Module 3

Introduction to Machine Learning

Course Learning Objectives:

To impart knowledge of machine learning methods a
s applications.

Course outcome

Illustrate the concepts of Machine Learning,
pplications, and its advantages over human learning.

ontents

- troduction to Human Learning,
- ypes of Human Learning,
- achine Learning,
- ypes of Machine Learning,
- on-Machine Learning Problems,
- pplications of Machine Learning.
- extbook 2:Chapter 1
- kat Dutt, Subramanian Chandramouli, Amit Kumar
s, “Machine Learning”, Pearson Education India,2018.

Introduction to Human Learning

rning is typically referred to as the process of gaining information through
ervation.

y do we need to learn?

our daily life, we need to carry out multiple activities. It may be a task as sim
valking down the street or doing homework. Or it may be some complex tas
iding the angle in which a rocket should be launched so that it can have a
ticular trajectory.

do a task in a proper way, we need to have prior information on one or more
ngs related to the task.

o, as we keep learning more or in other words acquiring more information, t
ciency in doing the tasks keep improving

For example, with more knowledge, the ability to do homewor
with less number of mistakes increases.

In the same way, information from past rocket launches helps i
aking the right precautions and makes more successful rocket
aunches. Thus, with more learning, tasks can be performed mo
efficiently.

TYPES OF HUMAN LEARNING

nan learning happens in one of the three ways:

) either somebody who is an expert in the subject directly teac

 we build our own notion indirectly based on what we have le
om the expert in the past, or

 we do it ourselves, maybe after multiple attempts, some bein
successful.

The first type of learning, we may call, falls under the category of learning directly under expert guidance

The second type falls under learning guided by knowledge gained from experts and

The third type is learning by self or self-learning

1 Learning under expert guidance

n infant may inculcate certain traits and characteristics, learnin

raight from its guardians.

e calls his hand, a 'hand', because that is the information he ge

om his parents.

he sky is 'blue' to him because that is what his parents have ta

m. We say that the baby 'learns' things from his parents.

e next phase of life is when the baby starts going to school. In
ool, he starts with basic familiarization of alphabets and digits
en the baby learns how to form words from the alphabets and
nbers from the digits.

owly more complex learning happens in the form of sentences,
agraphs, complex mathematics, science, etc.

e baby is able to learn all these things from his teacher who
eady has knowledge on these areas.

n starts higher studies where the person learns about more
nplex, application-oriented skills.

ineering students get skilled in one of the disciplines like civil,
nputer science, electrical, mechanical, etc. medical students
rn about anatomy, physiology, pharmacology, etc.

re are some experts, in general the teachers, in the respective
d who have in-depth subject matter knowledge, who help the
dents in learning these skills.

en the person starts working as a professional in some field.

ough he might have gone through enough theoretical learning

e respective field, he still needs to learn more about the hands

plication of the knowledge that he has acquired.

e professional mentors, by virtue of the knowledge that they h

ned through years of hands-on experience, help all newcome

e field to learn on-job.

all phases of life of a human being, there is an element of guid

rning.

is learning is imparted by someone, purely because of the fact

t he/she has already gathered the knowledge by virtue of his/h

perience in that field.

o guided learning is the process of gaining information from a

rson having sufficient knowledge due to the past experience.

2 Learning guided by knowledge gained from exper

essential part of learning also happens with the knowledge w

s been imparted by teacher or mentor at some point of time i

me other form/context.

ample, a baby can group together all objects of same colour e

nis parents have not specifically taught him to do so.

is able to do so because at some point of time or other his pa

ve told him which colour is blue, which is red, which is green, grown-up kid can select one odd word from a set of words cause it is a verb and other words being all nouns.

e could do this because of his ability to label the words as verb ouns, taught by his English teacher long back.

a professional role, a person is able to make out to which stomers he should market a campaign from the knowledge ab eference that was given by his boss long back.

ll these situations, there is no direct learning. It is some past ormation shared on some different context, which is used as a ning to make decisions

.3 Learning by self

many situations, humans are left to learn on their own.

lassic example is a baby learning to walk through obstacles. H mps on to obstacles and falls down multiple times till he learn t whenever there is an obstacle, he needs to cross over it.

faces the same challenge while learning to ride a cycle as a k ve a car as an adult.

t all things are taught by others. A lot of things need to be lea y from mistakes made in the past.

e tend to form a check list on things that we should do, and th t we should not do, based on our experiences

WHAT IS MACHINE LEARNING?

lachine Learning is an application of Artificial intelligence (AI) that provides stems the ability to automatically learn and improve from experience without being explicitly programmed.

stated by Tom M. Mitchell, Professor of Machine Learning Department, Schoo
nputer Science, Carnegie Mellon University. Tom M. Mitchell has defined mach
ning as

‘A computer program is said to learn from experience E with respect to some c
of tasks T and performance measure P, if its performance at tasks in T, as meas
by P, improves with experience E.’

Vhat this essentially means is that a machine can be considere

o learn if it is able to gather experience by doing a certain task

nd improve its performance in doing the similar tasks in the

uture.

past experience, it means past data related to the task. This da

s an input to the machine from some source.

he context of learning to play checkers, E represents the experience of pla game, T represents the task of playing checkers and P is the performance nsure indicated by the percentage of games won by the player.

same mapping can be applied for any other machine learning problem, f mple, image classification problem.

ontext of image classification, E represents the past data with images hav els or assigned classes (for example whether the image is of a class cat or or a class elephant etc.), T is the task of assigning class to new, unlabelled ges and P is the performance measure indicated by the percentage of ima ectly classified.

1.3.1 How do machines learn?

basic machine learning process can be divided into three part

Data Input: Past data or information is utilized as a basis for future decision-making

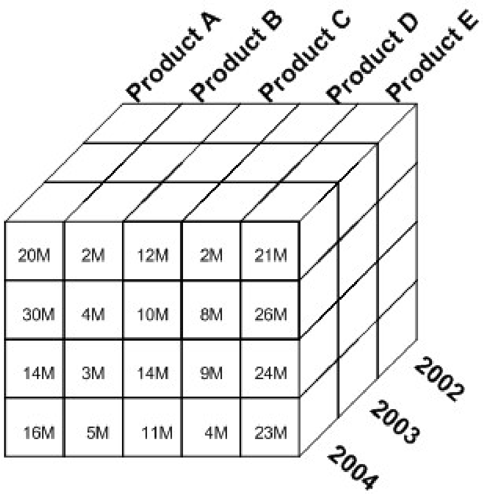
Abstraction: The input data is represented in a broader way through the underlying algorithm

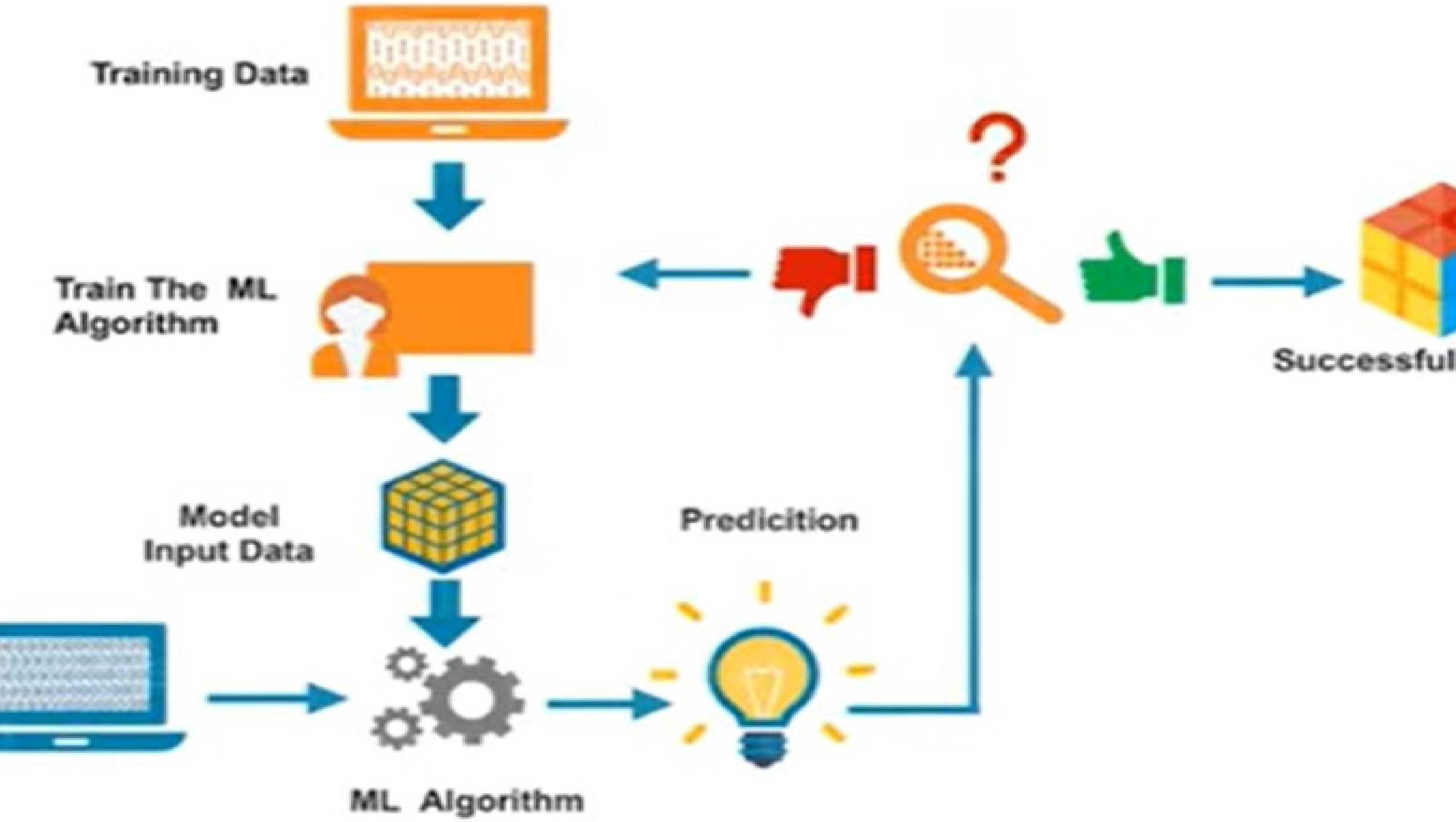
Generalization: The abstracted representation is generalized to form a framework for making decisions

Figure 1.2 is a schematic representation of the machine
arning process.

