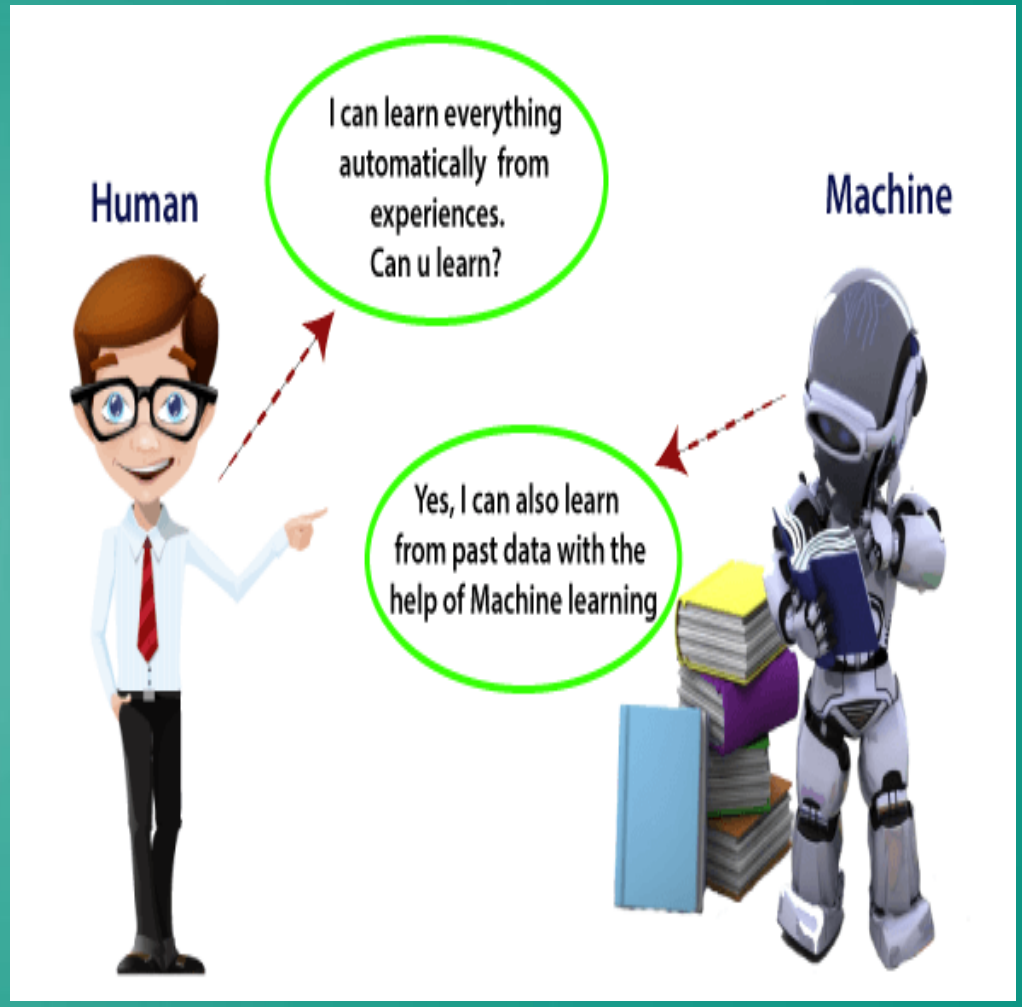
ASSIGNMENT –I
DEEP LEARNING

TEJAAL M
ENG19CS0334
SEM V
F section



DAYANANDA SAGAR UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF ENGINEERING
Dayananda Sagar University
Kudlu Gate
Bangalore – 560068

REPORT
ON
"MACHINE LEARNING"

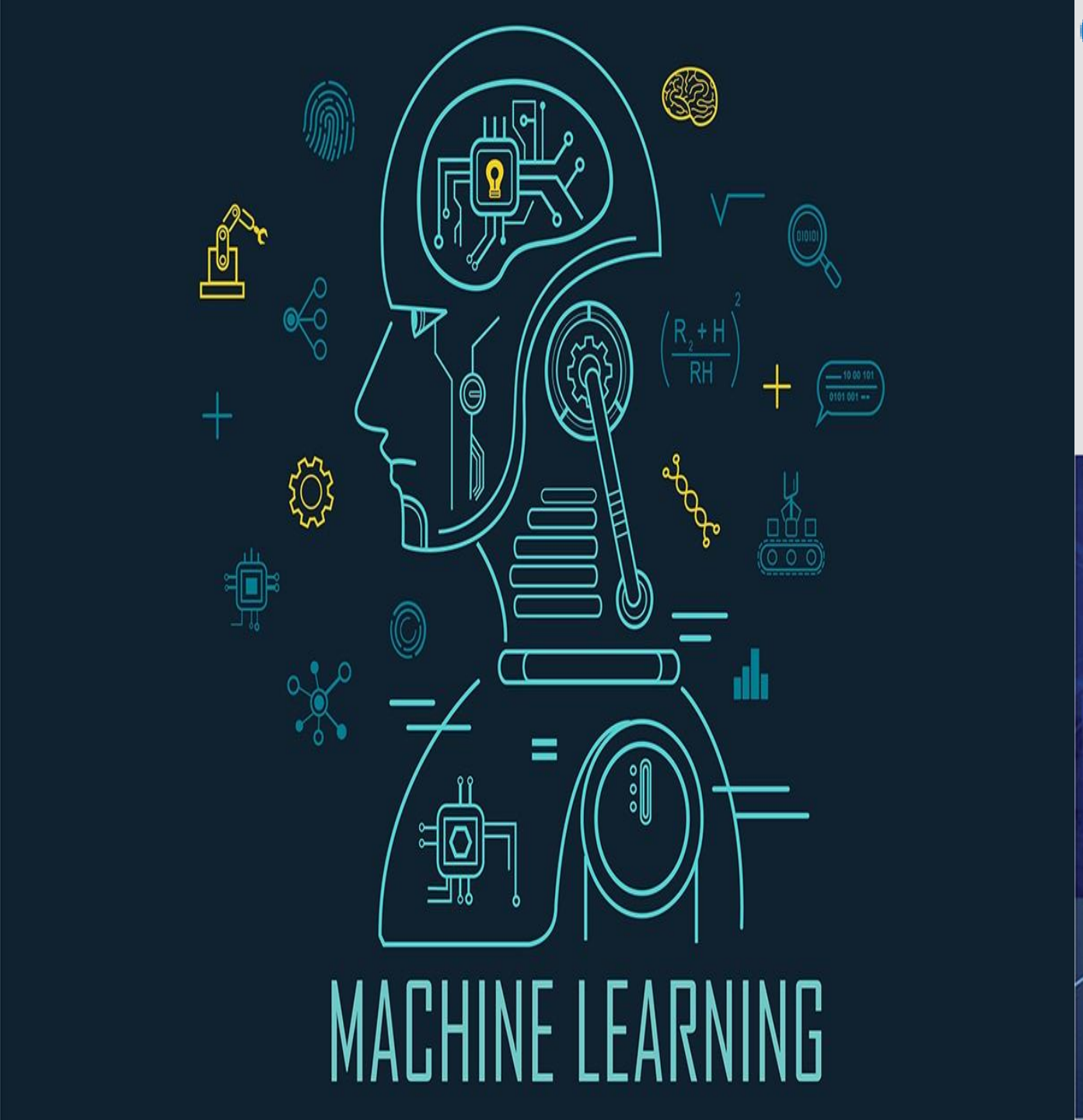
SUBMITTED TO THE 5TH SEMESTER DEEP LEARNING COURSE
BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

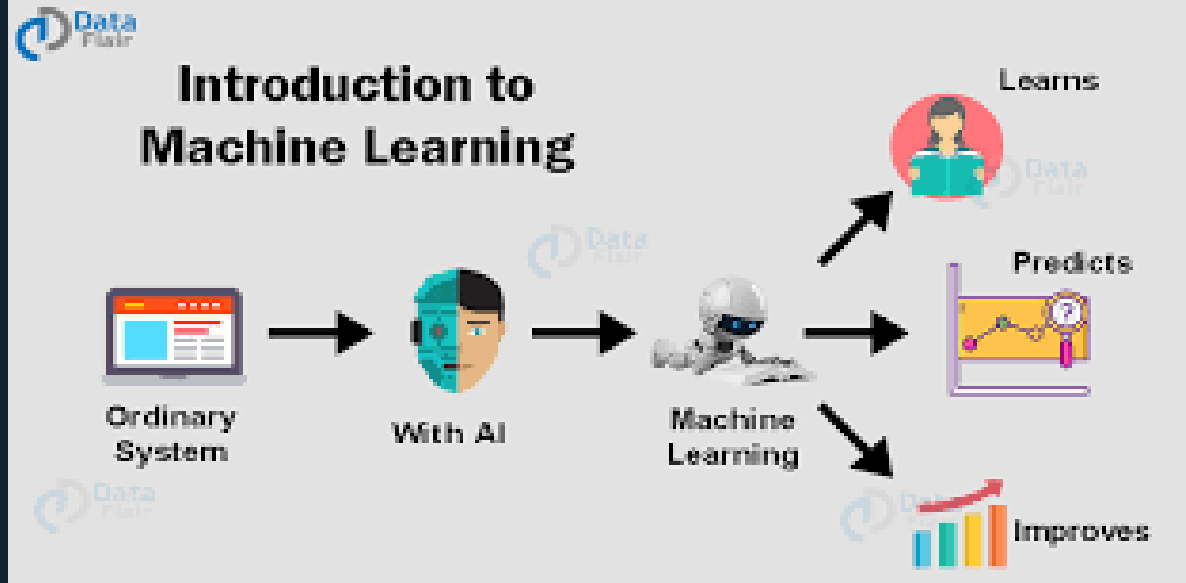
SUBMITTED BY
TEJAAL M
ENG19CS0334
SEM V
F SECTION

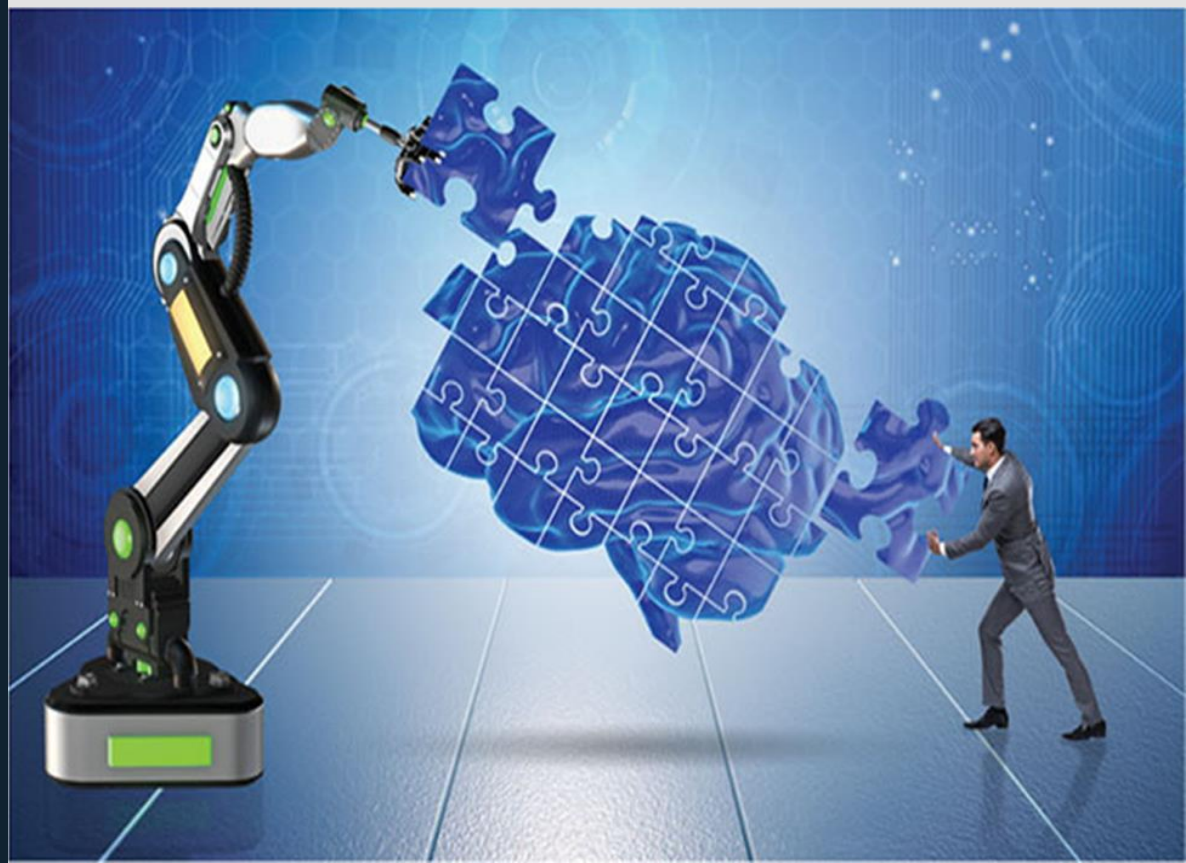
UNDER THE SUPERVISION OF
MR. CVSN REDDY

CONTENTS

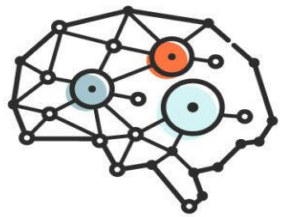
1	INTRODUCTION TO MACHINE LEARNING
2	SUPERVISED, UNSUPERVISED AND REINFORCEMENT LEARNING
3	PROCESS OF DATA WAREHOUSE AND DATA MINING
4	MACHINE LEARNING ALGORITHMS ❑ LINEAR REGRESSION ❑ LOGISTIC REGRESSION ❑ DECISION TREE ❑ RANDOM FOREST ❑ KNN ❑ SVM
5	APPLICATION OF ML IN DIFFERENT DOMAINS- HEALTH CARE, TELECOM


MACHINE LEARNING

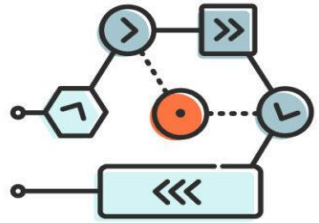

Introduction to
Machine Learning
Learns
Predicts
Improves
Ordinary System
With AI
Machine Learning



