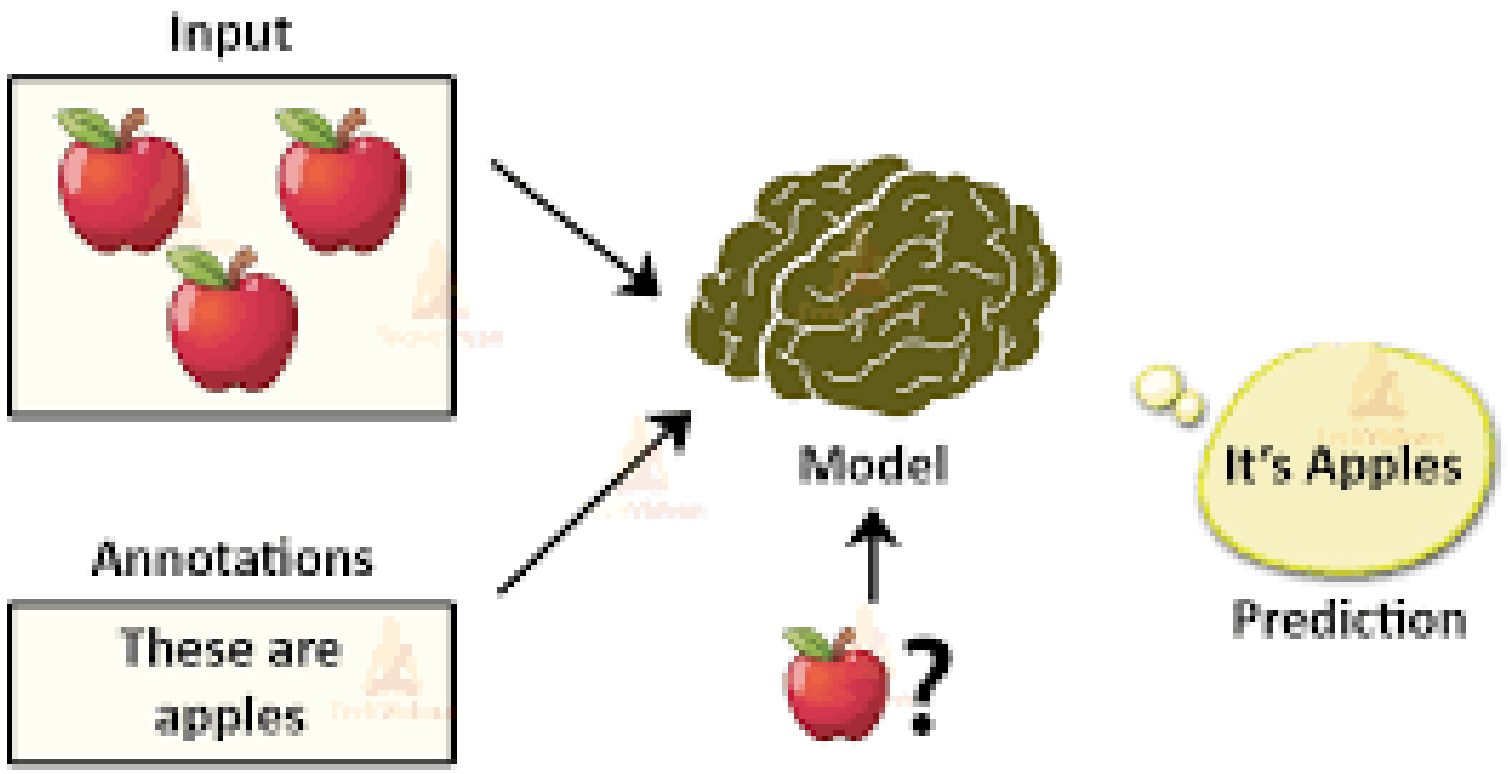


Types

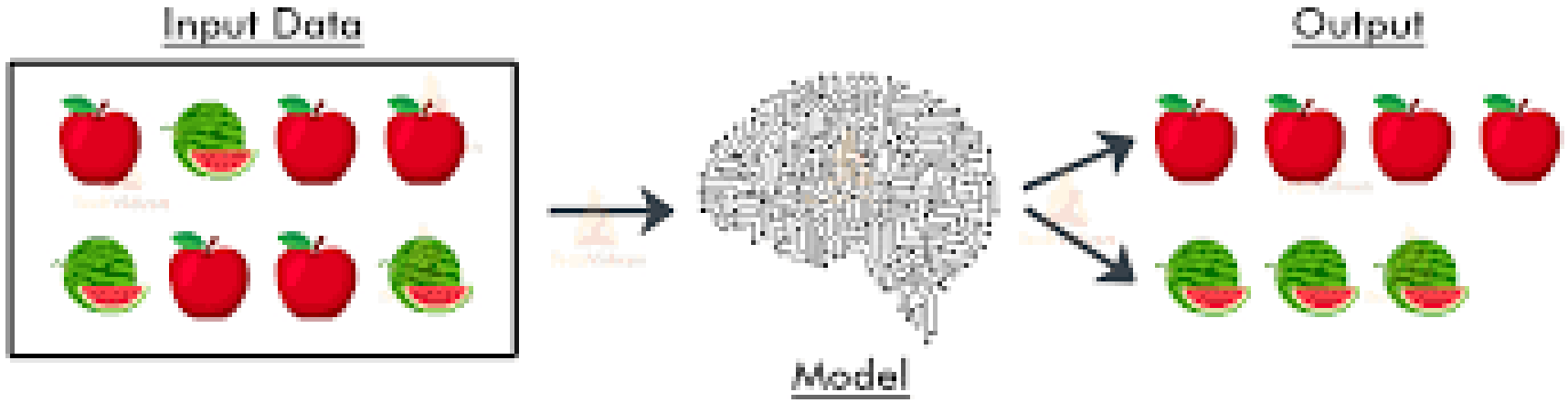

Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Feature Elicitation
Big data Visualistaion
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating life expectancy
Population Growth Prediction
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