FIG. 1.2 Process of machine learning

w do machines learn...





ta input

ata is gathered from environment using sensors and/or past data
ken from dataset.

How Human Learning Process takes place : Student
Memorizing & Perfect Recall Does not help when questions are c

better learning strategy needs to be adopted:

to be able to deal with the vastness of the subject matter and the
lated issues in memorizing it

to be able to answer questions where a direct answer has not been
arnt.

gure out the key points or ideas amongst a vast pool of knowledge.

is helps in creating an outline of topics and a conceptual mapping
those outlined topics with the entire knowledge pool

vertebrate: Do not have backbones and skeletons

rtebrate

1. Fishes: Always live in water and lay eggs
2. Amphibians: Semi-aquatic i.e. may live in water or land; sm skin; lay eggs
3. Reptiles: Semi-aquatic like amphibians; scaly skin; lay eggs cold-blooded
4. Birds: Can fly; lay eggs; warm-blooded
5. Mammals: Have hair or fur; have milk to feed their young; warm-blooded

Abstraction

ring the machine learning process, knowledge is fed in the form
out data.

e data cannot be used in the original shape and form.

straction helps in deriving a conceptual map based on the input
ta.

is map, or a model is known in the machine learning paradigm, is
mmarized knowledge representation of the raw data.

e model may be in any one of the following forms

- Computational blocks like if/else rules
- Mathematical equations
- Specific data structures like trees or graphs
- Logical groupings of similar observations

e choice of the model used to solve a specific learning problem.

e **decision** related to the choice of model is taken based on multiple
ects, some of which are listed below:

e type of problem to be solved:

ether the problem is related to forecast or prediction, analysis of trend,
derstanding the different segments or groups of objects, etc.

ture of the input data:

w exhaustive (completeness) the input data is, and the data types, etc.

main of the problem:

ritical domain with a high rate of data input and need for immediate
cision making.

. fraud detection problem in banking domain.

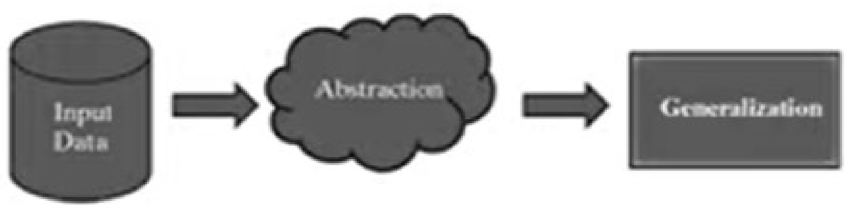
. Generalization

The abstraction process, or **training the model**, used for abstract the knowledge which comes as input data in the form of a model.

The **generalization** is, the abstracted knowledge to a form which can be used to take future decisions.

The model is trained based on a finite set of data, which may possess a limited set of characteristics.

Apply the model to take decision on a set of unknown data, usually termed as test data, then some problems occurs.



then there are two problems:

1. The trained model is aligned with the training data too much, hence may not portray the actual trend.

2. The test data have sometimes certain characteristics unknown to the training data.

TYPES OF MACHINE LEARNING

ighlighted in Figure 1.3, Machine learning can be classified int
e broad categories:

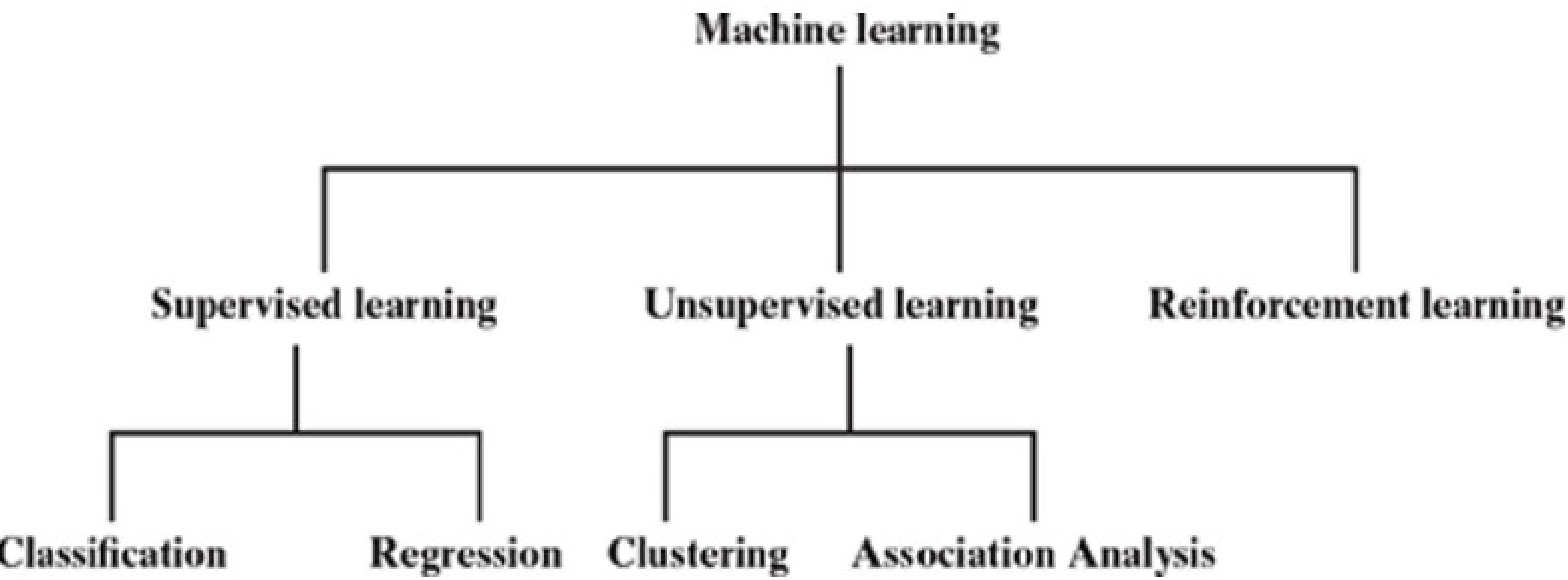


FIG. 1.3 Types of machine learning

1. Supervised learning – Also called predictive learning. A machine predicts the class of unknown objects based on prior class-related information of similar objects.

2. Unsupervised learning – Also called descriptive learning. A machine finds patterns in unknown objects by grouping simila bjects together.

3. Reinforcement learning – A machine learns to act on its owr o achieve the given goals

Supervised learning

rn from past information.

at kind of past information does the machine need for
ervised learning?

chine is getting images of different objects as input and the
k is to segregate the images by either shape or colour of the
ect.

t is by shape, the images which are of round-shaped objects
d to be separated from images of triangular-shaped objects,

he segregation needs to happen based on colour, images of
e objects need to be separated from images of green objects.

w can the machine know what is round shape, or triangular shape?
w can the machine distinguish image of an object based on whether it
e or green in colour?

achine needs the basic information to be provided to it. This basic inp
he experience in the paradigm of machine learning, is given in the forr
ning data .
ning data is the past information on a specific task. In context of the
ge segregation problem, training data will have past data on different
ects or features on a number of images, along with a tag on whether t
ge is round or triangular, or blue or green in colour.
e tag is called ‘ label’ and we say that the training data is labelled in cas
ervised learning.