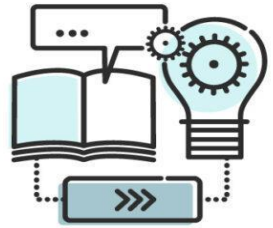
MACHINE LEARNING



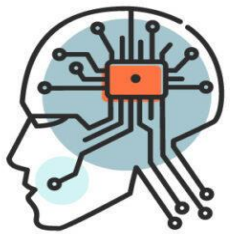
DEEP LEARNING



ALGORITHM

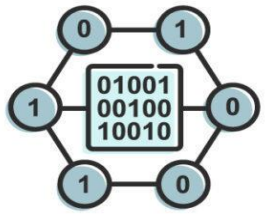


LEARNING

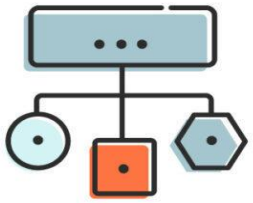


AI

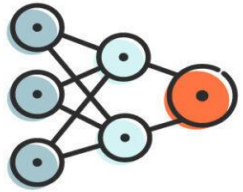
IMPROVES



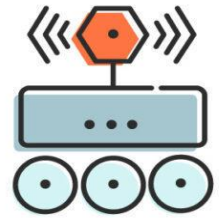
DATA MINING



CLASSIFICATION



NEURAL NETWORKS

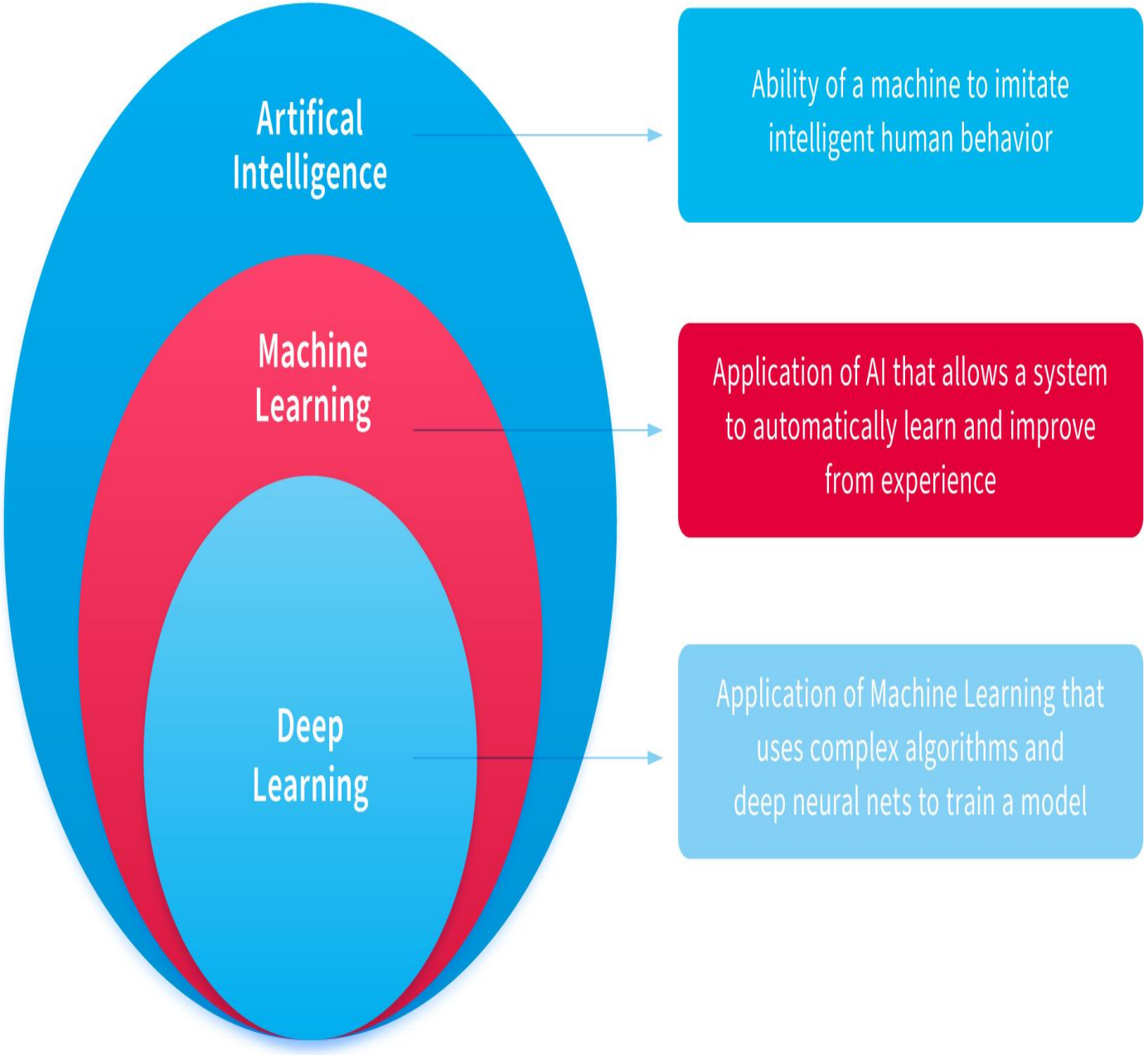


AUTONOMUS

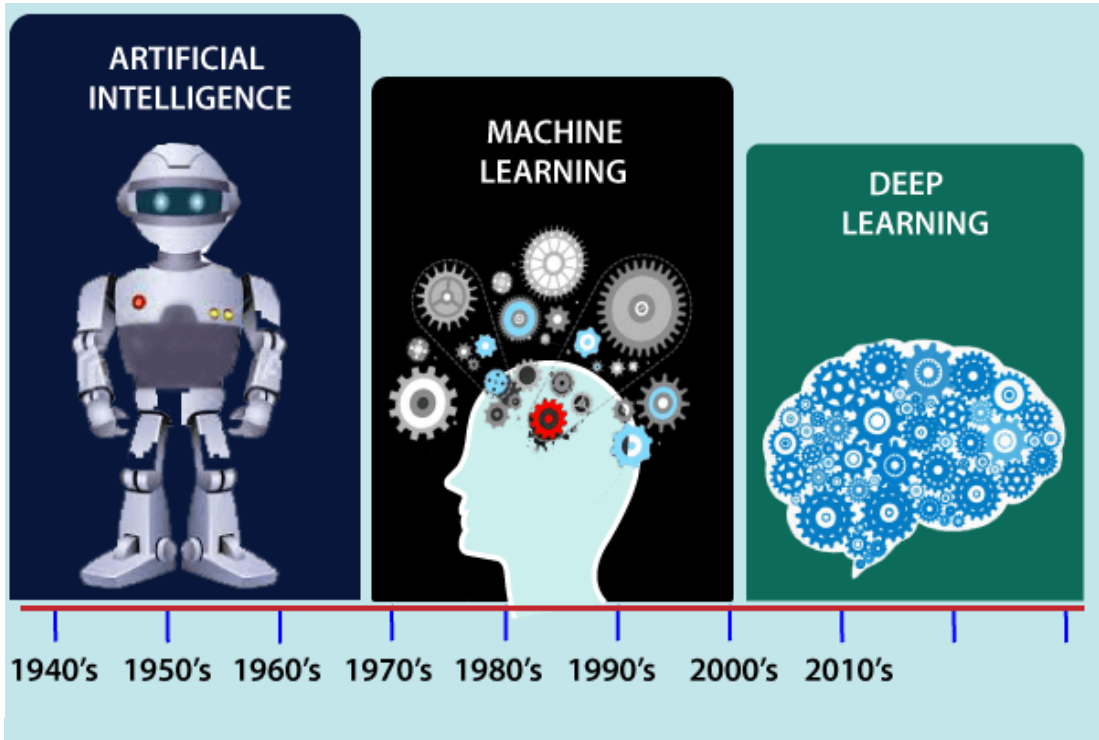
ANALYZE

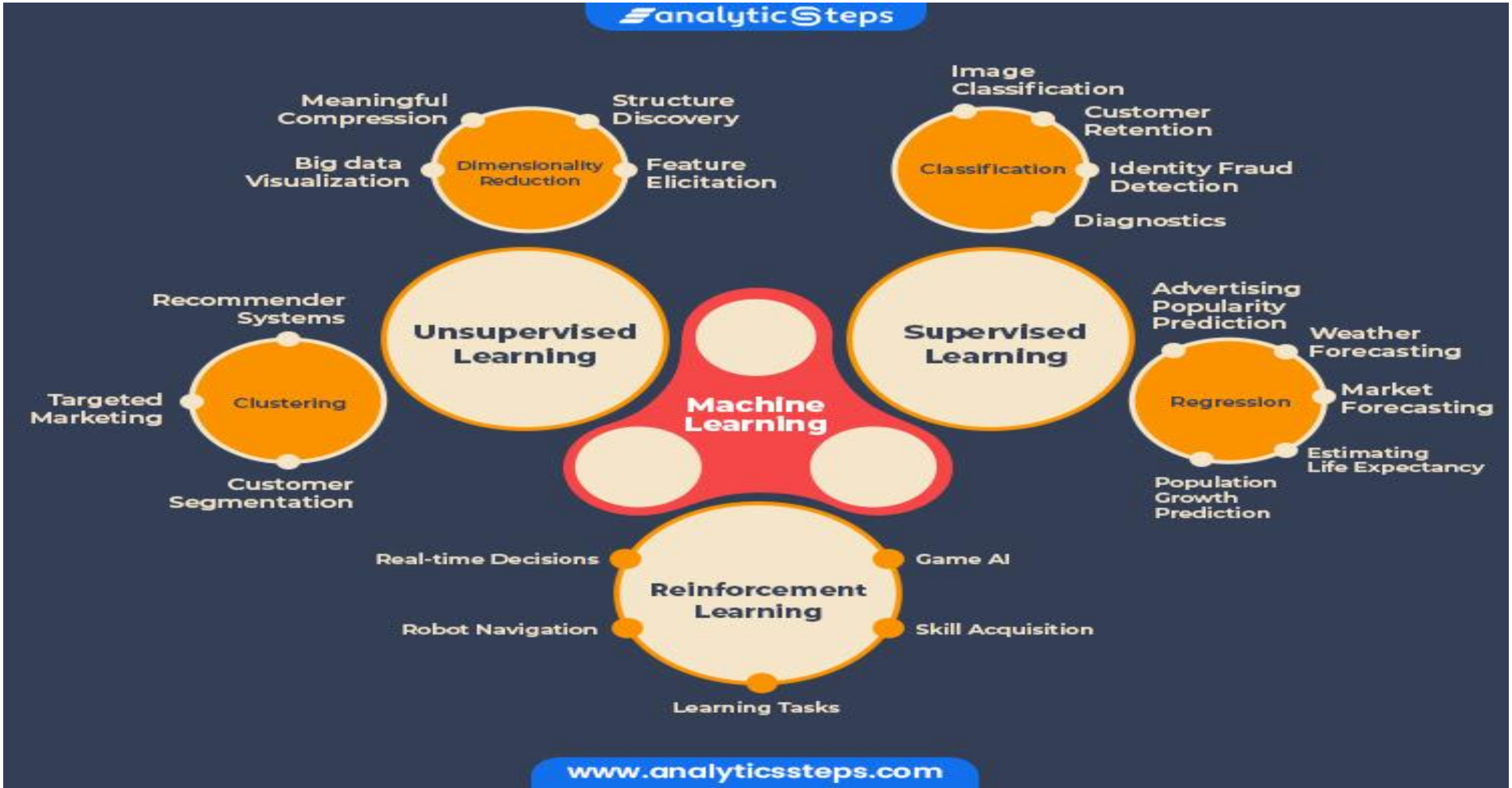
STEPS USED IN MACHINE LEARNING

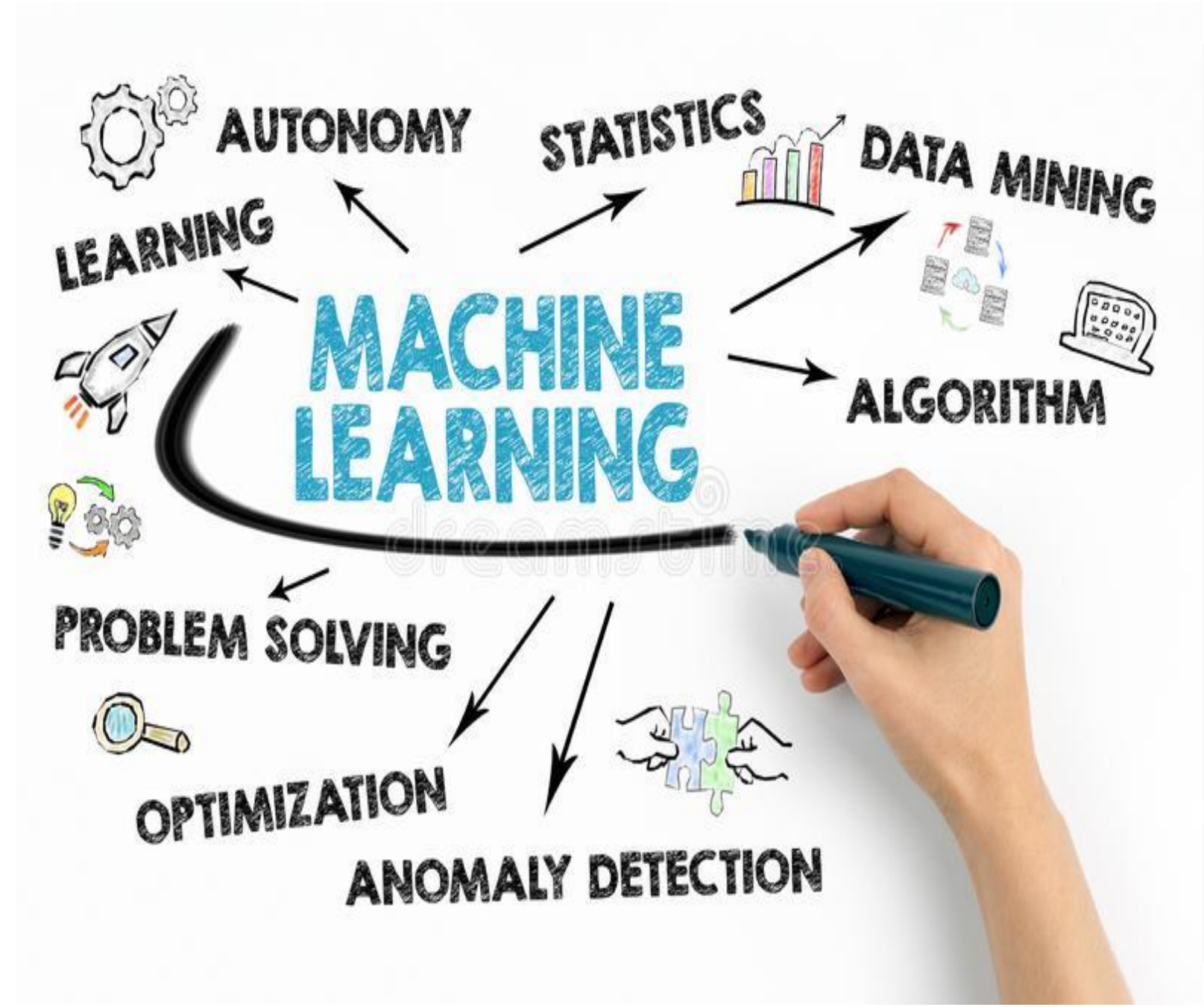
- **Collecting data**: Be it the raw data from excel, access, text files etc., this step (gathering past data) forms the foundation of the future learning. The better the variety, density and volume of relevant data, better the learning prospects for the machine becomes.
- **Preparing the data**: Any analytical process thrives on the quality of the data used. One needs to spend time determining the quality of data and then taking steps for fixing issues such as missing data and treatment of outliers. Exploratory analysis is perhaps one method to study the nuances of the data in details thereby burgeoning the nutritional content of the data.
- **Training a model**: This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts – train and test (proportion depending on the prerequisites); the first part (training data) is used for developing the model. The second part (test data), is used as a reference.
- **Evaluating the model**: To test the accuracy, the second part of the data (holdout / test data) is used. This step determines the precision in the choice of the algorithm based on the outcome. A better test to check accuracy of model is to see its performance on data which was not used at all during model build.
- **Improving the performance**: This step might involve choosing a different model altogether or introducing more variables to augment the efficiency. That's why significant amount of time needs to be spent **in data collection and preparation.**