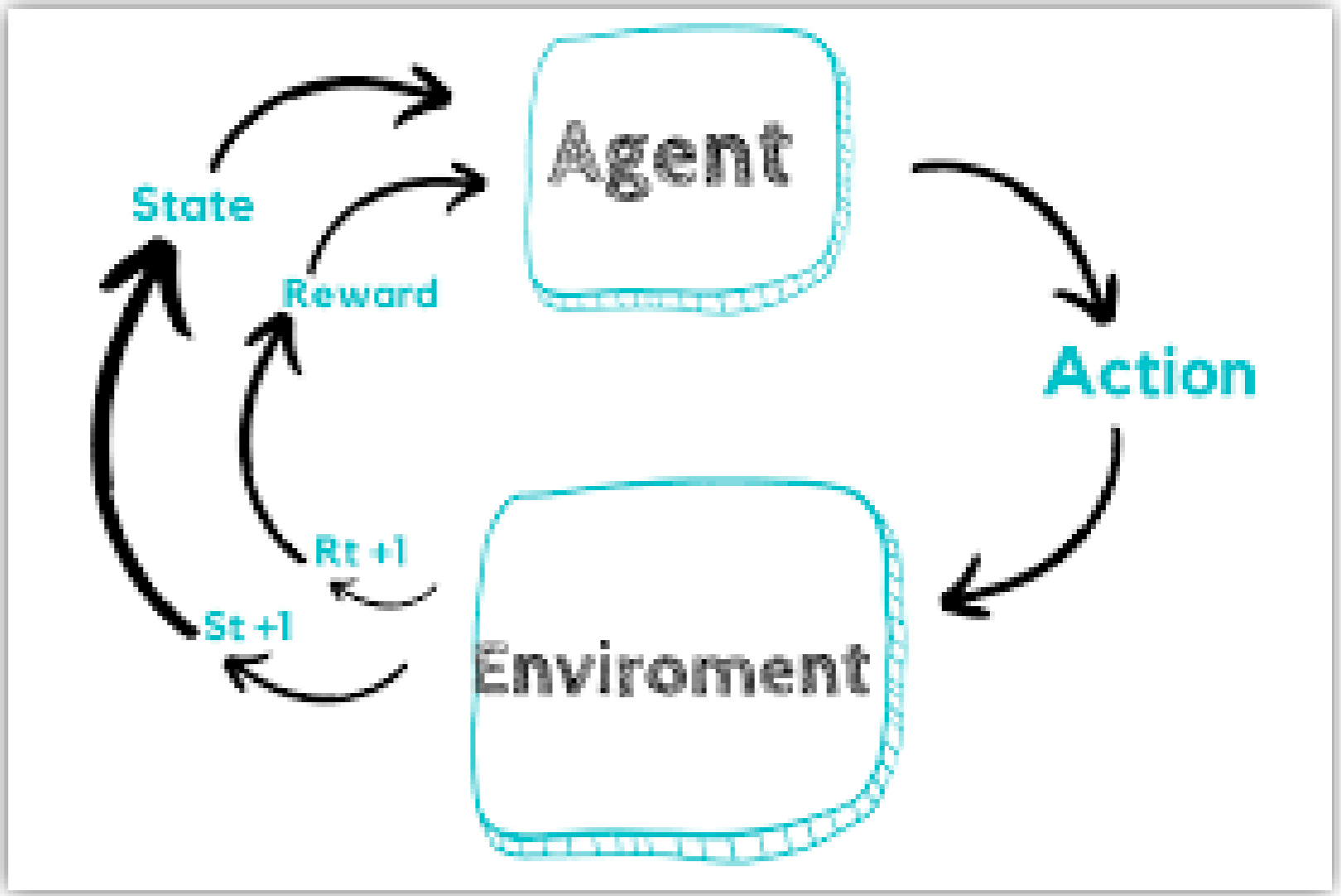

Supervised Learning
Unsupervised Learning

Supervised Learning in ML



Unsupervised Learning in ML





Reinforcement Learning

Definitions

- Supervised Learning – A type of learning where features and labels are given to the ML algorithm and prediction takes place.
- Unsupervised Learning – A type of learning where only features and given to the ML algorithm and prediction takes place.
- Reinforcement Learning – A type of learning which is about taking the right actions in order to maximize cumulative reward.

What do you want the machine learning system to do?

I want to see if there are natural clusters or dimensions in the data I have about different situations.

I want to learn what actions to take in different situations.

Do you want the ML system to be active or passive?

ACTIVE

The system's own actions will affect the situations it sees in the future.

PASSIVE

The system will learn from data I give it.

Do you have access to data that describes a lot of examples of situations and appropriate actions for each situation?

- Yes → Could a knowledgeable human decide what actions to take based on the data you have about the situation?
  - Yes → **MACHINE LEARNING IS NOT USEFUL**
  - No → Will the system be able to gather a lot of data by trying sequences of actions in many different situations and seeing the results?
    - No → **MACHINE LEARNING IS NOT USEFUL**
    - Yes → **REINFORCEMENT LEARNING MAY BE APPROPRIATE**
- No → Could there be patterns in these situations that humans haven't recognized before?
  - Yes → **SUPERVISED LEARNING MAY BE APPROPRIATE**