re 1.4 a simple depiction of the supervised learning process

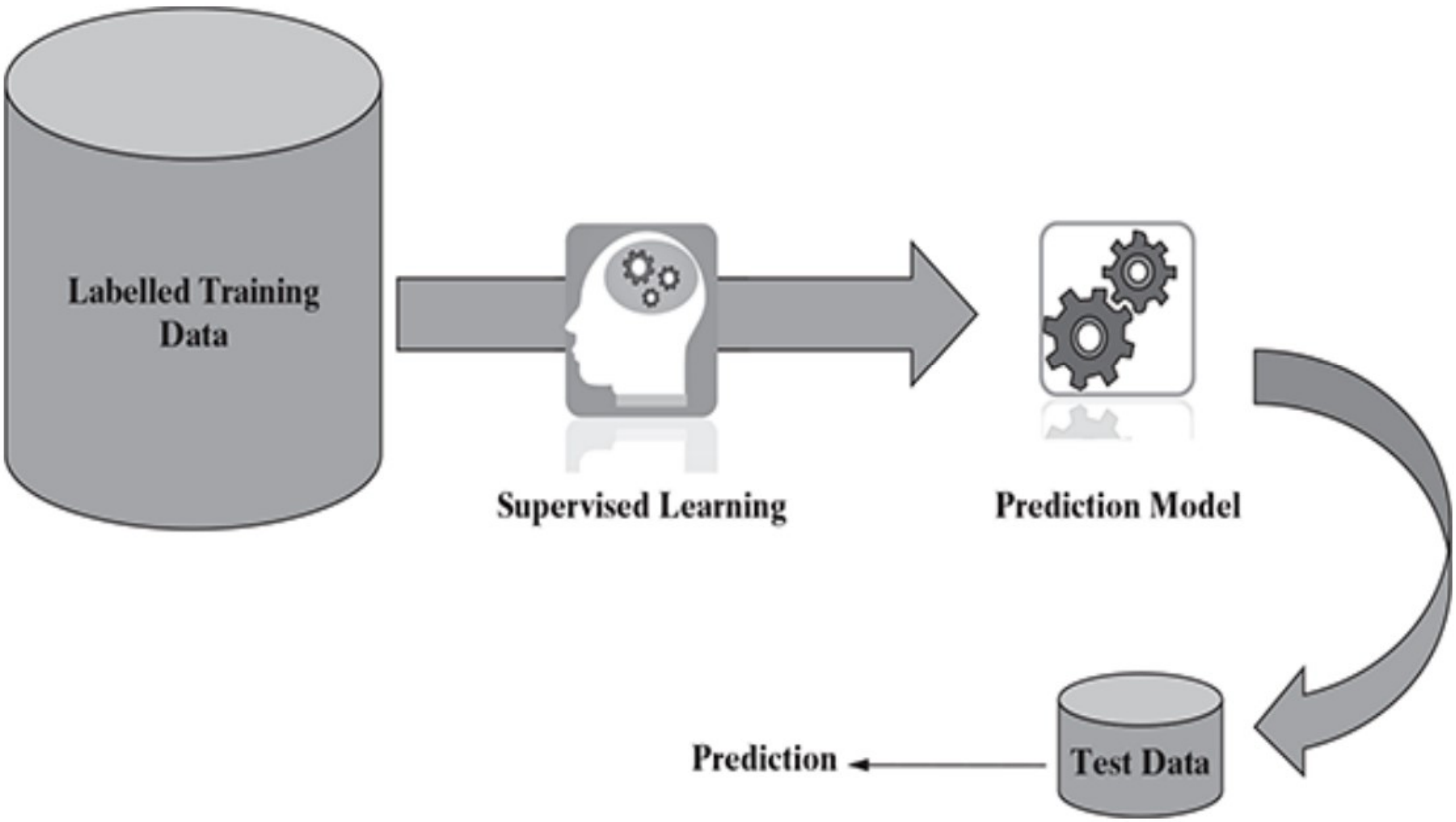


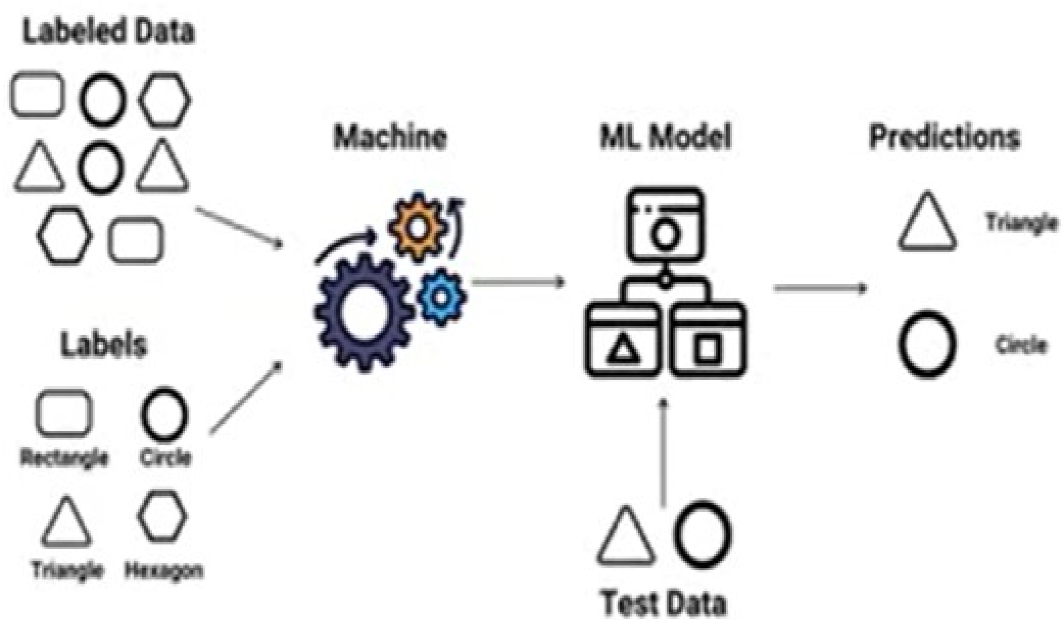
FIG. 1.4 Supervised learning

pervised machine learning is based on supervision.

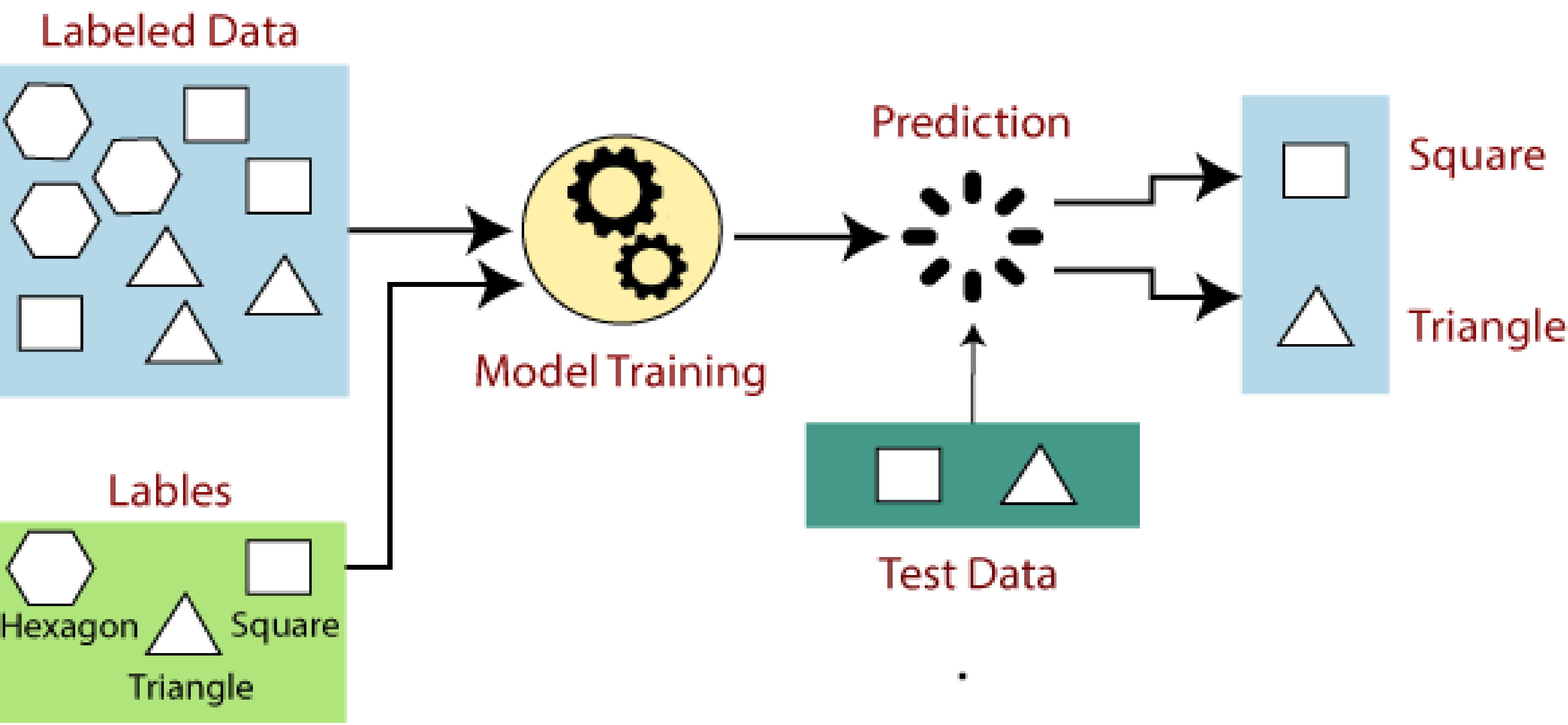
train the machines using the "labelled" dataset, and based on the training, the achine predicts the output.

e labelled data specifies that some of the inputs are already mapped to the tput.

ain the machine with the input and corresponding output, and then the achine will predict the output using the test dataset.



Example 1:


Labeled Data
Prediction
Square
Triangle
Model Training
Lables
Hexagon
Square
Triangle
Test Data

mple:


"Dog"
"Cat"


New Data
It's a Cat
</>
Model Training
Prediction
Output
Past Labeled Data

me examples of supervised learning are
redicting the results of a game
redicting whether a tumour is malignant or benign
redicting the price of domains like real estate, stocks, etc.
lassifying texts such as classifying a set of emails as spam
r non-spam

Steps Involved in Supervised Learning:

- rst Determine the type of training dataset
- ollect/Gather the labelled training data.
- plit the training dataset into training **dataset, test dataset,** **alidation dataset**.
- etermine the input features of the training dataset, which should h nough knowledge so that the model can accurately predict the outp
- etermine the suitable algorithm for the model, such as support ve achine, decision tree, etc.
- xecute the algorithm on the training dataset. Sometimes we n alidation sets as the control parameters, which are the subse aining datasets.
- valuate the accuracy of the model by providing the test set. If odel predicts the correct output, which means our model is accurat

Types of Supervised Learning

Supervised Learning

Classification

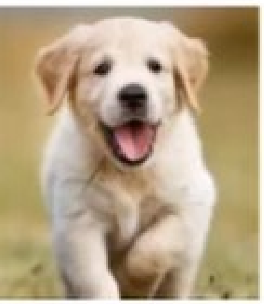
Classification is about predicting a class or discrete values
Eg: Male or Female; True or False

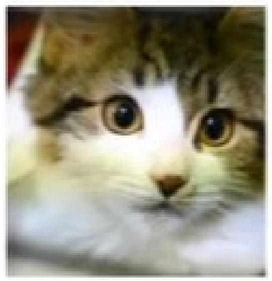
Regression

Regression is about predicti quantity or continuous valu
Eg: Salary; age; Price.

Types of Supervised Learning

ssification:





(Dog or Cat)

Dog

Cat

gression:

Rainfall in cm

mperature

Rainfall in cm

assification

Classification algorithms are used when the output variab categorical, which means there are two classes such as Yes Male-Female, True-false, etc.

Spam Filtering,

Random Forest

Decision Trees

Logistic Regression

Support vector Machines

Algorithms

lassification:

- Decision Tree Classification
- Random Forest Classification
- K-nearest Neighbor

egression:

- Logistic Regression
- Polynomial Regression
- Support Vector Machines

ression

gression algorithms are used if there is a relationship betweer e input variable and the output variable. It is used for the ediction of continuous variables, such as Weather forecasting arket Trends, etc. Below are some popular Regressior gorithms which come under supervised learning:

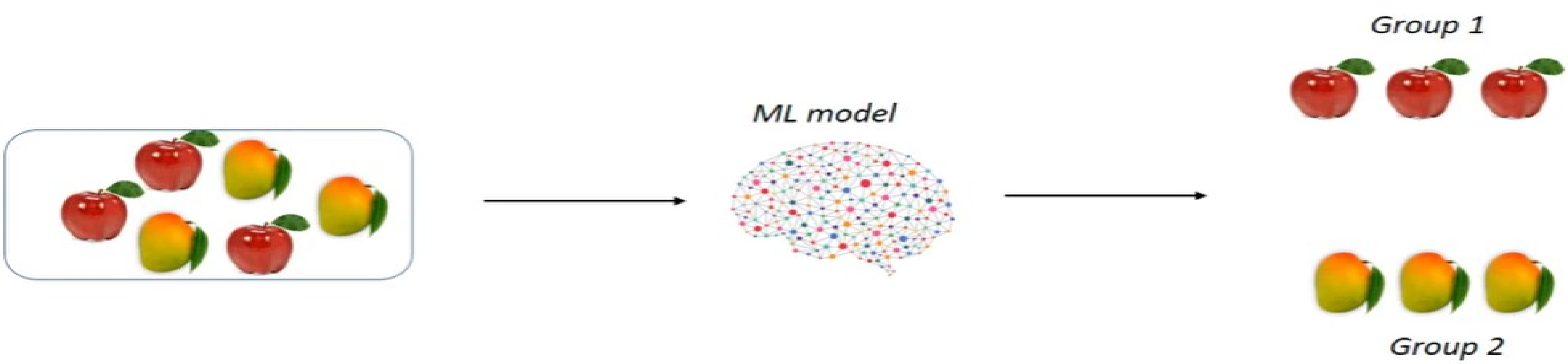
- near Regression
- gression Trees
- n-Linear Regression
- yesian Linear Regression
- lynomial Regression