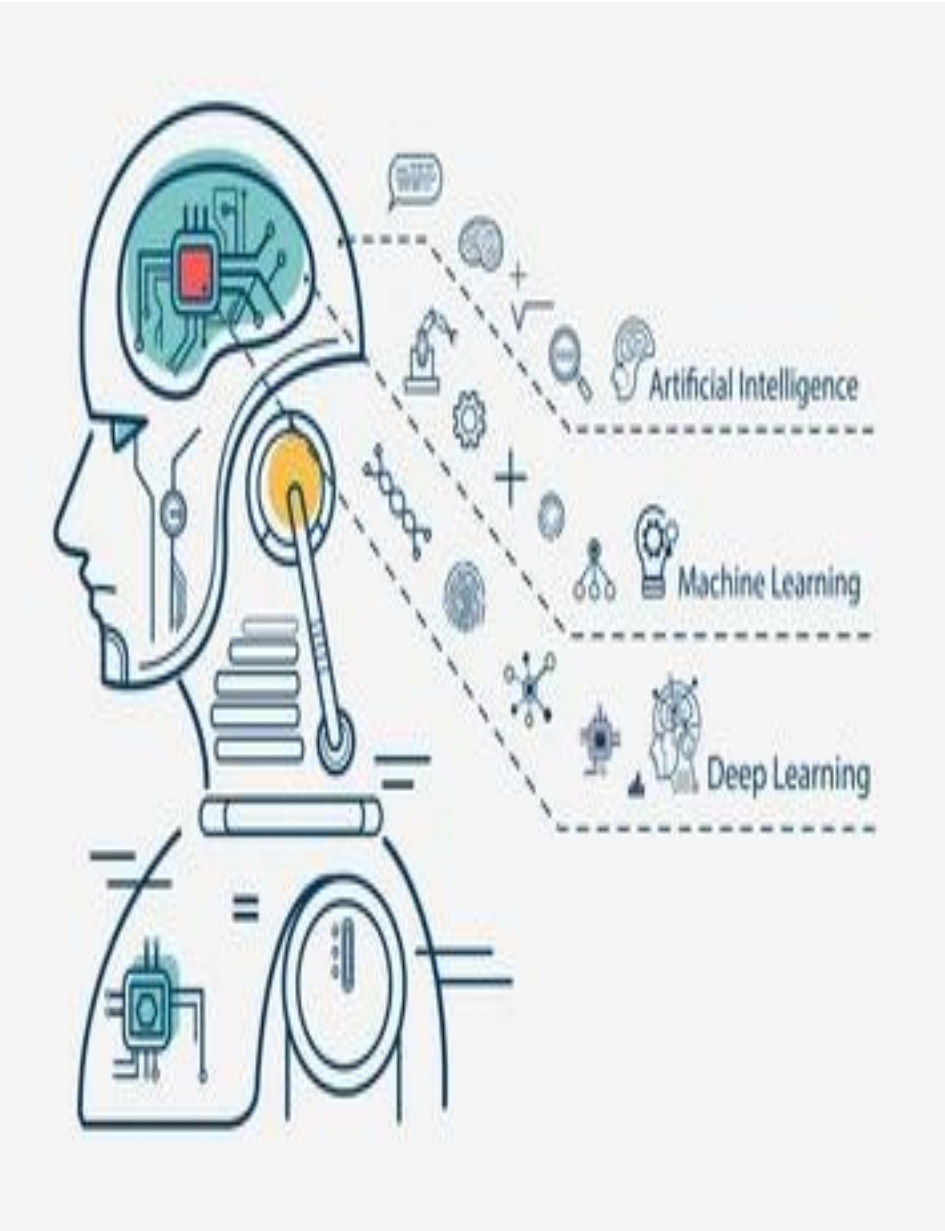


Machine Learning is the field of computer science that uses statistical techniques to give computer systems the ability to learn with Data, without being explicitly programmed.




analyticSteps
Machine Learning
Unsupervised Learning
Supervised Learning
Reinforcement Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualization
Feature Elicitation
Classification
Image Classification
Customer Retention
Identity Fraud Detection
Diagnostics
Clustering
Recommender Systems
Targeted Marketing
Customer Segmentation
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating Life Expectancy
Population Growth Prediction
Real-time Decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks
www.analyticssteps.com


AUTONOMY
STATISTICS
DATA MINING
LEARNING
MACHINE LEARNING
ALGORITHM
PROBLEM SOLVING
OPTIMIZATION
ANOMALY DETECTION


Artificial Intelligence
Machine Learning
Deep Learning

Classical Machine Learning

Task Driven

Data Driven

Supervised Learning

(Pre Categorized Data)

Classification

(Divide the socks by Color)

Eg. Identity Fraud Detection

Regression

(Divide the Ties by Length)

Eg. Market Forecasting

Unsupervised Learning

(Unlabelled Data)

Clustering

(Divide by Similarity)

Eg. Targeted Marketing

Association

(Identify Sequences)

Eg. Customer Recommendation

Dimensionality Reduction

(Wider Dependencies)

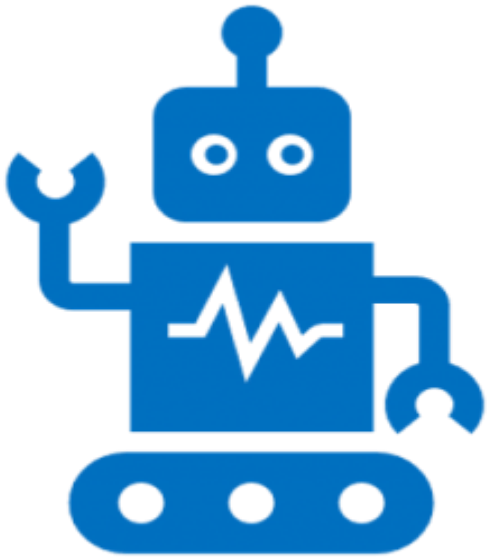
Eg. Big Data Visualization

Obj: Predications & Predictive Models

Pattern/ Structure Recognition

Reinforcement Learning

Machine



Takes Action

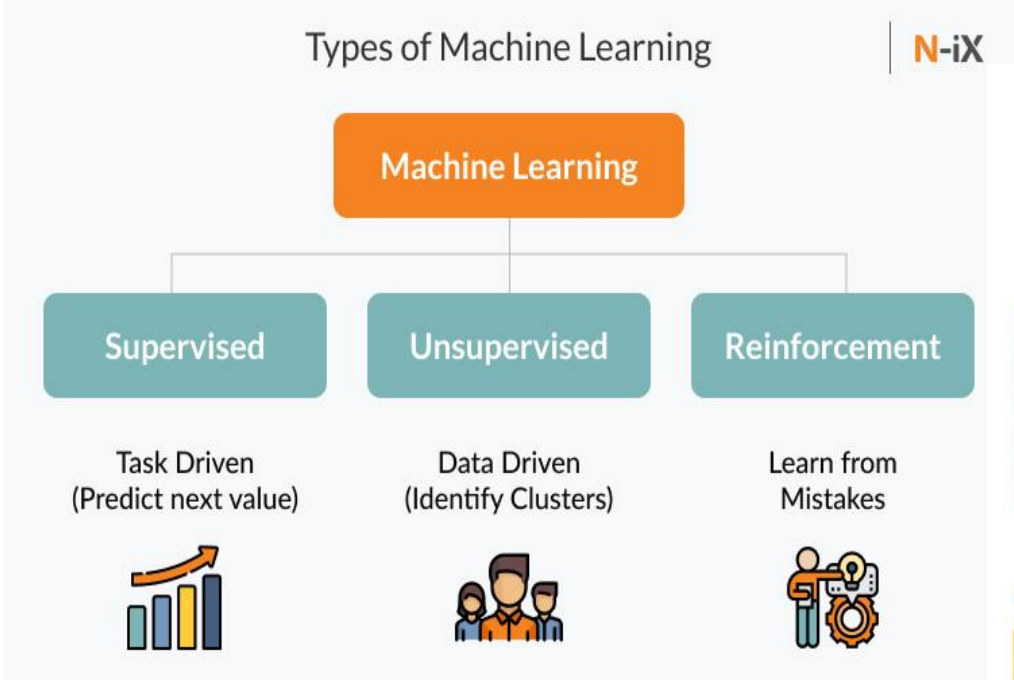
Gives Reward

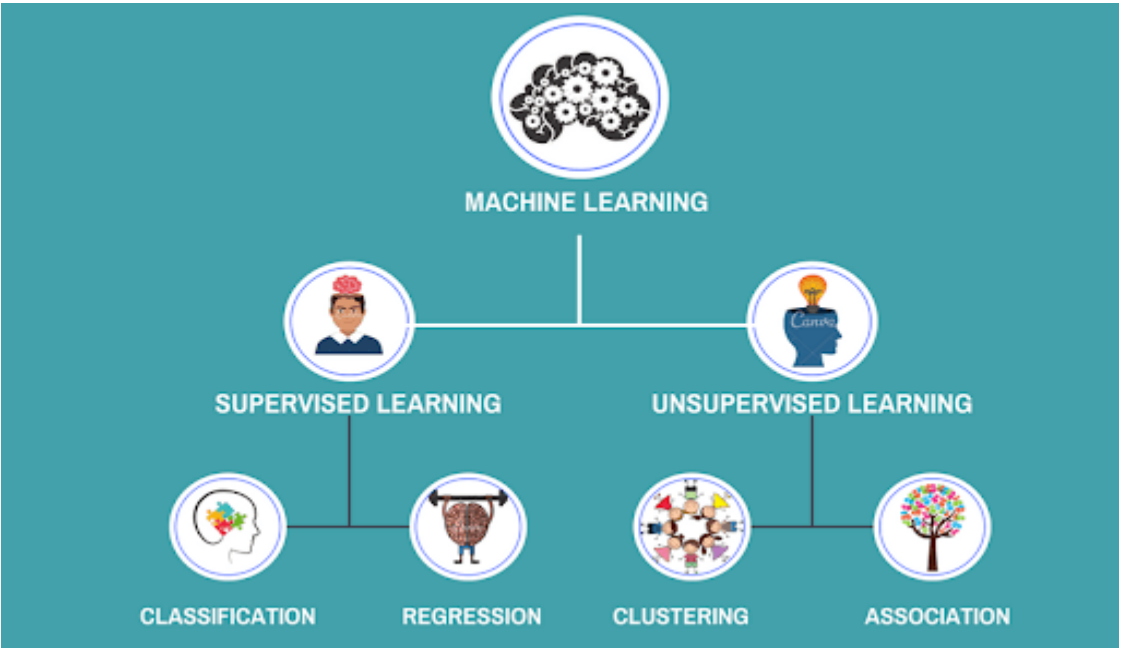
Environment

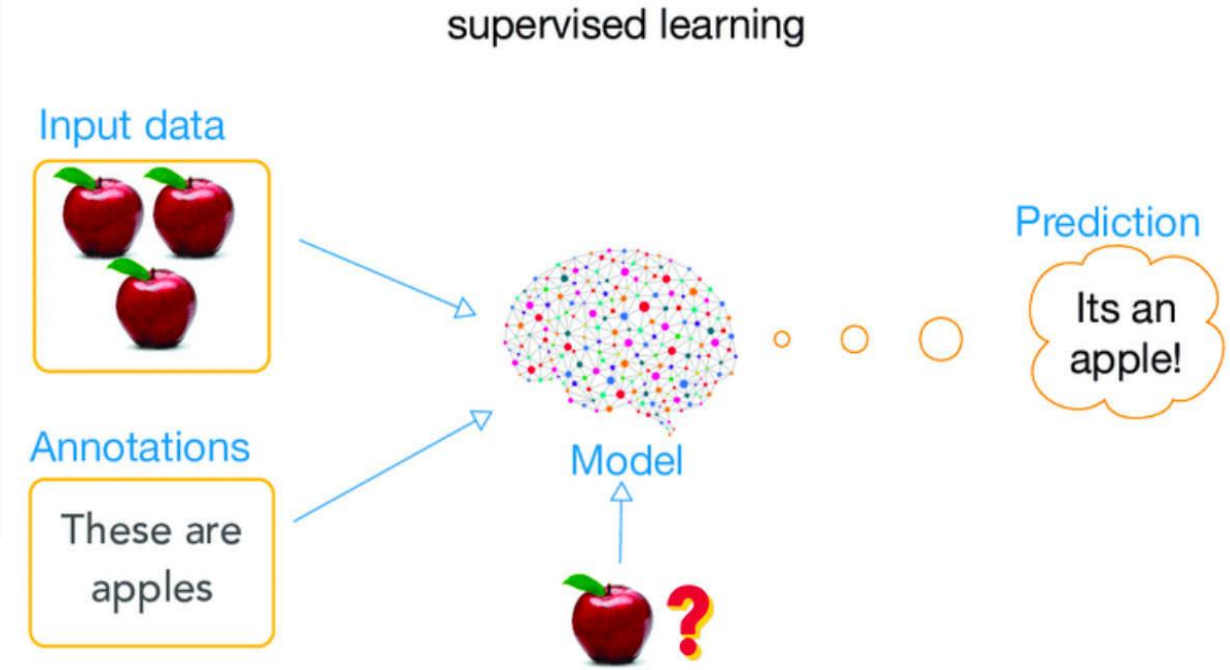
The machine keeps on taking actions based on each reward it gets. This process keeps on iterating until a desired level of learning is not reached.