neural nets
support vector machines
regression
recommender systems

  - No → **MACHINE LEARNING IS NOT USEFUL**

UNSUPERVISED LEARNING MAY BE APPROPRIATE

clustering
anomaly detection

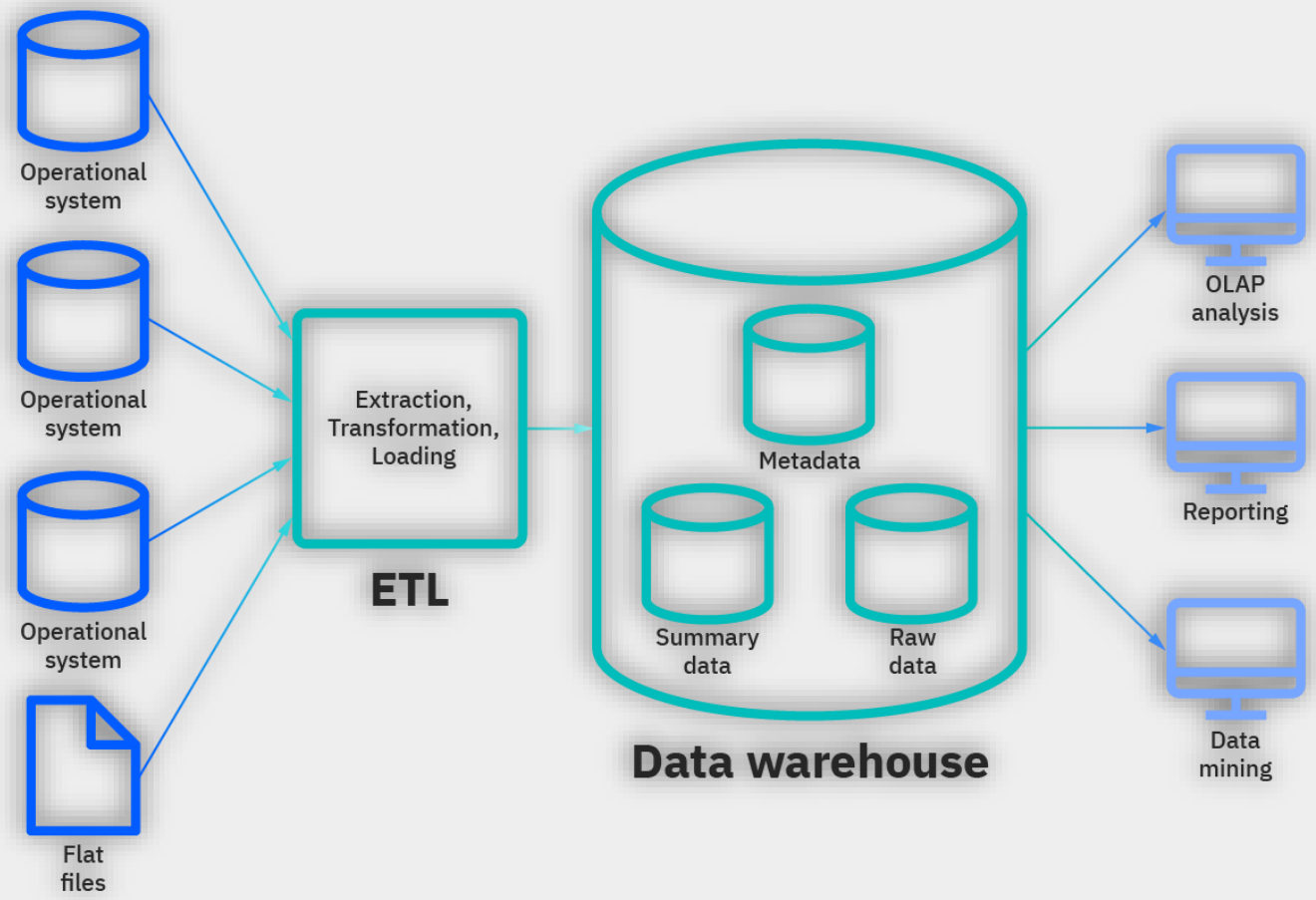
Credit: Thomas Malone, MIT Sloan | Design: Laura Wentzel

Data Warehouse

- Data warehousing is the process of constructing and using a data warehouse.
- A data warehouse is constructed by collecting data from multiple sources which can be used to
  - Analyse reports
  - Structured queries
  - Decision making

Data Mining

- It is the process of extracting data from data warehouse to discover hidden patterns in large data set using methods such as machine learning, statistics and DBMS.
- Some tools such as, Xplenty, IBM Cognos, Apache Kylin are used.