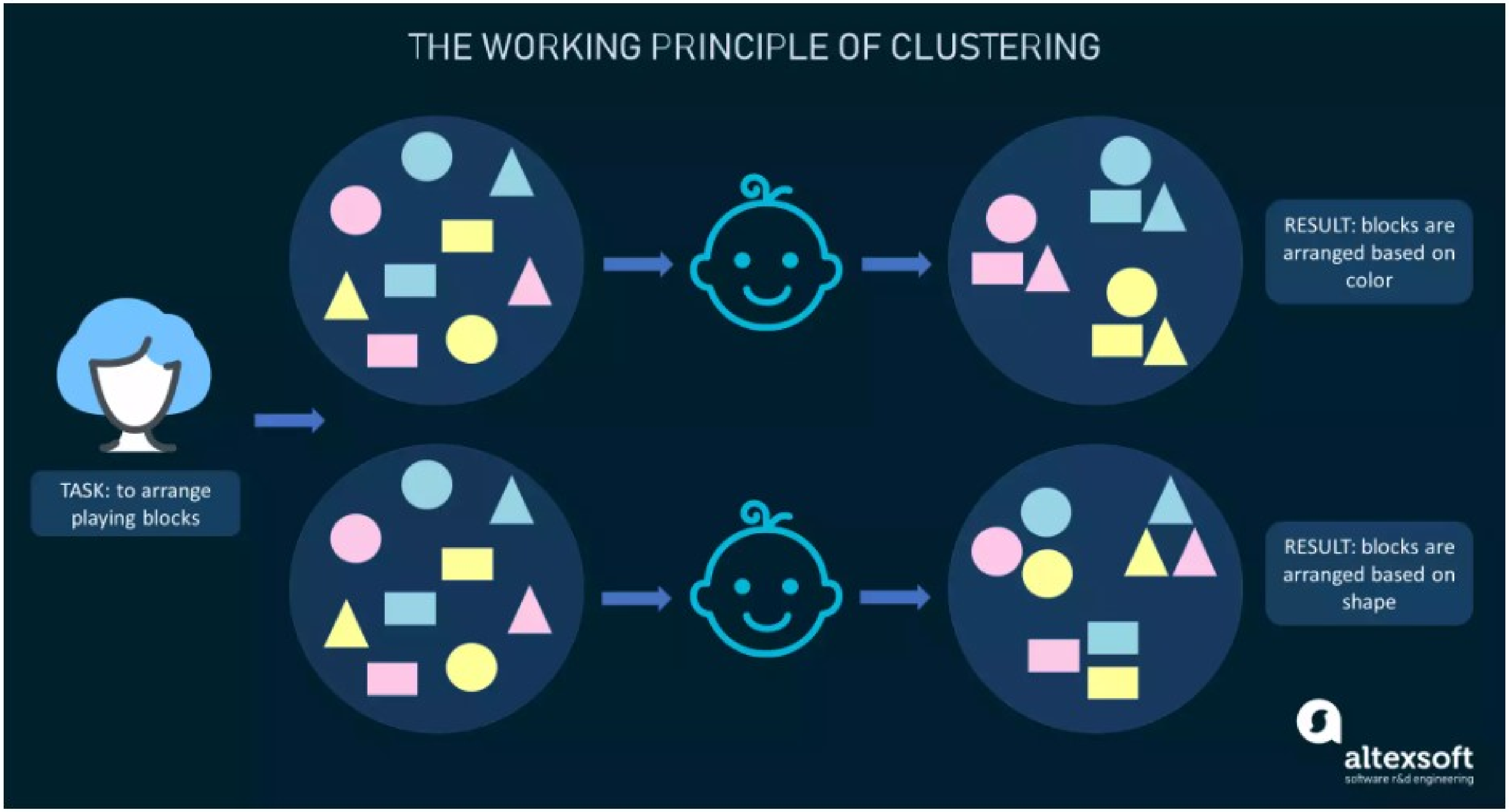
2 Unsupervised learning

nsupervised learning, the objective is to take a dataset as input and try t natural groupings or patterns within the data elements or records.

upervised learning is often termed as descriptive model and the process pervised learning is referred as pattern discovery or knowledge overy.

In Unsupervised Learning, the Machine Learning algorithm learns from Unlabelled Data




THE WORKING PRINCIPLE OF CLUSTERING
TASK: to arrange playing blocks
RESULT: blocks are arranged based on color
RESULT: blocks are arranged based on shape
altexsoft
software r&d engineering

Types of Unsupervised Learning

Unsupervised Learning

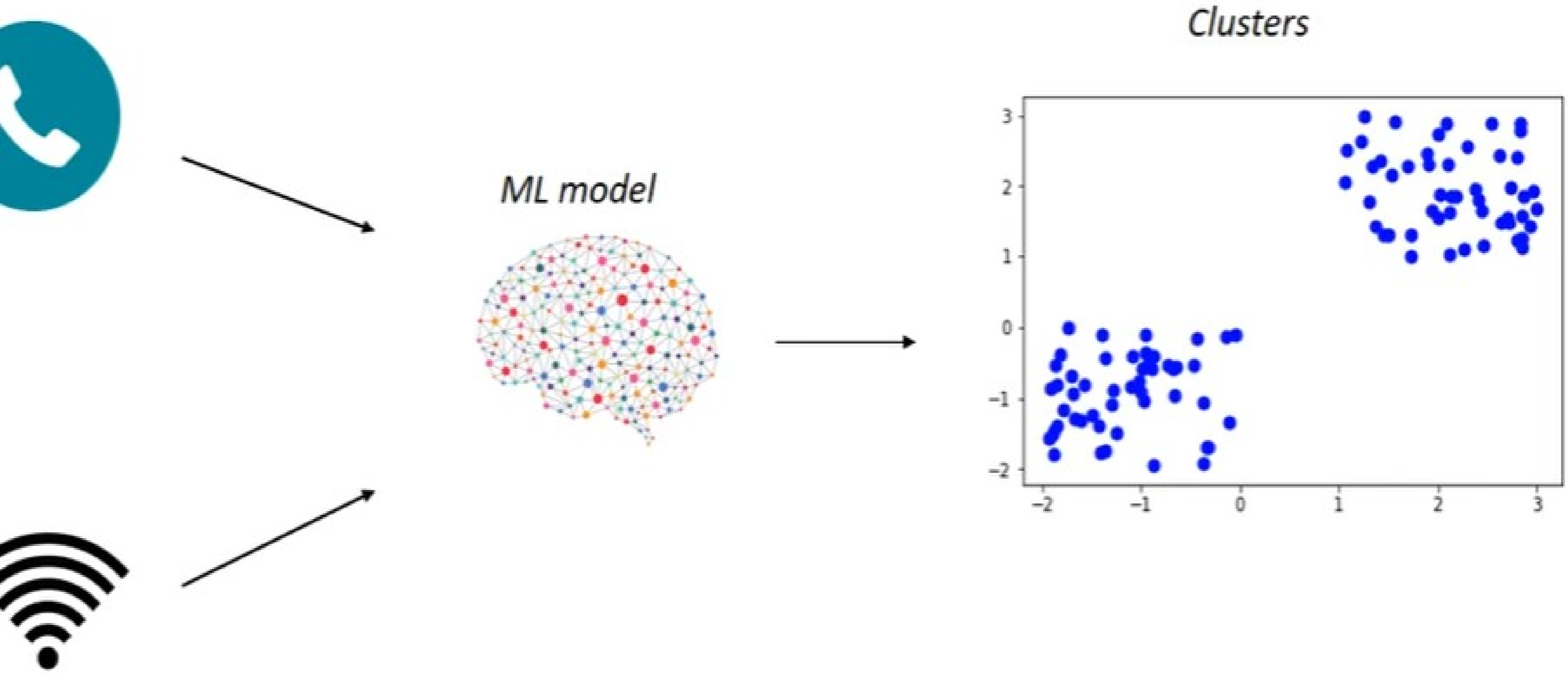
Clustering

ustering is an unsupervised
sk which involves grouping the
milar data points.

Association

Association is an unsupervise
task that is used to find
important relationship betwe
data points

Clustering



Association

Customer 1

- *Bread*
- *Milk*
- *Fruits*
- *wheat*

Customer 2

- *Bread*
- *Milk*
- *Rice*
- *Butter*

Customer 3

Now, when customer 3 goes and buys bre
it is highly likely that he will also buy milk.

4.3 Reinforcement learning

einforcement Learning is a feedback-based Machine learning techniq
hich an agent learns to behave in an environment by performing the ac
d seeing the results of actions. For each good action, the agent gets pos
edback, and for each bad action, the agent gets negative feedback
nalty.

Reinforcement Learning, the agent learns automatically using feedb
thout any labeled data

nce there is no labeled data, so the agent is bound to learn by its exper
ly.

ample of the child learning to walk.