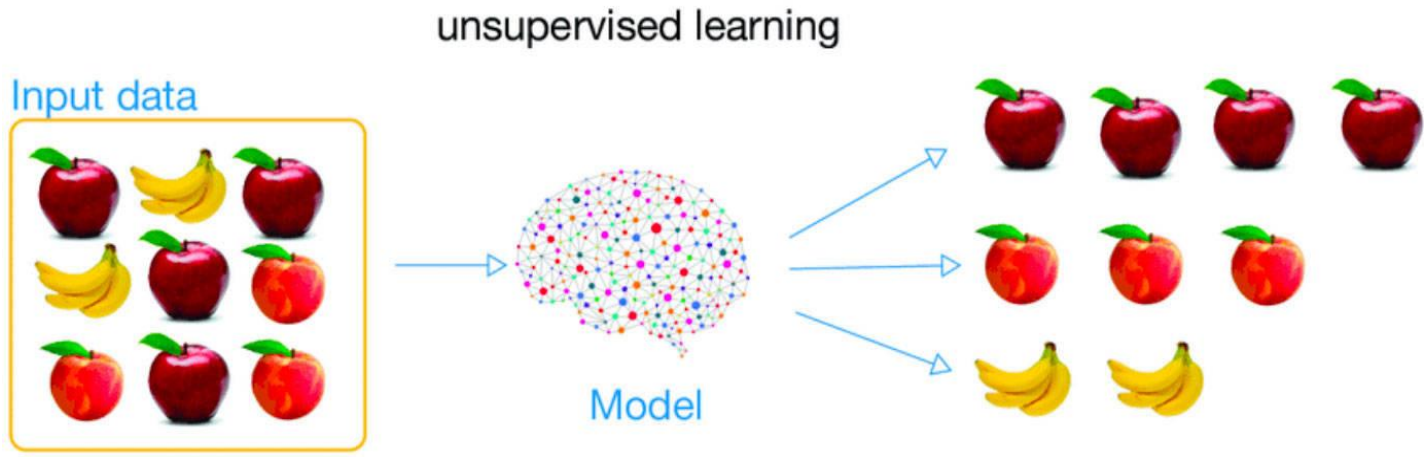
Reinforcement Learning

- It is the ability of an agent to interact with the environment and find out what is the best outcome. It follows the concept of hit and trial method.
- The agent is rewarded or penalized with a point for a correct or a wrong answer, and on the basis of the positive reward points gained the model trains itself. And again once trained it gets ready to predict the new data presented to it.


Types of Machine Learning
N-iX
Machine Learning
Supervised
Unsupervised
Reinforcement
Task Driven
(Predict next value)
Data Driven
(Identify Clusters)
Learn from
Mistakes


MACHINE LEARNING
SUPERVISED LEARNING
UNSUPERVISED LEARNING
CLASSIFICATION
REGRESSION
CLUSTERING
ASSOCIATION


supervised learning
Input data
Annotations
These are
apples
Model
Prediction
Its an
apple!


unsupervised learning
Input data
Model

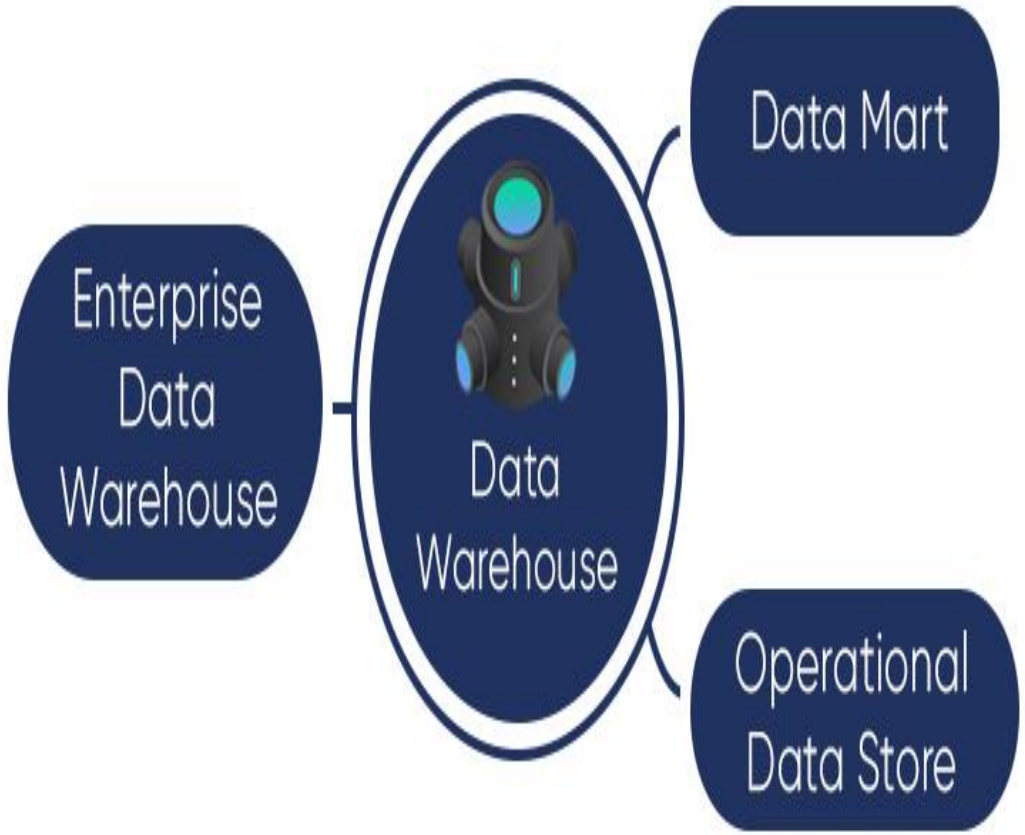
Data Warehousing

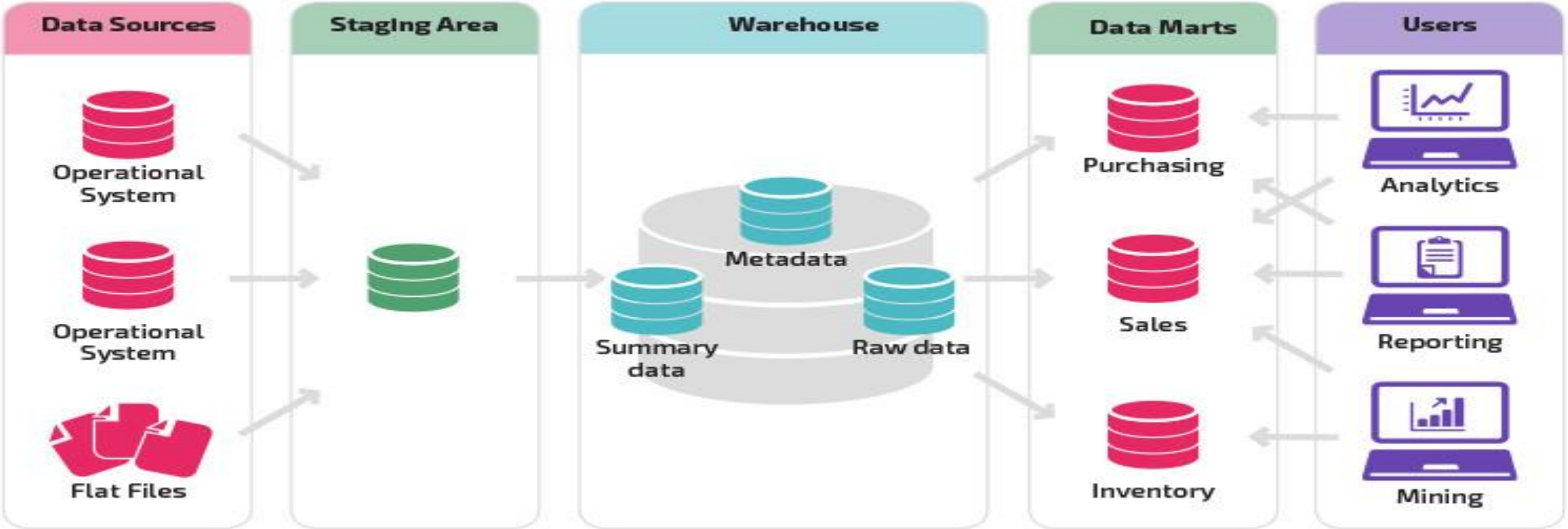
- **Data warehouse** refers to the process of compiling and organizing data into one common database.
- **General stages** of the use of a data warehouse:
- **Offline Operational Database:** In this stage, data is copied to a server from an operating system so that loading, processing, and reporting the data does not impact the performance of the operational system.
- **Offline Data Warehouse:** The data stored in the warehouse is regularly updated from the operational database to derive useful business insights.
- **Real-time Data Warehouse:** Whenever a transaction takes place in the operational database, the same is updated in the data warehouse.
- **Integrated Data Warehouse:** Every transaction taking place in the operational database is updated simultaneously in the data warehouse. Then, the warehouse generates transactions that are forwarded to the operational database.

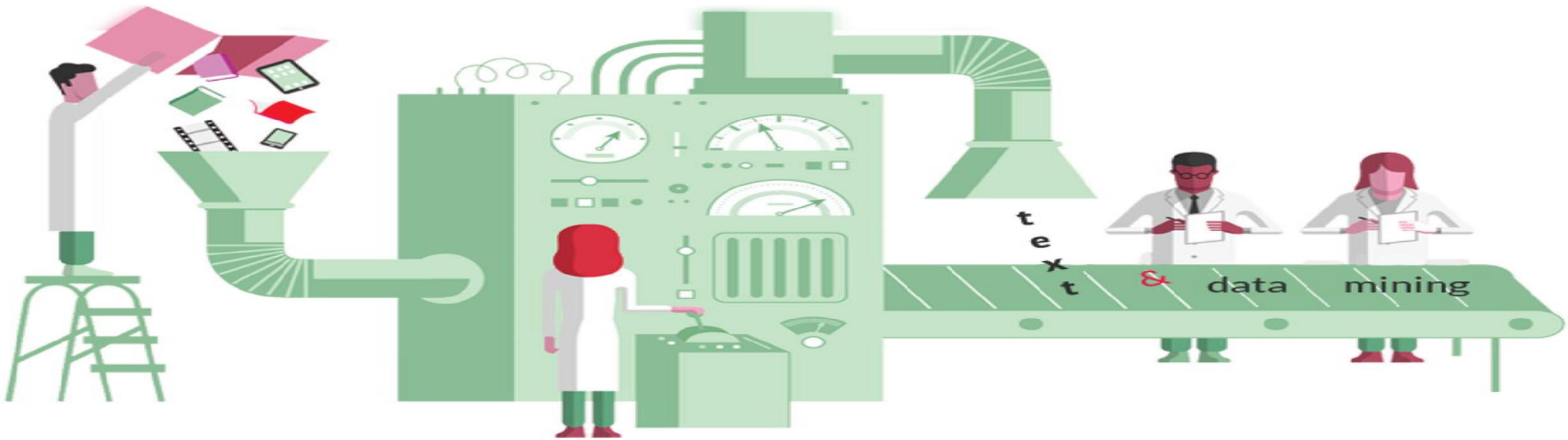
Stages of Data Warehouse

- Operational Database
- Offline Data Warehouse
- Real-time Data Warehouse
- Integrated Data Warehouse