Operational system
Operational system
Operational system
Flat files
Extraction, Transformation, Loading
ETL
Metadata
Summary data
Raw data
Data warehouse
OLAP analysis
Reporting
Data mining

ML algorithms

- Linear Regression
- Logistic Regression
- Decision Tree
- Random Forest
- KNN
- SVM

Linear Regression

- Used to predict the value of a variable based on the value of previous continuous data.
- Sample code:

```
from sklearn import linear_model

x = [[4], [5], [6], [7], [8], [9]]

y = [12, 15, 18, 21, 24, 27]

classifier = linear_model.LinearRegression()

classifier.fit(x, y)

X_marks = [[13]]

print(classifier.predict(X_marks))
```

```
print(classifier.predict(X_marks))

[39.]
```

Logistic Regression

- Logistic regression is a process of modeling the probability of a discrete outcome given an input variable.
- That is pass or fail, true or false, cat or dog, etc.
- Sample code:

```
from sklearn.linear_model import LogisticRegression
x = [[30], [40], [50], [60], [20], [10], [70]]
y = [0, 1, 1, 1, 0, 0, 1]
classifier = LogisticRegression()
classifier.fit(x,y)
X_marks = [[50]]
print(classifier.predict(X_marks))
```

```
X_marks = [[50]]
print(classifier.predict(X_marks))

[1]
```

Decision tree

- It has a tree like structure which builds the best attribute as root, then splits the dataset into subsets.
- Usually used to predict class or value of target variable.
- Sample code:

```
from sklearn.tree import DecisionTreeClassifier

X = [[30], [40], [50], [60], [20], [10], [70]]

y = [0, 1, 1, 1, 0, 0, 1]

RandomForestRegModel = DecisionTreeClassifier(criterion='entropy',random_state=0)

RandomForestRegModel.fit(X, y)

print(RandomForestRegModel.predict([50]))
```

```
X_marks = [[50]]
print(RandomForestRegModel.predict(X_marks))
```

```
[1]
```

Random Forest

- It comes under supervised classification algorithm.
- Multiple number of decision trees together forms a random forest. It uses the rule of each randomly created decision tree and stores the predicted outcome and the most voted prediction is the final.
- Sample code:

```
from sklearn.ensemble import RandomForestRegressor
        X = [[30], [40], [50], [60], [20], [10], [70]]
        y = [0, 1, 1, 1, 0, 0, 1]
        RandomForestRegModel= RandomForestRegressor()
        RandomForestRegModel.fit(X, y)
        X_marks = [[50]]
        print(RandomForestRegModel.predict(X_marks))
```

```
X_marks = [[50]]
print(RandomForestRegModel.predict(X_marks))

[0.98]
```

KNN k-Nearest Neighbours

- Used for classification and regression.
- It simply compares the stored cases with the new data for majority of ‘k’- neighbours.
- Sample code:

```
from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

X = [[30], [40], [50], [60], [20], [10], [70]]

y = [0, 1, 1, 1, 0, 0, 1]

classifier = KNeighborsClassifier(n_neighbors=5,
                                  metric='minkowski', p=2)

classifier.fit(X, y)

X_marks = [[50]]

print(classifier.predict(X_marks))
```

```
X_marks = [[50]]
print(classifier.predict(X_marks))

[1]
```

SVM

- Stands for Support Vector Machine.
- It ais a binary classifier. It differentiates the to closest given data-points.
- Sample code:

```
from sklearn.svm import SVC
X = [[30], [40], [50], [60], [20], [10], [70]]
y = [0, 1, 1, 1, 0, 0, 1]
classifier = SVC(kernel='linear', random_state=0)
classifier.fit(X, y)
X_marks = [[55]]
print(classifier.predict(X_marks))
```

```
X_marks = [[50]]
print(classifier.predict(X_marks))

[1]
```

But why?

- Machine learning is important because of its wide range of applications and its incredible ability to adapt and provide solutions to complex problems efficiently, effectively and quickly.
- It can increase the value of your embedded analytics in many areas, such as natural language interfaces, automatic outlier detection, recommendation system and many more.
- All of these features help speed user insights and reduce decision bias.

Applications of ML

- They are used from day-to-day life
  - Speech Recognition:
    - Virtual Assistant – Siri, Alexa
    - Text-to-speech and vice versa
  - Computer Vision:
    - Face Detection
    - Hand Writing and fingerprint
  - Business Intelligence
    - Customer Support
    - Product Recommendation
    - Business Analytics

Applications of ML

- Banking:
  - Fraud Monitoring
  - Data Security
  - Financial Trading
- Health Care:
  - Disease Identification and Diagnosis
  - Robotic Surgery
  - Wearable Tech