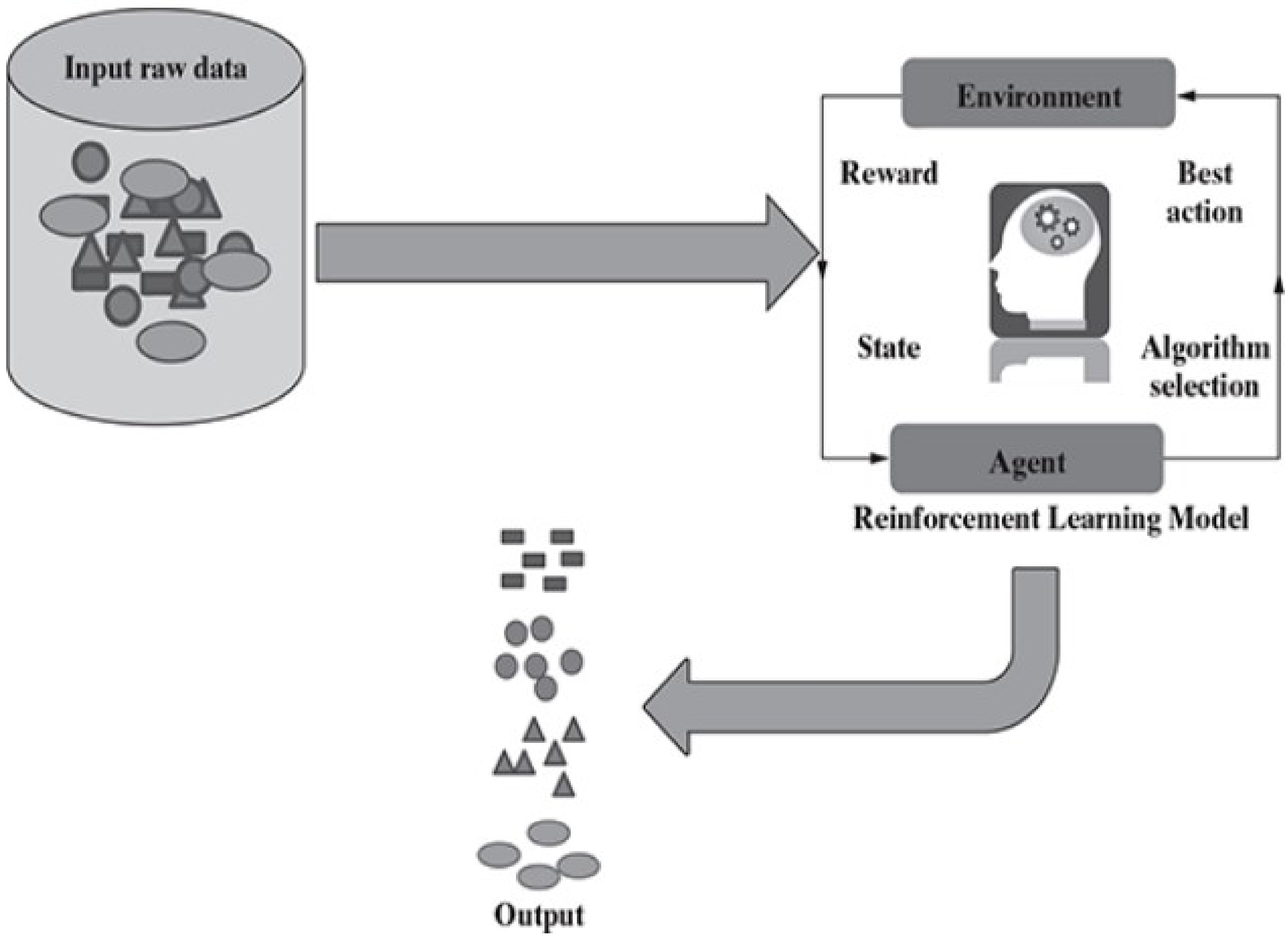
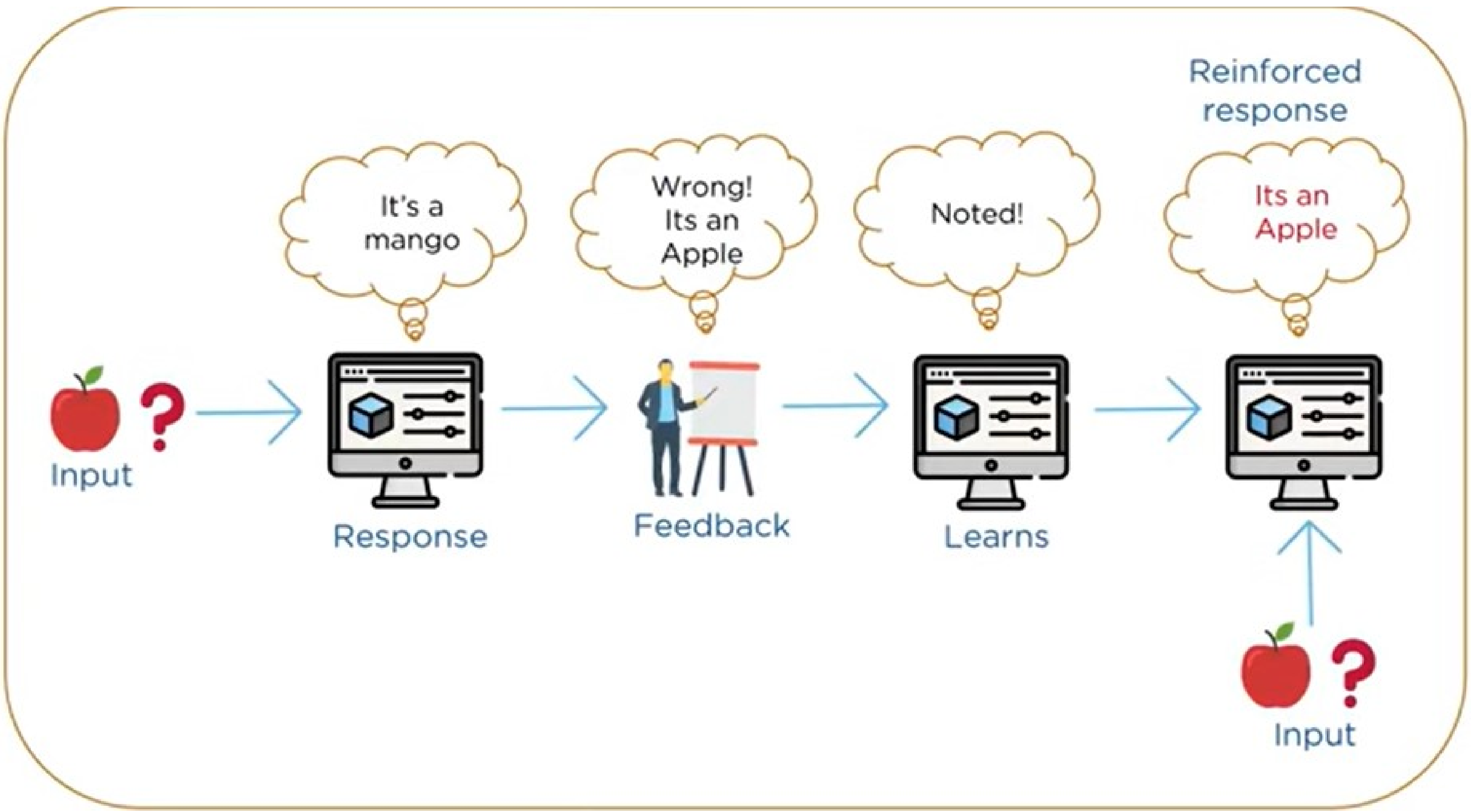


FIG. 1.10 Reinforcement learning


Reinforced
response
It's a
mango
Wrong!
Its an
Apple
Noted!
Its an
Apple
Input
Response
Feedback
Learns
Input

he agent interacts with the environment and identifies the possible actions he can

erform.

he primary goal of an agent in reinforcement learning is to perform actions by looking at

he environment and get the maximum positive rewards.

n Reinforcement Learning, the agent learns automatically using feedbacks without any

abeled data, unlike supervised learning.

ince there is no labeled data, so the agent is bound to learn by its experience only.

einforcement Learning is used to solve specific type of problem where decision making is

equential, and the goal is long-term, such as game-playing, robotics, etc.

- Reinforcement learning is the training of machine learning models to make a sequence of decisions.
- The agent learns to achieve a goal in an uncertain, potentially complex environment.
- In reinforcement learning, an agent faces a game-like situation.
- The agent employs trial and error, to come up with a solution to the problem.
- The agent gets either **rewards or penalties** for the actions it performs.
- Its goal is to maximize the total reward.
- In many complex domains, reinforcement learning is the only feasible way to train a program to perform at high levels

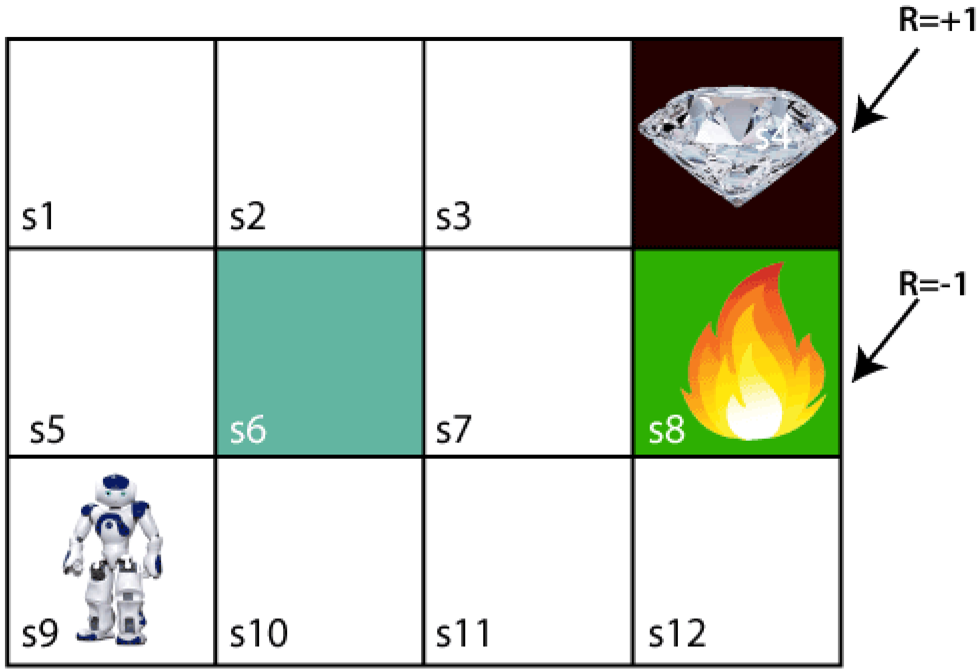
SUPERVISED	UNSUPERVISED	REINFORCEMENT
This type of learning is used when you know how to classify a given data, or in other words classes or labels are available.	This type of learning is used when there is no idea about the class or label of a particular data. The model has to find pattern in the data.	This type of learning is used when there is no idea about the class or label of a particular data. The model has to do the classification – it will get rewarded if the classification is correct, else get punished.
Labelled training data is needed. Model is built based on training data.	Any unknown and unlabelled data set is given to the model as input and records are grouped.	The model learns and updates itself through reward/ punishment.
The model performance can be evaluated based on how many misclassifications have been done based on a comparison between predicted and actual values.	Difficult to measure whether the model did something useful or interesting. Homogeneity of records grouped together is the only measure.	Model is evaluated by means of the reward function after it had some time to learn.
There are two types of supervised learning problems – classification and regression.	There are two types of unsupervised learning problems – clustering and association.	No such types.
Simplest one to understand.	More difficult to understand and implement than supervised learning.	Most complex to understand and apply.