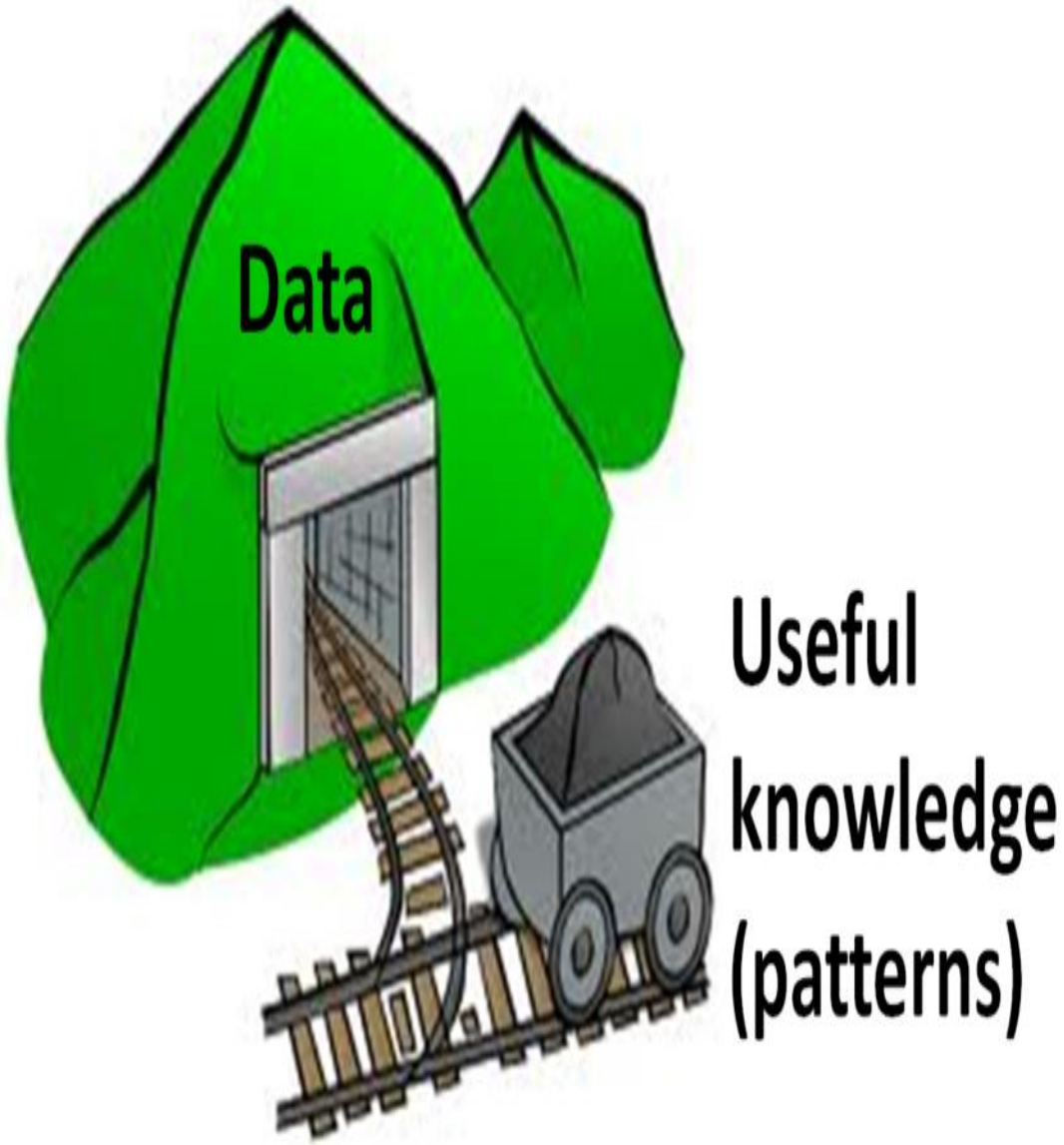
Type Of Data Warehousing



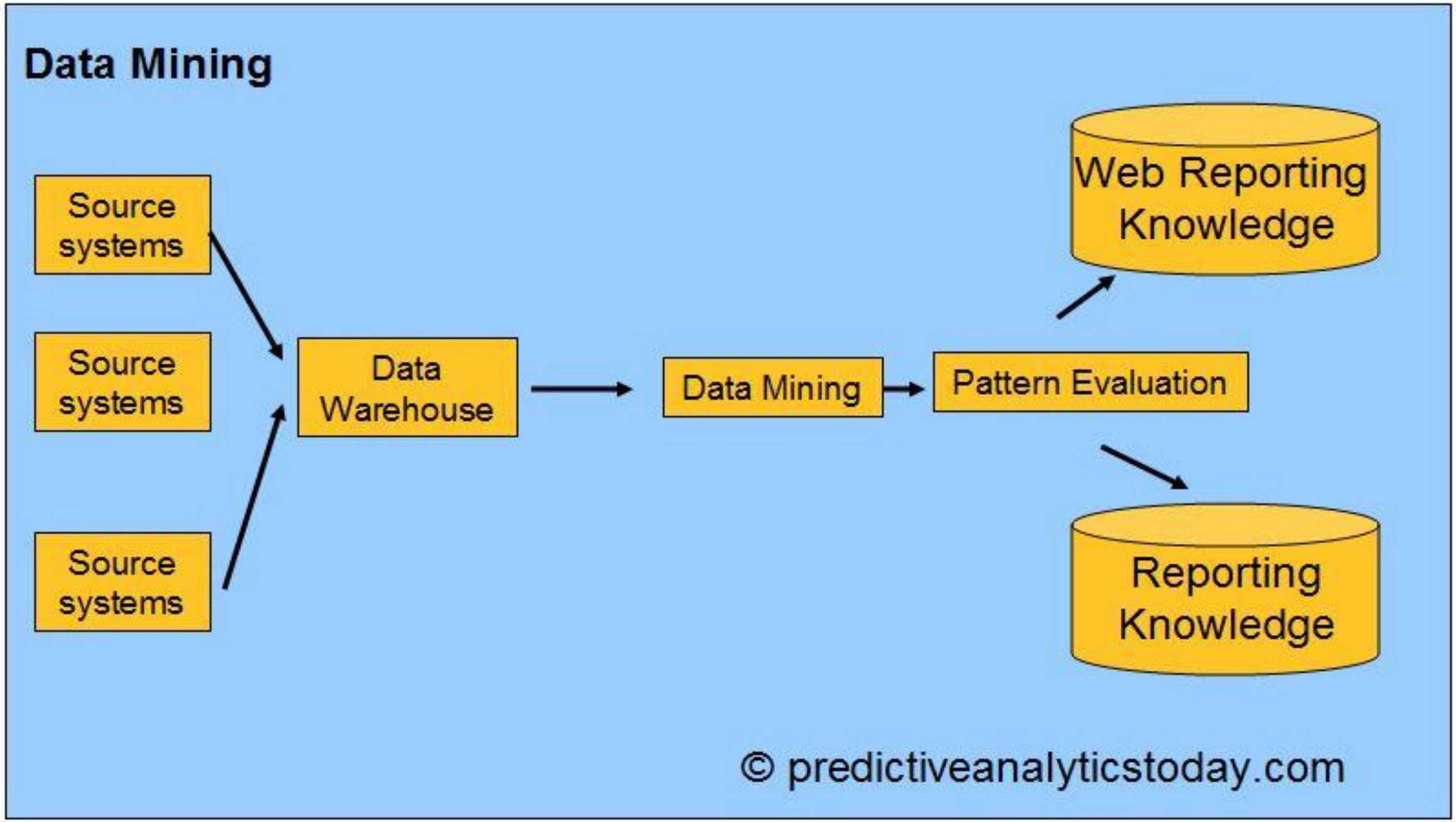

Data Sources
Staging Area
Warehouse
Data Marts
Users
Operational System
Operational System
Flat Files
Metadata
Summary data
Raw data
Purchasing
Sales
Inventory
Analytics
Reporting
Mining

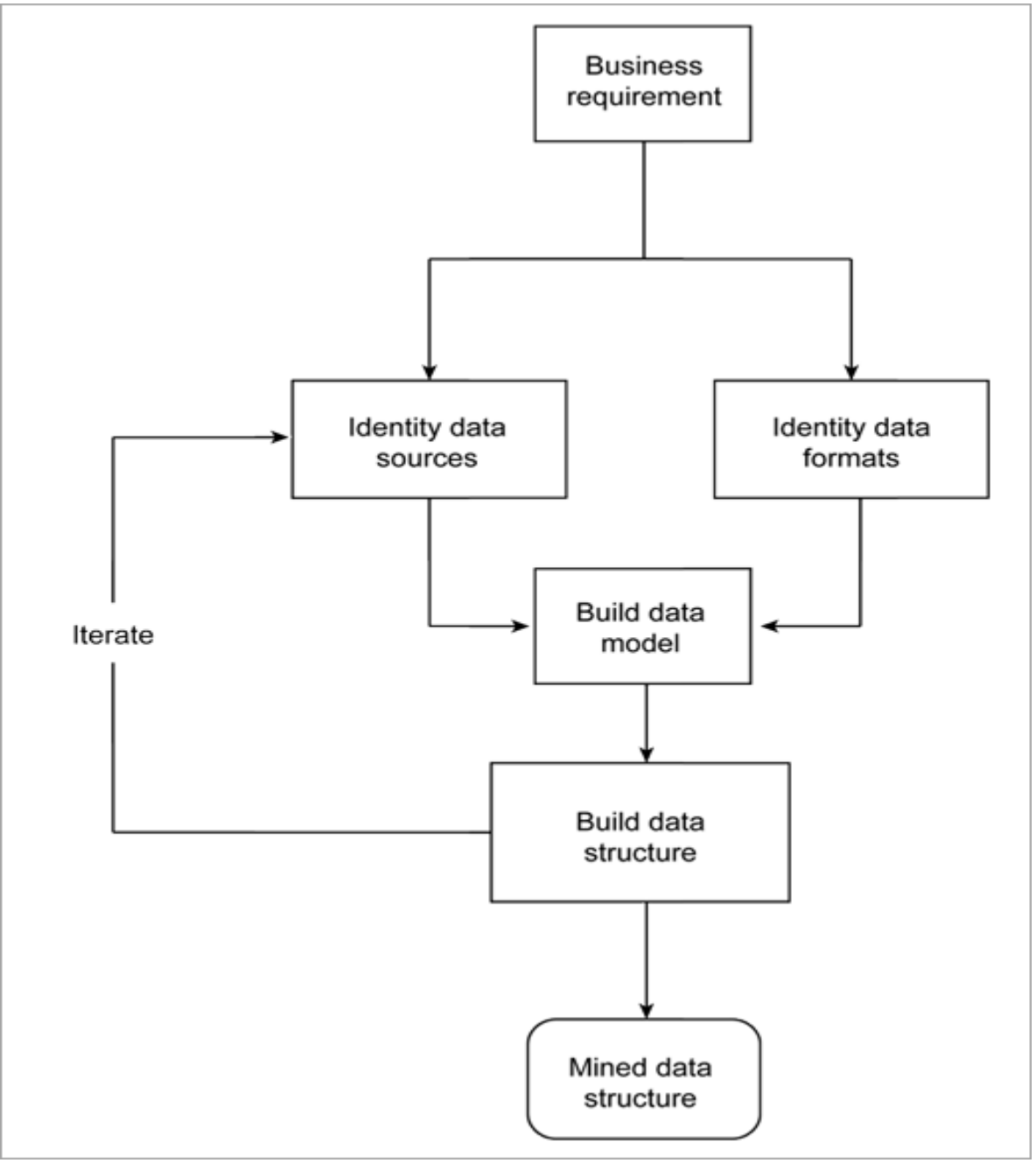

t e x t & data mining

DATA MINING



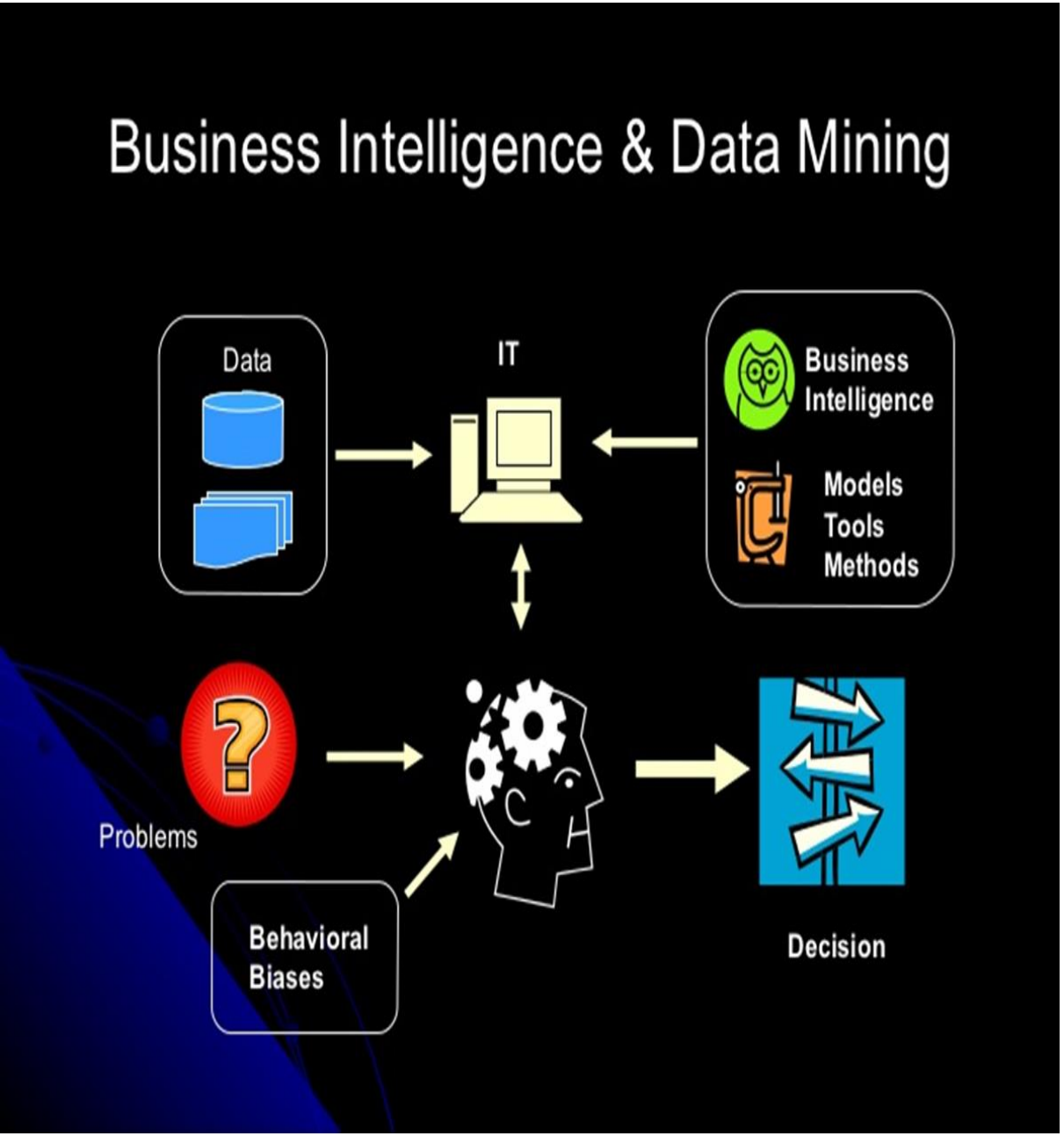
- Data mining is concerned with the analysis and picking out relevant information
- Data mining analysts collect as much as raw data as they can to arrive at reliable conclusions and decisions.
- As mining operations where large amounts of low grade materials are cleaned and processed in order to find something of value.

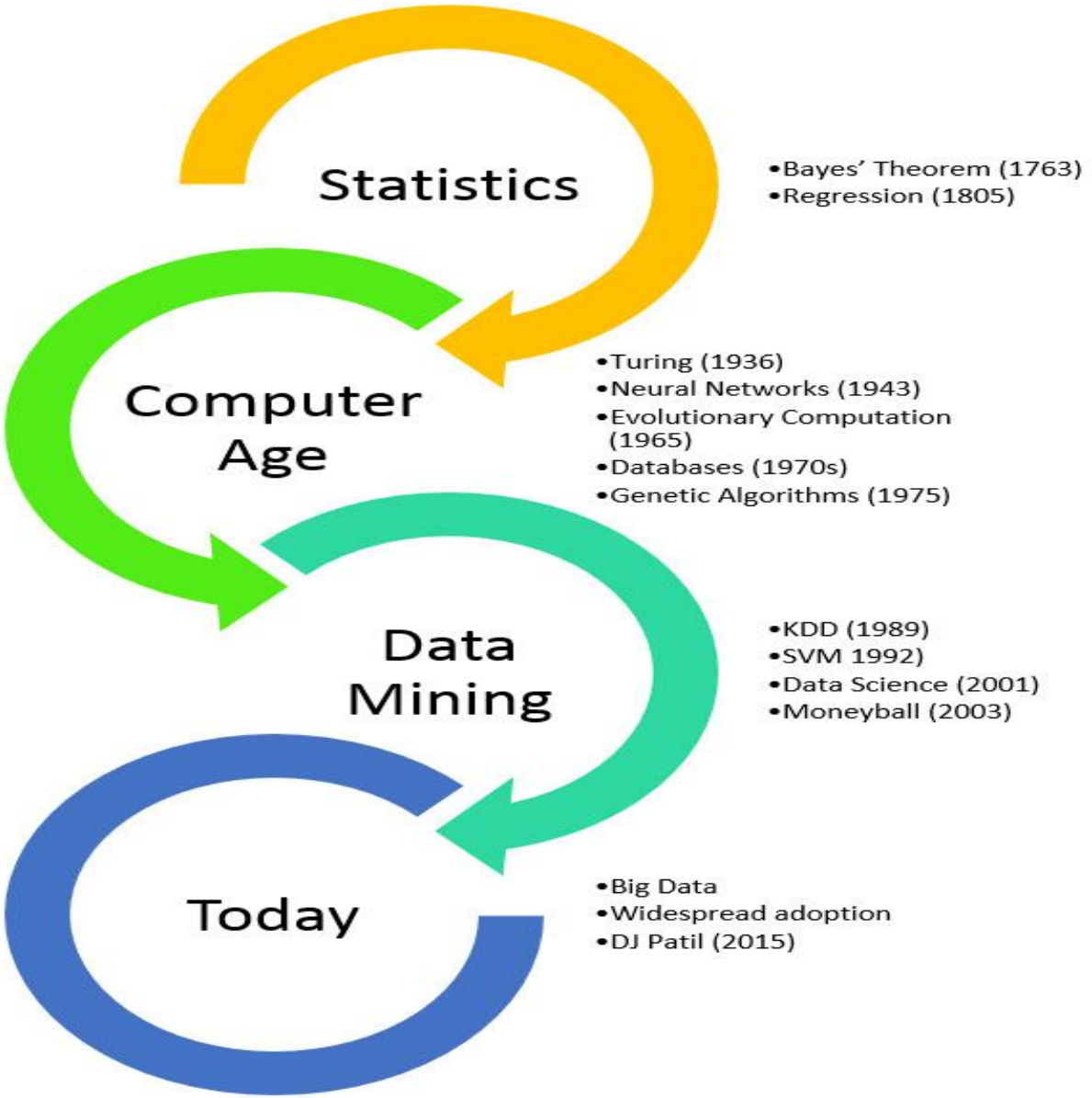

Data Mining
Source systems
Source systems
Source systems
Data Warehouse
Data Mining
Pattern Evaluation
Web Reporting Knowledge
Reporting Knowledge
© predictiveanalyticstoday.com



STAGES OF DATA PREPROCESSING

- Data Cleaning
- Data Integration
- Data Reduction
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Representation


Business Intelligence & Data Mining
Data
IT
Business Intelligence
Models
Tools
Methods
Problems
Behavioral Biases
Decision


Statistics
•Bayes' Theorem (1763)
•Regression (1805)
Computer Age
•Turing (1936)
•Neural Networks (1943)
•Evolutionary Computation (1965)
•Databases (1970s)
•Genetic Algorithms (1975)
Data Mining
•KDD (1989)
•SVM 1992)
•Data Science (2001)
•Moneyball (2003)
Today
•Big Data
•Widespread adoption
•DJ Patil (2015)

Machine Learning - Algorithms

Linear Regression: Linear regression is used in which value of dependent variable is predicted through independent variables. A relationship is formed by mapping the dependent and independent variable on a line and that line is called regression line which is represented by Y= a*X + b.

Logistic Regression: In logistic regression we have lot of data whose classification is done by building an equation. This method is used to find the discrete dependent variable from the set of independent variables. Its goal is to find the best fit set of parameters. In this classifier, each feature is multiplied by a weight and then all are added. Then the result is passed to sigmoid function which produces the binary output.

Decision Tree: It belongs to supervised learning algorithm. Decision tree can be used to classification and regression both having a tree like structure. In a decision tree building algorithm first the best attribute of dataset is placed at the root, then training dataset is split into subsets. Splitting of data depends on the features of datasets. This process is done until the whole data is classified and we find leaf node at each branch. Information gain can be calculated to find which feature is giving us the highest information gain. Decision trees are built for making a training model which can be used to predict class or the value of target variable.

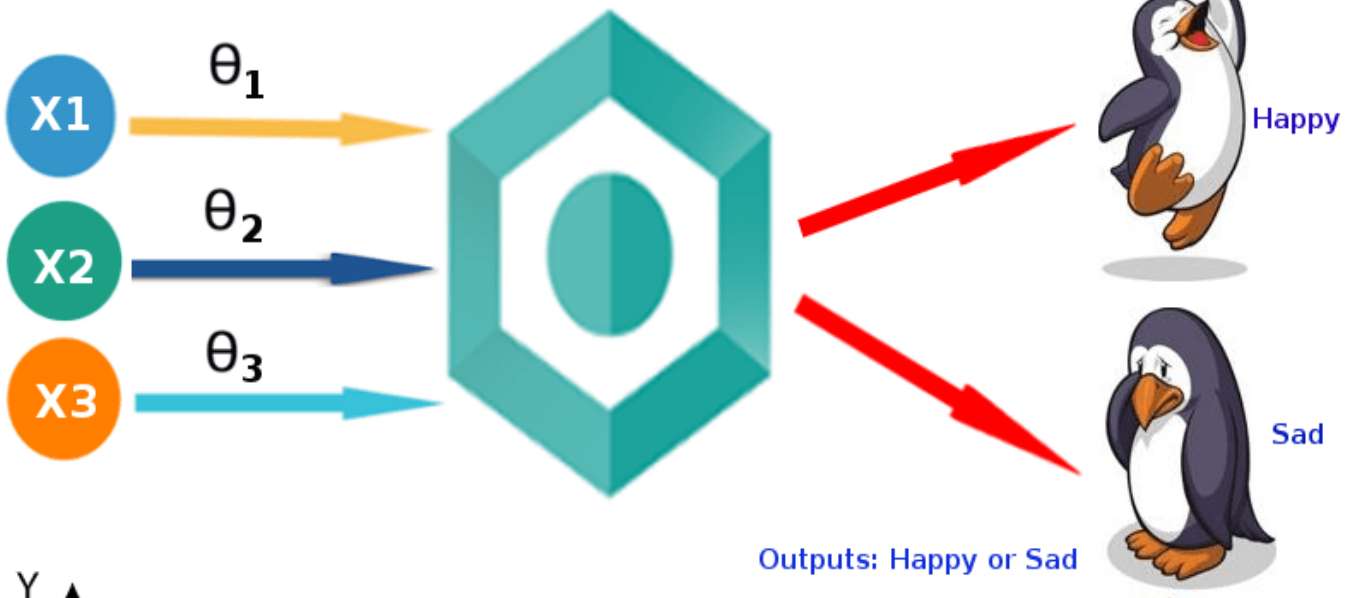
Machine Learning - Algorithms

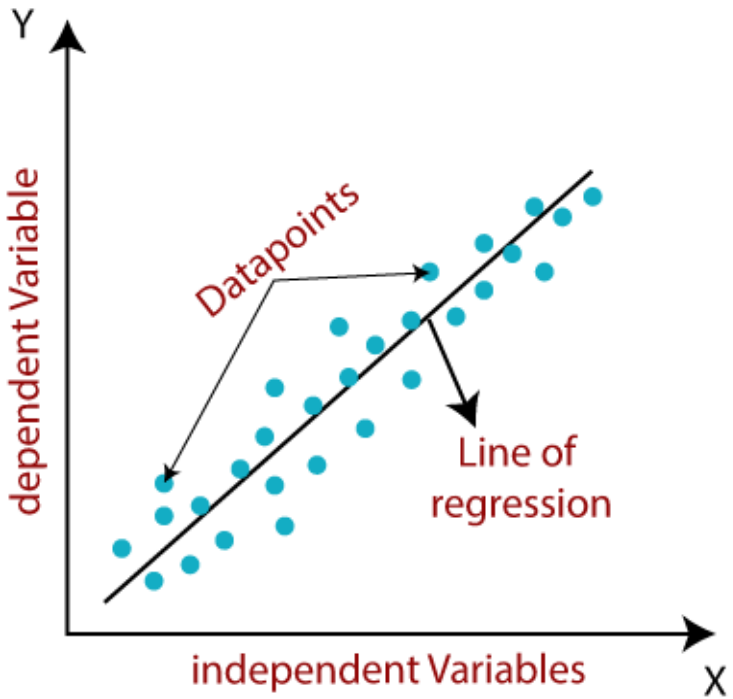
Support vector machine: Support vector machine is a binary classifier. Raw data is drawn on the n- dimensional plane. In this a separating hyperplane is drawn to differentiate the datasets. The line drawn from centre of the line separating the two closest data-points of different categories is taken as an optimal hyperplane. This optimised separating hyperplane maximizes the margin of training data. Through this hyperplane, new data can be categorised

KNN: This method is used for both classification and regression. It is among the simplest method of machine learning algorithms. It stores the cases and for new data it checks the majority of the k neighbours with which it resembles the most. KNN makes predictions using the training dataset directly.

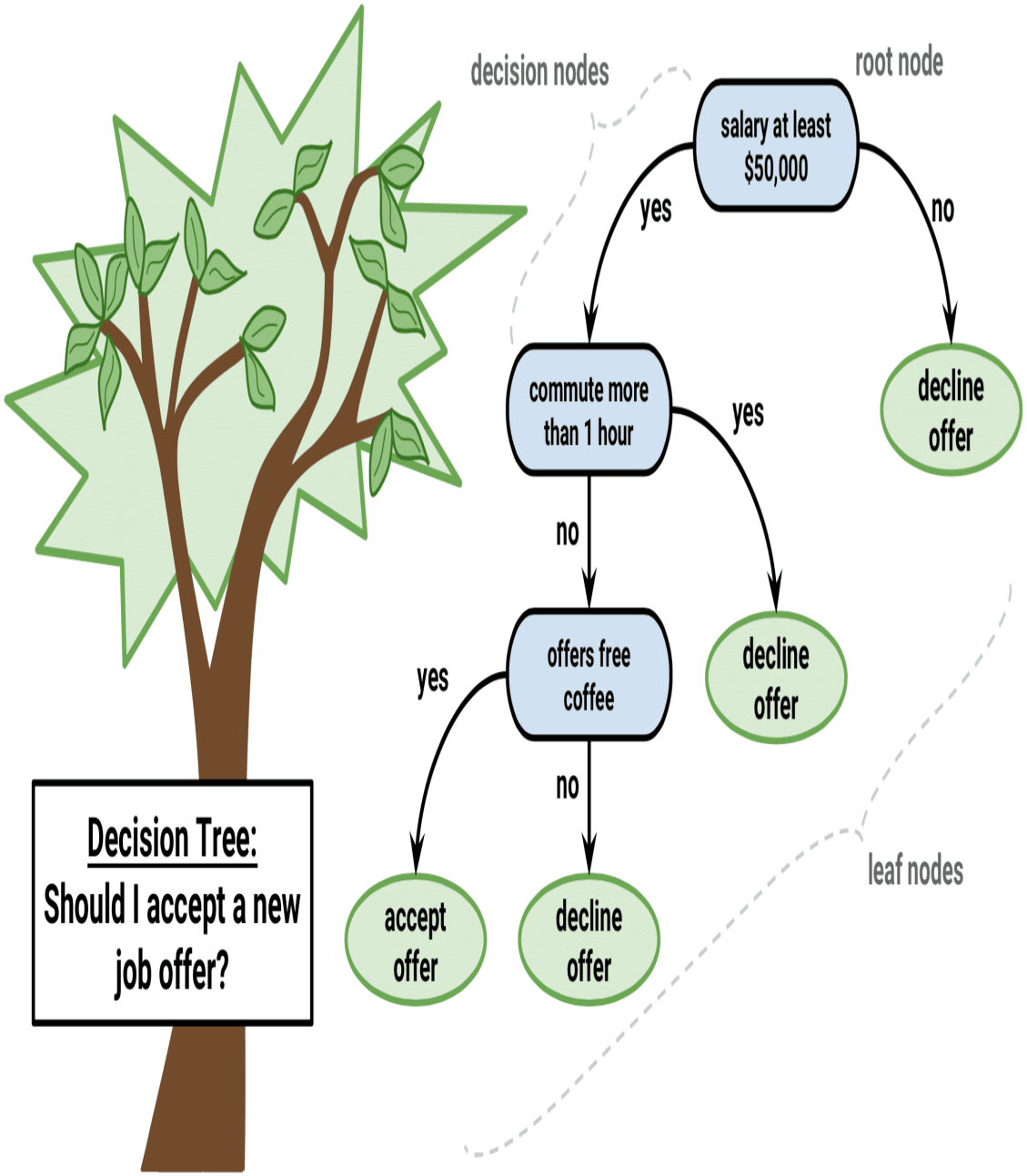
Random Forest: It is a supervised classification algorithm. Multiple number of decision trees taken together forms a random forest algorithm i.e the collection of many classification tree. It can be used for classification as well as regression. Each decision tree includes some rule based system. For the given training dataset with targets and features, the decision tree algorithm will have set of rules. In random forest unlike decision trees there is no need to calculate information gain to find root node. It use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome. Further it calculates the vote for each predicted target. Thus high voted prediction is considered as the final prediction from the random forest algorithm.