Example

self-driving cars.

- The critical information which it needs to take care of are speed and speed limit in different road segments, traffic conditions, road conditions, weather conditions, etc. The tasks that have to be taken care of are start/stop, accelerate/decelerate, turn to left / right, etc

mples



Lorem ipsum

APPLICATIONS OF MACHINE LEARNING

- Banking and finance
- Insurance
- Healthcare

king and finance

the banking industry, fraudulent transactions, especially the ones relat
credit cards, are extremely prevalent. Since the volumes as well as
locity of the transactions are extremely high, high performance machin
rning solutions are implemented by almost all leading banks across th
obe.

e models work on a real-time basis, i.e. the fraudulent transactions are
otted and prevented right at the time of occurrence.

is helps in avoiding a lot of operational hassles in settling the disputes
at customers will otherwise raise against those fraudulent transactions

urance

rance industry is extremely data intensive. For that reason, machine learning is
nsively used in the insurance industry.

major areas in the insurance industry where machine learning is used are risk
iction during new customer onboarding and claims management.

ng customer onboarding, based on the past information the risk profile of a new
omer needs to be predicted.

ed on the quantum of risk predicted, the quote is generated for the prospective
omer.

en a customer claim comes for settlement, past information related to historic clai
g with the adjustor notes are considered to predict whether there is any possibilit
claim to be fraudulent.

er than the past information related to the specific customer, information related t
ar customers, i.e. customer belonging to the same geographical location, age grou
ic group, etc., are also considered to formulate the model.

lthcare

rable device data form a rich source for applying machine learning and predict the
th conditions of the person real time.

se there is some health issue which is predicted by the learning model, immediate
person is alerted to take preventive action.

se of some extreme problem, doctors or healthcare providers in the vicinity of the
on can be alerted. Suppose an elderly person goes for a morning walk in a park
e to his house. Suddenly, while walking, his blood pressure shoots up beyond a
ain limit, which is tracked by the wearable.

wearable data is sent to a remote server and a machine learning algorithm is
tantly analyzing the streaming data. It also has the history of the elderly person ar
ons of similar age group.

model predicts some fatality unless immediate action is taken. Alert can be sent to
person to immediately stop walking and take rest. Also, doctors and healthcare
iders can be alerted to be on standby

Decision tree

Thus, a decision tree consists of three types of nodes:

Root Node

Branch Node

Leaf Node

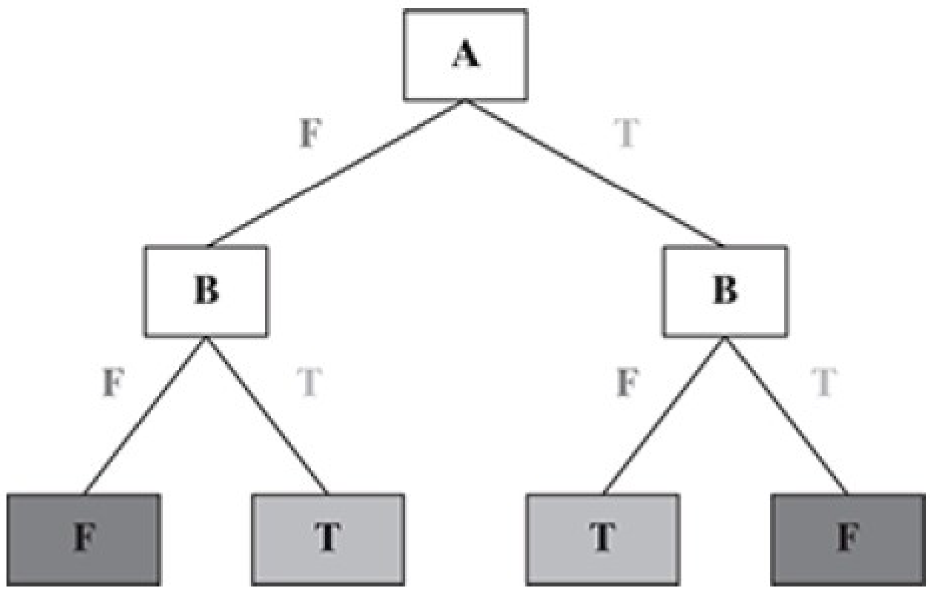


FIG. 7.8 Decision tree structure

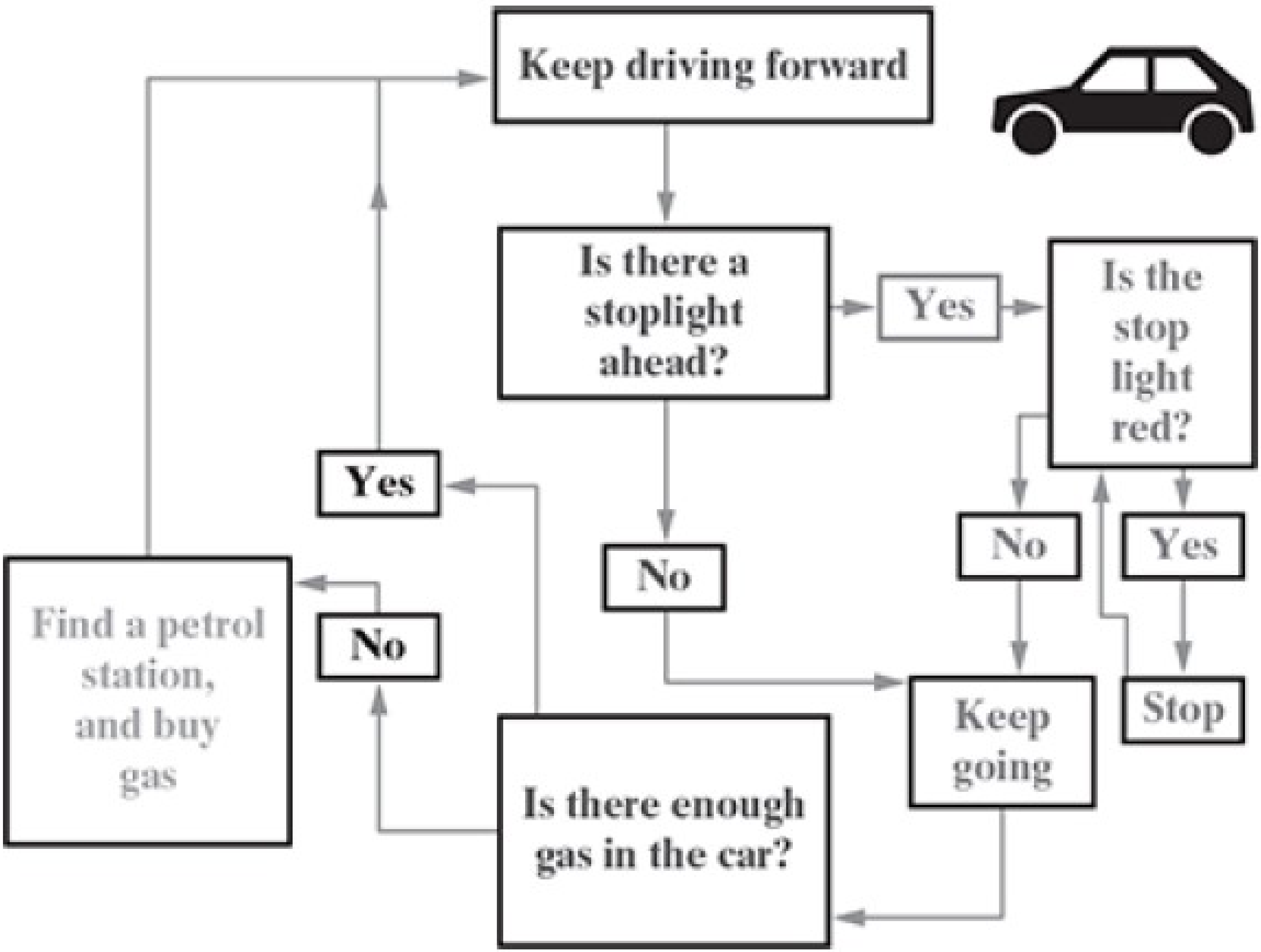

Keep driving forward
Is there a stoplight ahead?
Yes
Is the stop light red?
No
Yes
No
Yes
Keep going
Stop
Yes
No
Find a petrol station, and buy gas
Is there enough gas in the car?

FIG. 7.8 D[illegible]

Building a decision tree

Decision trees are built corresponding to the training data following an approach called recursive partitioning.

The approach splits the data into multiple subsets on the basis of the feature values.

It starts from the root node, which is nothing but the entire data set.

It first selects the feature which predicts the target class in the strongest way.

The decision tree splits the data set into multiple partitions, with data in each partition having a distinct value for the feature based on which the partitioning has happened.

This is the first set of branches.

Likewise, the algorithm continues splitting the nodes on the basis of the feature which helps in the best partition.

This continues till a stopping criterion is reached.

The usual stopping criteria are –

1. All or most of the examples at a particular node have the same class
2. All features have been used up in the partitioning
3. The tree has grown to a pre-defined threshold limit

ra : Student
High
unication – Bad;
de – High;
mming skills – Bad
fered - ?