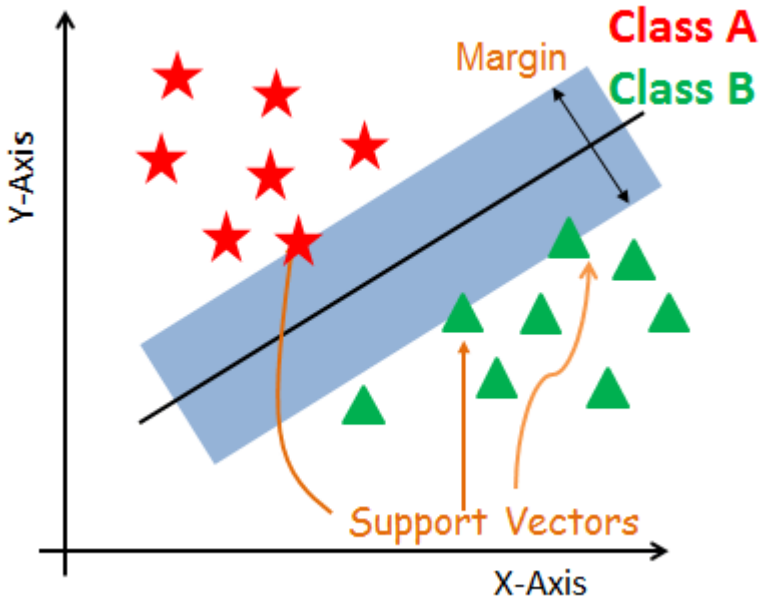
Logistic Regression Model



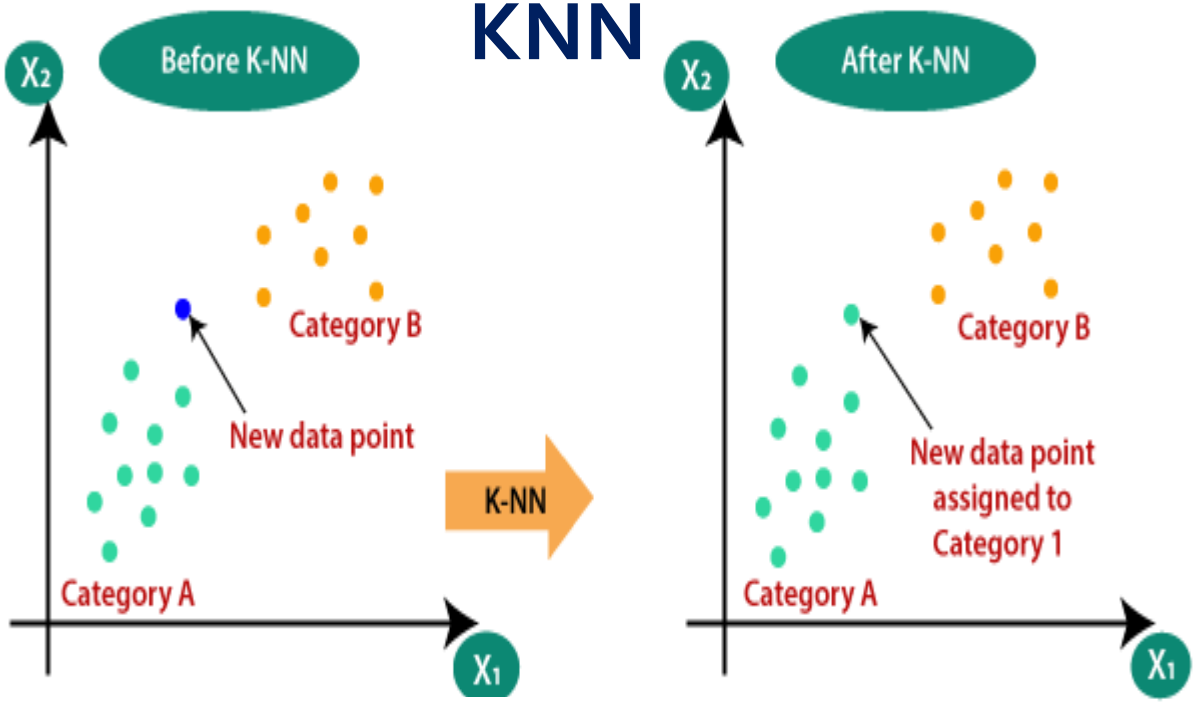


LINEAR REGRESSION

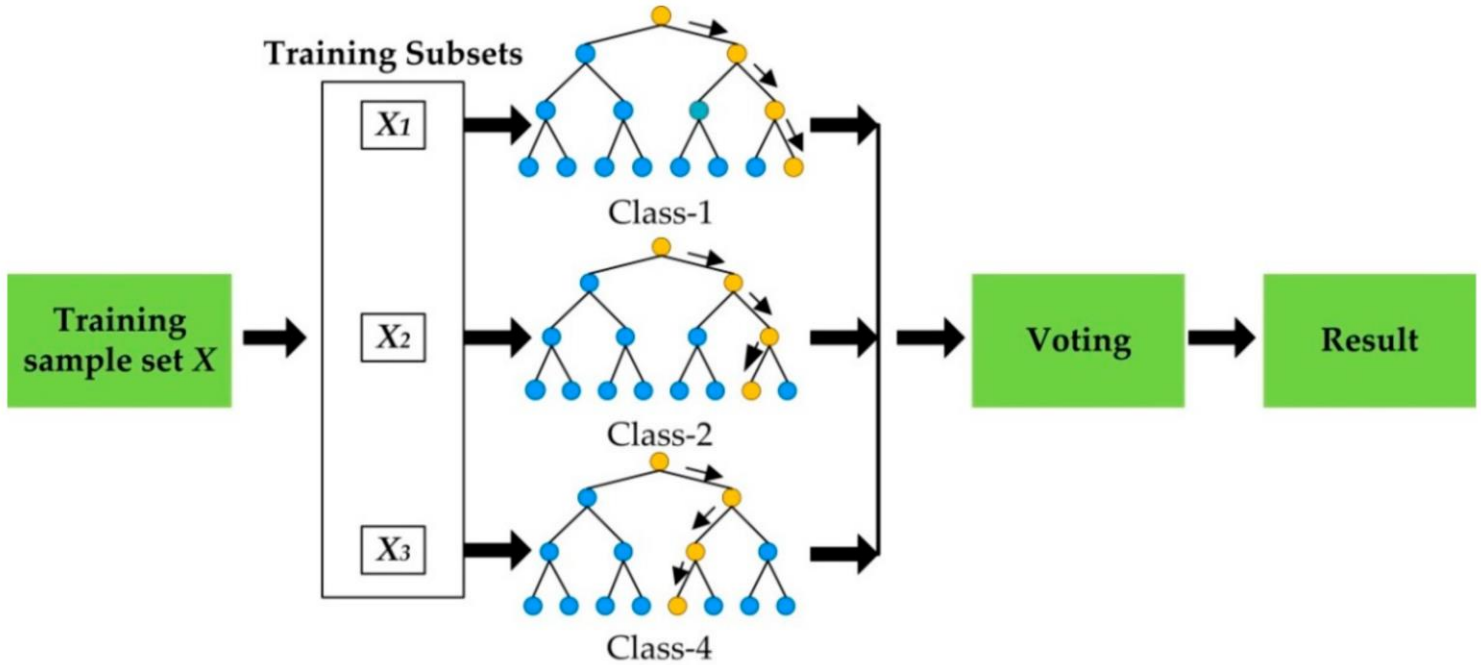




KNN



SVM

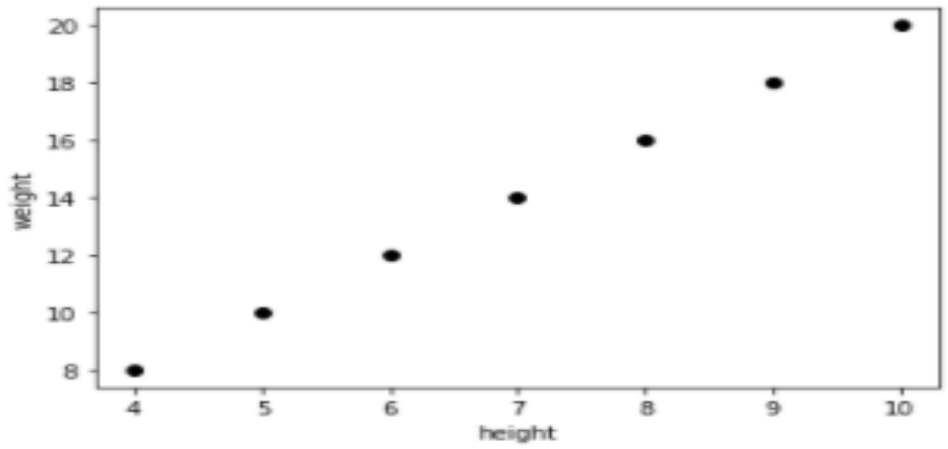


RANDOM FOREST

1.LINEAR REGRESSION

```
#-----------------------------------------------------------LINEAR REGRESSION
import matplotlib
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import sklearn
from sklearn import linear_model
height=[[4.0],[5.0],[6.0],[7.0],[8.0],[9.0],[10.0]]
weight=[  8, 10 , 12, 14, 16, 18, 20]
plt.scatter(height,weight,color='black')
plt.xlabel("height")
plt.ylabel("weight")
reg=linear_model.LinearRegression()
reg.fit(height,weight)
X_height=[[12.0]]
print(reg.predict(X_height))
```

```
[24.]
```



2.LOGISTIC REGRESSION

```
#-----------------------------------------------LOGISTIC REGRESSION
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
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=[[70]]
print(classifier.predict(X_marks))
```

```
[1]
```

3. DECISSION TREE

```
#-----------------------------------------------DECISION TREE
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
X = [[30],[40],[50],[60],[20],[10],[70]]
y = [0,1,1,1,0,0,1]
classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
classifier.fit(X,y)
X_marks=[[20]]
print(classifier.predict(X_marks))
```

```
[0]
```

4. RANDOM FOREST

```
#-----------------------------------------------------RANDOM FOREST
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
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)
classifier.fit(X,y)
X_marks=[[70]]
print(RandomForestRegModel.predict(X_marks))
```

```
[1.]
```

KNN

```
#-----------------------------------------------KNN
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
X = [[30],[40],[50],[60],[20],[10],[70]]
y = [0,1,1,1,0,0,1]
from sklearn.neighbors import KNeighborsClassifier
classifier= KNeighborsClassifier(n_neighbors=5, metric='minkowski', p=2 )
classifier.fit(X,y)
X_marks=[[50]]
print(classifier.predict(X_marks))
```

```
[1]
```

6.SVM

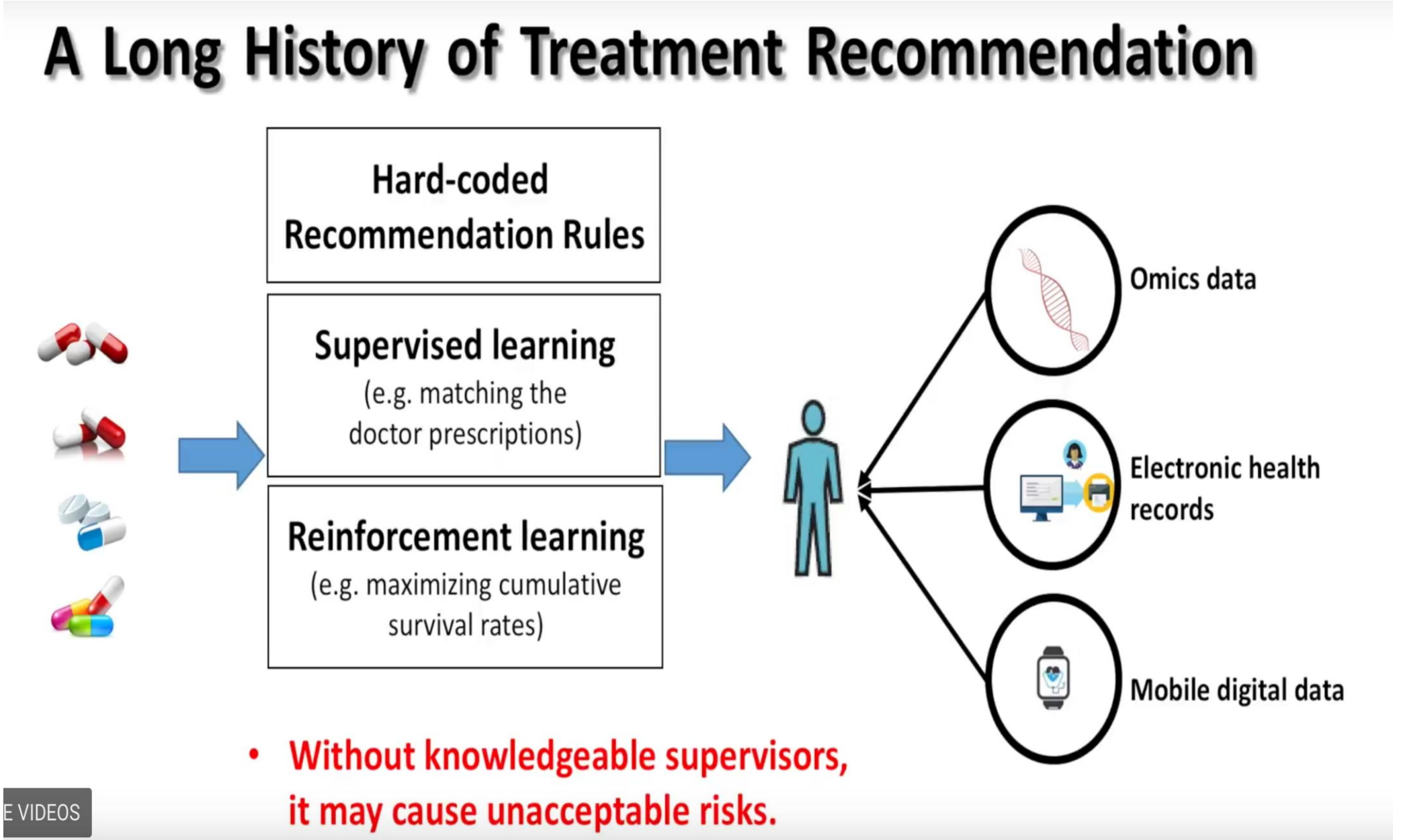
```
#----------------------------------------SVM
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.svm import SVC
X = [[30],[40],[50],[60],[20],[10],[70]]
y = [0,1,1,1,0,0,1]
classifier = SVC(kernel = 'linear', random_state = 0)
classifier.fit(X,y)
X_marks=[[21]]
print(classifier.predict(X_marks))
```

```
[0]
```

GIT HUB LINK- https://github.com/ENG19CS0334-TEJAAL-M/MACHINE-LEARNING-_ASSIGNMENT-1.git

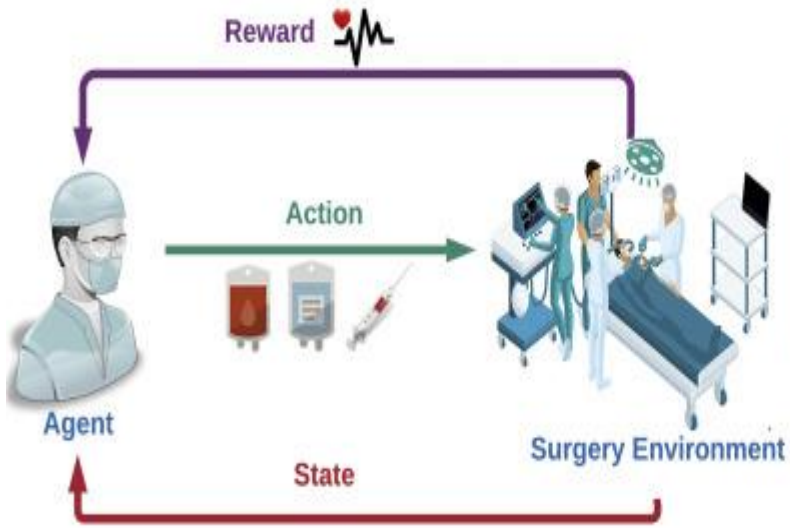
APPLICATIONS OF MACHINE LEARNING

APPLICATION OF ML IN HEALTH CARE



Challenges

* Determining reward function structure and parameters
* Adaptability of reward functions
* Limited reward evaluation methods
* Highly dependent on physician actions