CGPA	Communication	Aptitude	Programming Skill	Job o
High	Good	High	Good	
Medium	Good	High	Good	
Low	Bad	Low	Good	
Low	Good	Low	Bad	
High	Good	High	Bad	
High	Good	High	Good	
Medium	Bad	Low	Bad	
Medium	Bad	Low	Good	
High	Bad	High	Good	
Medium	Good	High	Good	
Low	Bad	High	Bad	
Low	Bad	High	Bad	
Medium	Good	High	Bad	
Low	Good	Low	Good	
High	Bad	Low	Bad	
Medium	Bad	High	Good	
High	Bad	Low	Bad	
Medium	Good	High	Bad	

Training Data from GTS Interview

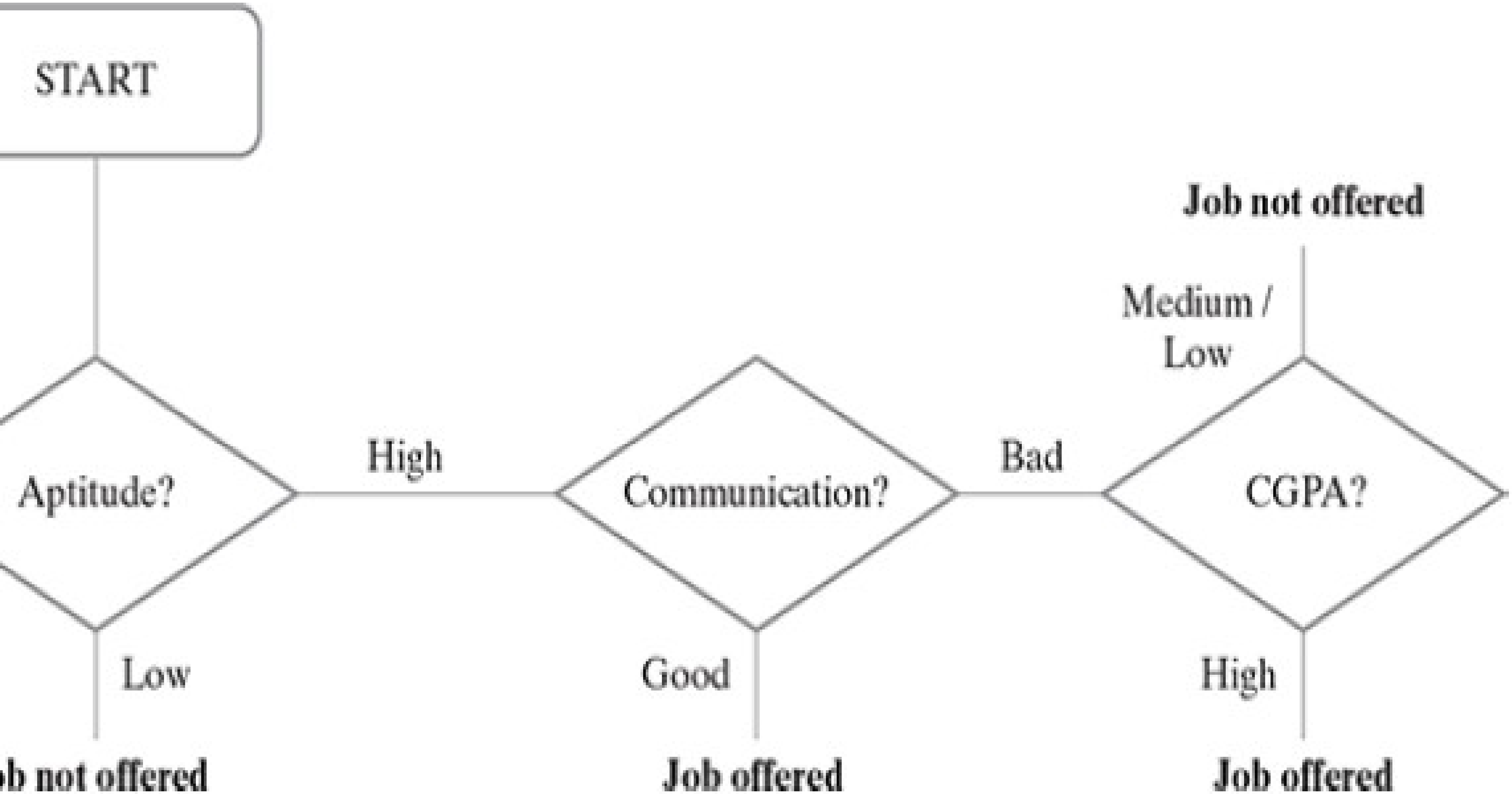


FIG. 7.11 Decision tree based on the training data

Building decision tree

The biggest challenge of a decision tree algorithm is to find out which feature to split upon.

The main driver for identifying the feature is that the data should be split in such a way that the partitions created by the split should contain examples belonging to a single class.

If that happens, the partitions are considered to be **pure.**

Entropy is a measure of impurity of an attribute or feature adopted by many algorithms such as ID3 and C5.0.

The information gain is calculated on the basis of the decrease in entropy (S) afte a data set is split according to a particular attribute (A).

Constructing a decision tree is all about finding an attribute that returns the highest information gain (i.e. the most homogeneous branches).

ntropy of a decision tree

Let us say S is the sample set of training examples. Then,
Entropy (S) measuring the impurity of S is defined as

$$\textbf{Entropy}(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

where c is the number of different class labels and p refers to the proportion of values falling into the *i*-th class label.

For example, with respect to the training data in Figure 7.10, we have two values for the target class 'Job Offered?' – Yes and No. The value of p_i for class value 'Yes' is 0.44 (i.e. 8/18) and that for class value 'No' is 0.56 (i.e. 10/18). So, we can calculate the entropy as

$$\textbf{Entropy}(S) = -0.44 \log_2(0.44) - 0.56 \log_2(0.56) = 0.99.$$

nformation gain of a decision tree

The information gain is created on the basis of the decrease in entropy (*S) after a data set is split according to a particular* attribute (*A).*

Constructing a decision tree is all about finding an attribute that returns the highest information gain (i.e. the most homogeneous branches).

If the information gain is 0, it means that there is no reduction in entropy due to split of the data set according to that particular feature.

On the other hand, the maximum amount of information gain which may happen is the entropy of the data set before the split.

Information gain for a particular feature A is calculated by the difference in entropy before a split (or S_{bs}) with the entropy after the split (S_{as}).

$$\textbf{Information Gain (S, A) = Entropy (S}_{\textbf{bs}}\textbf{)} - \textbf{Entropy (S}_{\textbf{as}}\textbf{)}$$

For calculating the entropy after split, entropy for all partitions needs to be considered. Then, the weighted summation of the entropy for each partition can be taken as the total entropy after split. For performing weighted summation, the proportion of examples falling into each partition is used as weight.

$$\textbf{Entropy}(S_{as}) = \sum_{i=1}^{n} w_i \, \text{Entropy}\,(p_i)$$

Communication	Aptitude	Programming Skill	Job offered?
Good	High	Good	Yes
Good	High	Good	Yes
Bad	Low	Good	No
Good	Low	Bad	No
Good	High	Bad	Yes
Good	High	Good	Yes
Bad	Low	Bad	No
Bad	Low	Good	No
Bad	High	Good	Yes
Good	High	Good	Yes
Bad	High	Bad	No
Bad	High	Bad	No
Good	High	Bad	Yes
Good	Low	Good	No
Bad	Low	Bad	No
Bad	High	Good	No
Bad	Low	Bad	No
Good	High	Bad	Yes

(a) Original data set:

	Yes	No	Tota
Count	8	10	18
pi	0.44	0.56	
-pi*log(pi)	0.52	0.47	0.9

Total Entropy = 0.99

$$\text{Information Gain (S, A)} = \text{Entropy } (S_{bs}) - \text{Entropy } (S_{a}$$

Splitted data set (based on the feature 'CGPA'):

PA = High

	Yes	No	Total
unt	4	2	6
	0.67	0.33	
*log(pi)	0.39	0.53	0.92

CGPA = Medium

	Yes	No	Total
Count	4	3	7
pi	0.57	0.43	
-pi*log(pi)	0.46	0.52	0.99

CGPA = Low

	Yes	No	T
Count	0	5	
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	

al Entropy = 0.69

Information Gain = 0.30

$$\text{ropy}(S_{as}) = \sum_{i=1}^{n} w_i \text{ Entropy } (p_i)$$

$$\textbf{Entropy}(S) = \sum_{i=1}^{c} - p_i \log_2$$

rforming weighted summation,

oportion of examples falling into each partition is used as weight.

Information Gain (S, A) = Entropy (S_{bs}) − Entropy (S_{a}

Splitted data set (based on the feature 'Communication'):

mmunication = Good

	Yes	No	Total
ount	7	2	9
	0.78	0.22	
i*log(pi)	0.28	0.48	0.76

tal Entropy = 0.63

Communication = Bad

	Yes	No	Tota
Count	1	8	9
pi	0.11	0.89	
-pi*log(pi)	0.35	0.15	0.50

Information Gain = 0.36

$$\text{ropy}(S_{as}) = \sum_{i=1}^{n} w_i \text{ Entropy } (p_i)$$

$$\textbf{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2$$

rforming weighted summation,

oportion of examples falling into each partition is used as weight.

Information Gain (S, A) = Entropy (S_{bs}) − Entropy (S_a

plitted data set (based on the feature 'Aptitude'):

ude = High

	Yes	No	Total
nt	8	3	11
	0.73	0.27	
og(pi)	0.33	0.51	0.85

Entropy = 0.52

Aptitude = Low

	Yes	No	Tota
Count	0	7	7
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	0.00

Information Gain = 0.47

Aptitude = Low split not be expanded sinc the rows belongs to c category of no job of

$$\text{ropy}(S_{as}) = \sum_{i=1}^{n} w_i \,\text{Entropy}\,(p_i)$$

$$\textbf{Entropy}(S) = \sum_{i=1}^{c} - p_i \log_2$$

rforming weighted summation,

oportion of examples falling into each partition is used as weight.

Information Gain (S, A) = Entropy (S_{bs}) − Entropy (S_{a}

(e) Splitted data set (based on the feature 'Programming Skill'):

Programming Skill = Good

	Yes	No	Total
Count	5	4	9
pi	0.56	0.44	
-pi*log(pi)	0.47	0.52	0.99

Total Entropy = 0.95

Programming Skill = Bad

	Yes	No	Total
Count	3	6	9
pi	0.33	0.67	
-pi*log(pi)	0.53	0.39	092

Information Gain = 0.04

$$\text{ropy}(S_{as}) = \sum_{i=1}^{n} w_i \text{ Entropy } (p_i)$$

$$\textbf{Entropy}(S) = \sum_{i=1}^{c} - p_i \log_2$$

rforming weighted summation,
oportion of examples falling into each partition is used as weight.

$$\text{Entropy}(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

(a) Original data set:

	Yes	No	Total
Count	8	10	18
pi	0.44	0.56	
-pi*log(pi)	0.52	0.47	0.99

Total Entropy = 0.99

l data set (based on the feature 'CGPA'):

h

Yes	No	Total
4	2	6
0.67	0.33	
0.39	0.53	0.92

py = 0.69

CGPA = Medium

	Yes	No	Total
Count	4	3	7
pi	0.57	0.43	
-pi*log(pi)	0.46	0.52	0.99

Information Gain = 0.30

CGPA = Low

	Yes	No	Total
Count	0	5	5
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	0.00

ted data set (based on the feature 'Programming Skill'):

ming Skill = Good

	Yes	No	Total
	5	4	9
	0.56	0.44	
pi)	0.47	0.52	0.99

ropy = 0.95

Programming Skill = Bad

	Yes	No	Total
Count	3	6	9
pi	0.33	0.67	
-pi*log(pi)	0.53	0.39	092

Information Gain = 0.04

(c) Splitted data set (based on the feature 'Communication'):

Communication = Good

	Yes	No	Total
Count	7	2	9
pi	0.78	0.22	
-pi*log(pi)	0.28	0.48	0.76

Total Entropy = 0.63

Communication = Ba

	Yes	N
Count	1	8
pi	0.11	0.
-pi*log(pi)	0.35	0.

Information Gain = 0.

(d) Splitted data set (based on the feature 'Aptitude'):

Aptitude = High

	Yes	No	Total
Count	8	3	11
pi	0.73	0.27	
-pi*log(pi)	0.33	0.51	0.85

Total Entropy = 0.52

Aptitude = Low

	Yes	N
Count	0	7
pi	0.00	1.0
-pi*log(pi)	0.00	0.0

Information Gain = (

Information Gain (S, A) = Entropy (S_{bs}) − Entro

$$\text{Entropy}(S_{as}) = \sum_{i=1}^{n} w_i \text{ Entropy } ($$

le = High

A	Communication	Programming Skill	Job offered?
	Good	Good	Yes
n	Good	Good	Yes
	Good	Bad	Yes
	Good	Good	Yes
	Bad	Good	Yes
n	Good	Good	Yes
	Bad	Bad	No
	Bad	Bad	No
n	Good	Bad	Yes
n	Bad	Good	No
n	Good	Bad	Yes

(a) Level 2 starting set:

	Yes	No	T
Count	8	3	
pi	0.73	0.27	
-pi*log(pi)	0.33	0.51	(

Total Entropy = 0.85

de = Low split need not be expanded since rows belongs to one category of no job d

$$\text{Entropy}(S) = \sum_{i=1}^{c} - p_i \log$$

Information Gain (S, A) = Entropy (S_{bs}) − Entropy (S_{a}

) Splitted data set (based on the feature 'CGPA'):

GPA = High

	Yes	No	Total
unt	4	0	4
	1.00	0.00	
'log(pi)	0.00	0.00	0.00

CGPA = Medium

	Yes	No	Total
Count	4	1	5
pi	0.80	0.20	
-pi*log(pi)	0.26	0.46	0.72

CGPA = Low

	Yes	No	Tota
Count	0	2	2
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	0.00

al Entropy = 0.33

Information Gain = 0.52

$$\text{ropy}(S_{as}) = \sum_{i=1}^{n} w_i \text{ Entropy } (p_i)$$

$$\text{Entropy}(S) = \sum_{i=1}^{c} - p_i \log$$

rforming weighted summation,

oportion of examples falling into each partition is used as weight.

Splitted data set (based on the feature 'Communication'):

ommunication = Good

	Yes	No	Total
ount	7	0	7
	1.00	0.00	
i*log(pi)	0.00	0.00	0.00

tal Entropy = 0.30

Communication = Bad

	Yes	No	Total
Count	1	3	4
pi	0.25	0.75	
-pi*log(pi)	0.50	0.31	0.81

Information Gain = 0.55

) Spitted data set (based on the feature 'Programming Skill'):

ogramming Skill = Good

	Yes	No	Total
ount	5	1	6
	0.83	0.17	
i*log(pi)	0.22	0.43	0.65

tal Entropy = 0.80

Programming Skill = Bad

	Yes	No	Total
Count	3	2	5
pi	0.60	0.40	
-pi*log(pi)	0.44	0.53	0.97

Information Gain = 0.05

Aptitude = High & Communication = Bad

CGPA	Programming Skill	Job offered?
High	Good	Yes
Low	Bad	No
Low	Bad	No
Medium	Good	No

(a) Level 2 starting set:

	Yes	No	Total
Count	1	3	4
pi	0.25	0.75	
-pi*log(pi)	0.50	0.31	0.81

Total Entropy = 0.81

itted data set (based on the feature 'CGPA'):

= High

	Yes	No	Total
	1	0	1
	1.00	0.00	
g(pi)	0.00	0.00	0.00

ntropy = 0.00

CGPA = Medium

	Yes	No	Total
Count	0	1	1
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	0.00

Information Gain = 0.81

CGPA = Low

	Yes	No
Count	0	2
pi	0.00	1.00
-pi*log(pi)	0.00	0.00

itted data set (based on the feature 'Programming Skill'):

mming Skill = Good

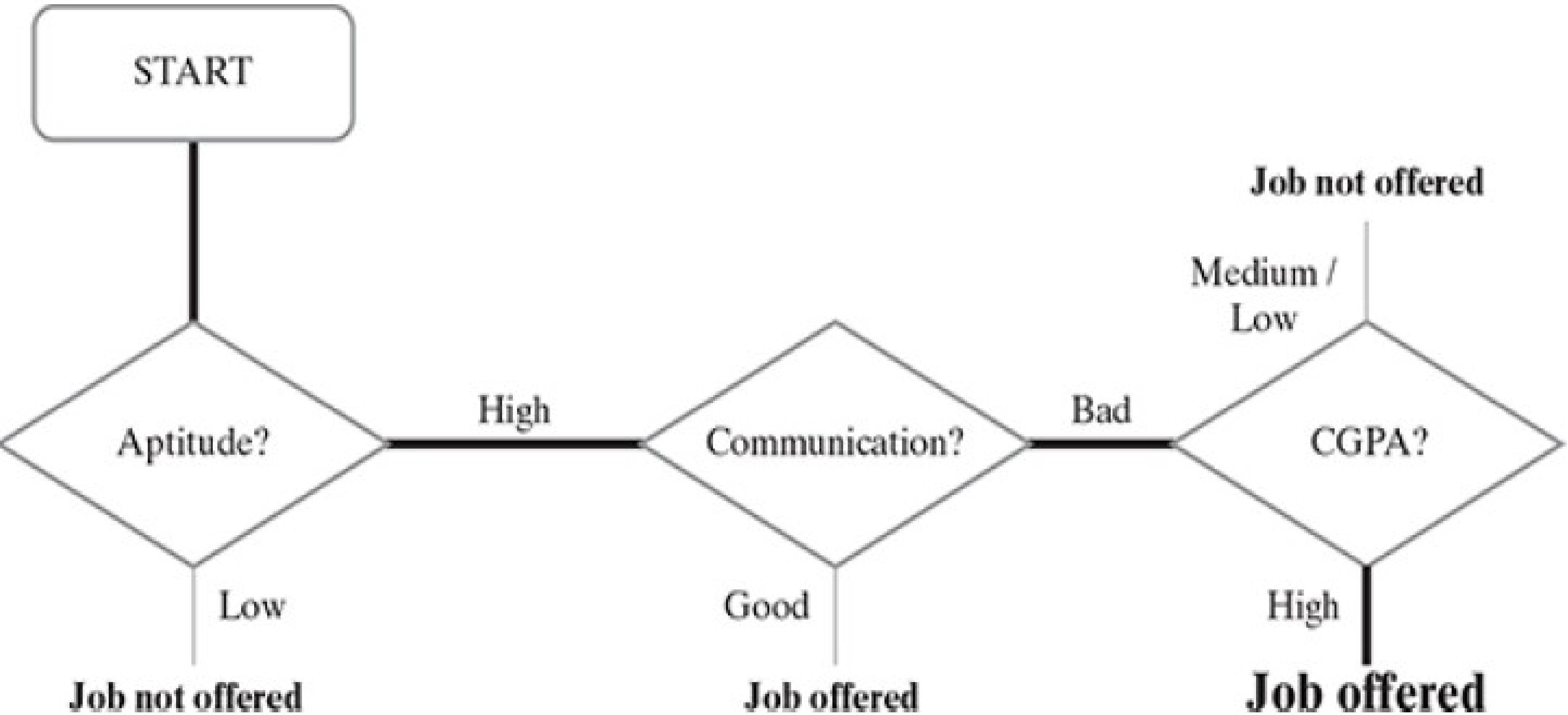
	Yes	No	Total
	1	1	2
	0.50	0.50	
g(pi)	0.50	0.50	1.00

ntropy = 0.50

Programming Skill = Bad

	Yes	No	Total
Count	0	2	2
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	0.00

Information Gain = 0.31


START
Aptitude?
High
Low
Job not offered
Communication?
Bad
Good
Job offered
CGPA?
Medium / Low
Job not offered
High
Job offered

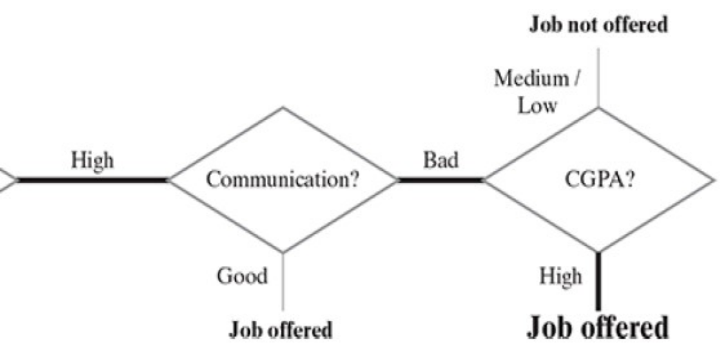
High & Communication = Bad

Programming Skill	Job offered?
Good	Yes
Bad	No
Bad	No
Good	No

starting set:

	Yes	No	Total
	1	3	4
	0.25	0.75	
	0.50	0.31	0.81

py = 0.81



(b) Splitted data set (based on the feature 'CGPA'):

CGPA = High

	Yes	No	Total
Count	1	0	1
pi	1.00	0.00	
-pi*log(pi)	0.00	0.00	0.00

CGPA = Medium

	Yes	No	Total
Count	0	1	1
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	0.00

CGPA = Low

	Yes	N
Count	0	2
pi	0.00	1.0
-pi*log(pi)	0.00	0.0

Total Entropy = 0.00

Information Gain = 0.81

(c) Splitted data set (based on the feature 'Programming Skill'):

Programming Skill = Good

	Yes	No	Total
Count	1	1	2
pi	0.50	0.50	
-pi*log(pi)	0.50	0.50	1.00

Programming Skill = Bad

	Yes	No	Total
Count	0	2	2
pi	0.00	1.00	
-pi*log(pi)	0.00	0.00	0.00

Total Entropy = 0.50

Information Gain = 0.31