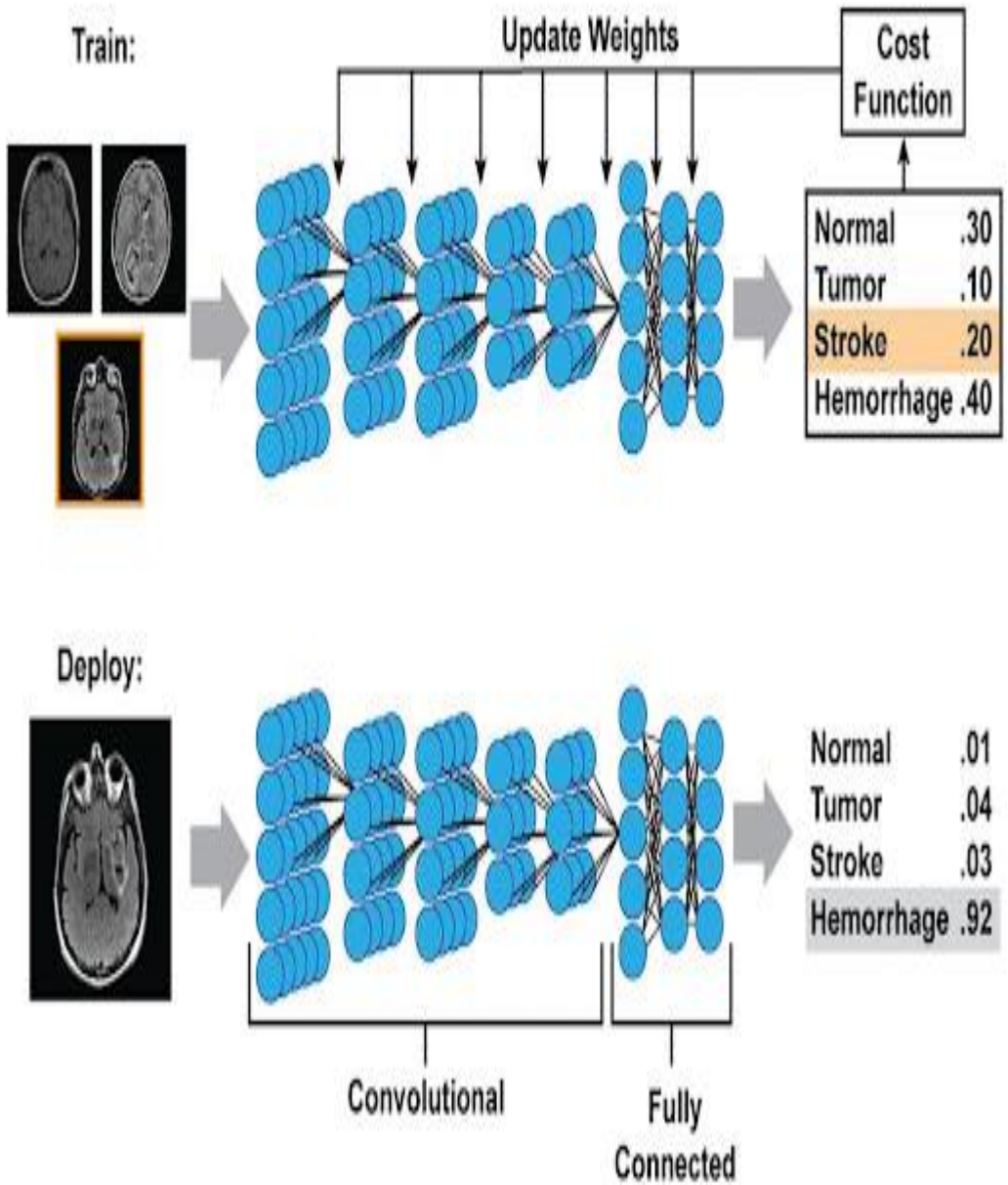
Challenges

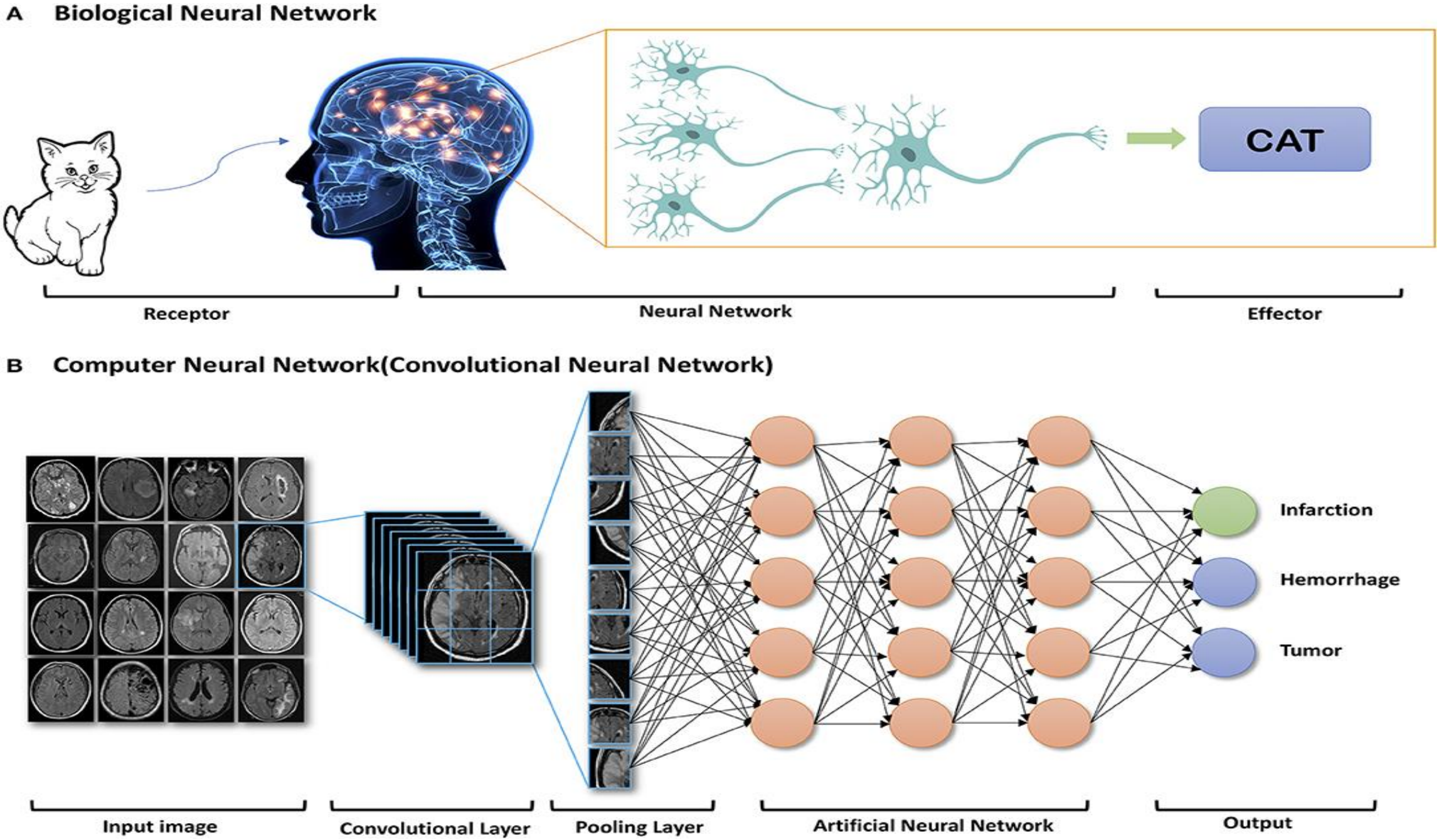
* Difficult to examine effect of suggested actions on final outcomes.

Challenges

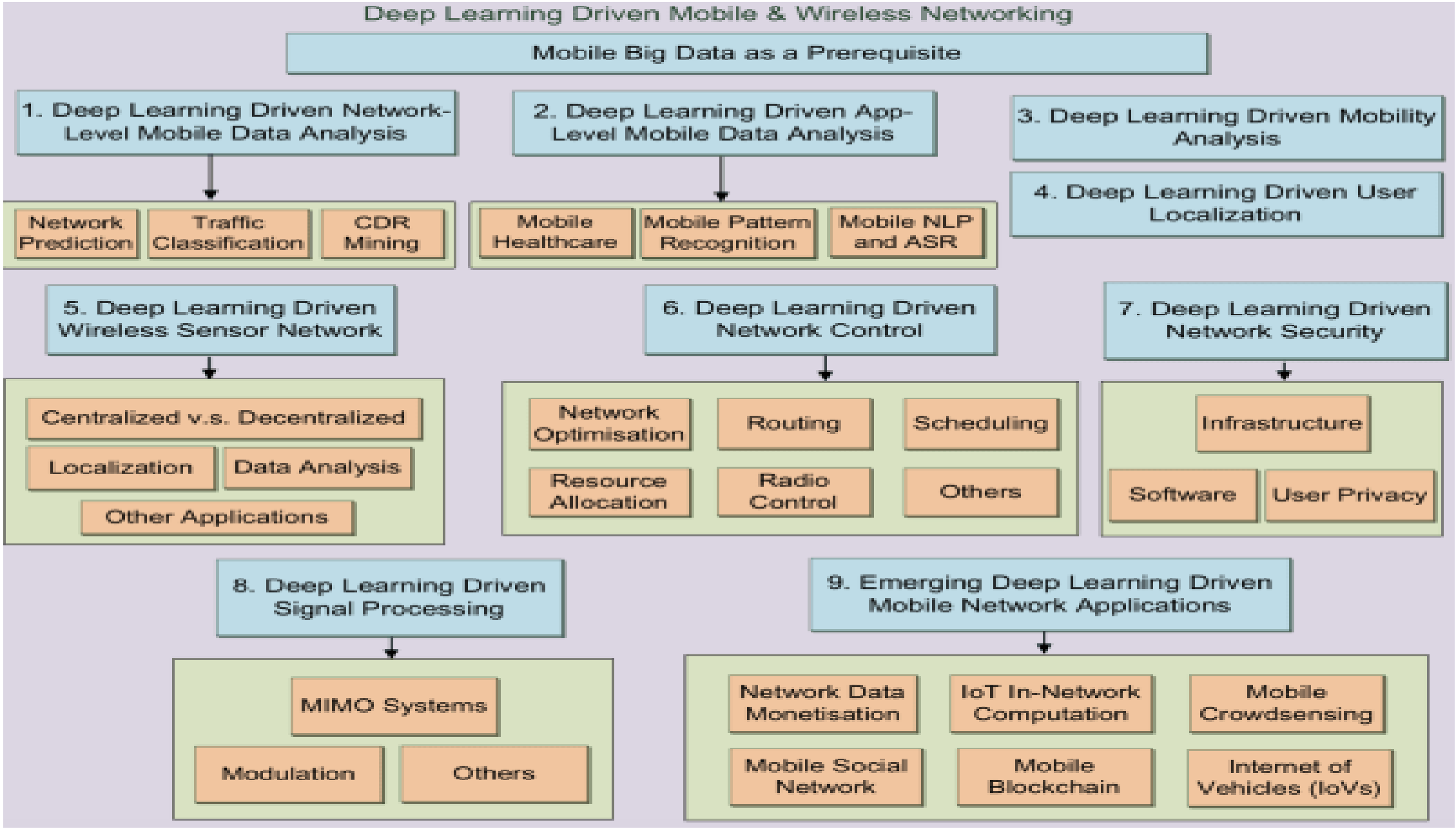
* Human physiology is partially observable, physiological and clinical data provide an incomplete picture.
* Potential issues with data quality
* Produced states may be susceptible to noise in vital signal readings

Convolutional Neural Networks

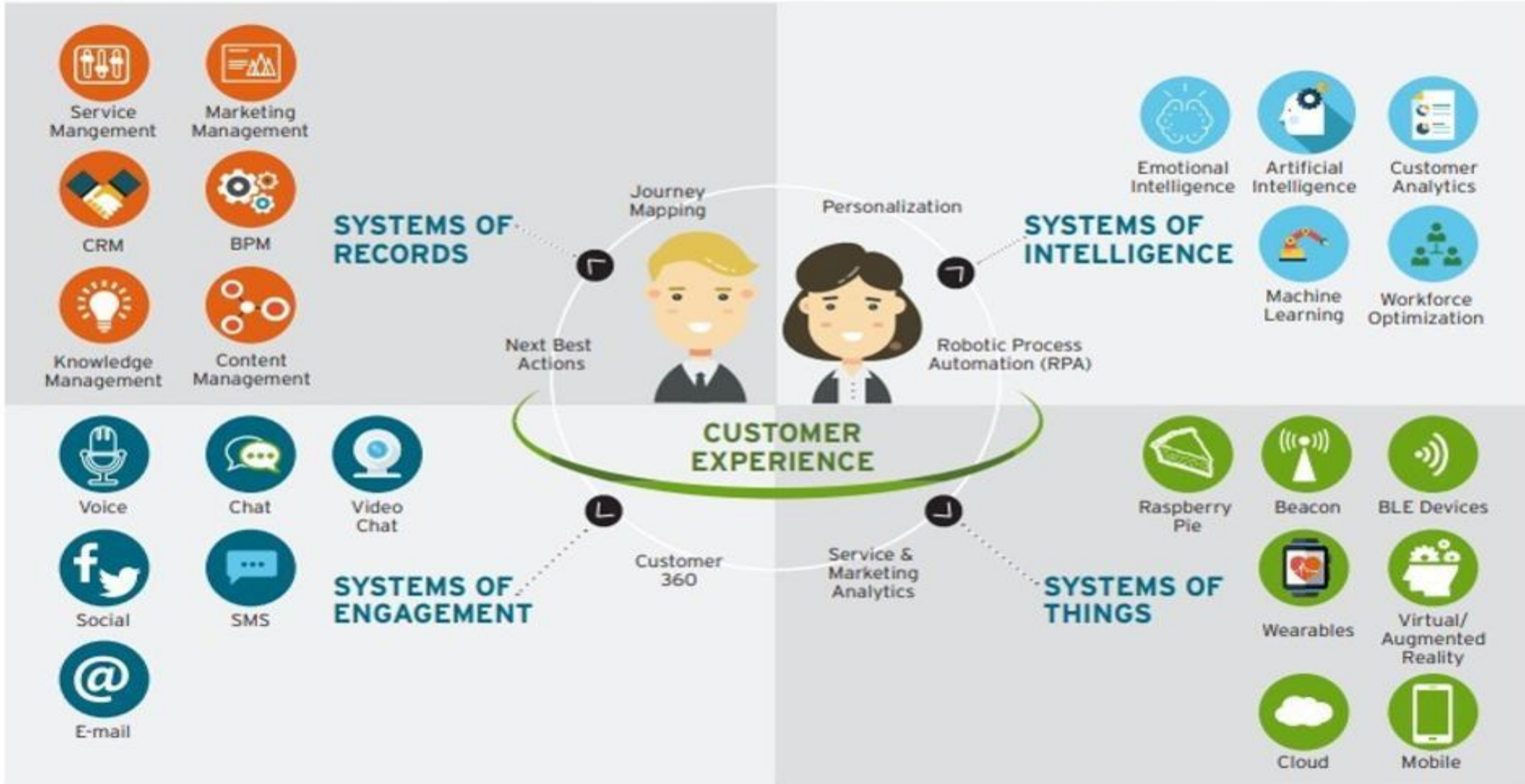



A Biological Neural Network
CAT
Receptor
Neural Network
Effector
B Computer Neural Network(Convolutional Neural Network)
Infarction
Hemorrhage
Tumor
Input image
Convolutional Layer
Pooling Layer
Artificial Neural Network
Output

APPLICATION OF ML IN TELECOM



Synchronizing Systems Around the Customer




THANK YOU
TEJAAL M
ENG19CS0334
B.Tech CSE
SEM V
F section