1) Explain the concept of K-means clustering with suitable Example?

K-Means clustering is an unsupervised machine learning Algorithm used for partitioning a dataset into a set of distinct, non-overlapping groups or clusters. The goal of the algorithm is to minimize the variance within each cluster and maximize the variance Between clusters, grouping similar data points together based on their features.

How K-Means clustering works:

→ Initialization
→ Assignment step
→ update step
→ update Centroids
→ Repeat
→ Result

Example of K means clustering

Data points {2,4,10,12,3,20,30,11,25}

Use K means clustering to cluster the following data in to two groups

Data points {2,4,10,12,3,20,30,11,25}

The distance function is used is Euclidean distance

Initial cluster centroid are: $M1 = 4, M2 = 11$

initial Centroids:

M1 : 4

M2 : 11

Initial Centroids:

M1: 4

M2: 11

Therefore

C1 = {2, 4, 3}

C2 = {10, 12, 20, 30, 11, 25}

New Centroids:

M1: 3

M2: 18

Data Points	Distance to M1	Distance to M2	cluster	New cluster
2	2	9	C1	
4	0	7	C1	
10	6	1	C2	
12	8	1	C2	
3	1	8	C1	
20	16	9	C2	
30	26	19	C2	
11	7	0	C2	
25	21	14	C2	

Current Centroids

M1: 3

M2: 18

Therefore

C1 = {2, 4, 10, 3}

C2 = {12, 20, 30, 11, 25}

New Centroids

M1: 4.75

M2: 19.6

Data Points	Distance to M1	Distance to M2	cluster	New cluster
2	1	16	C1	C1
4	1	14	C1	C1
10	7	8	C2	C1
12	9	6	C2	C2
3	0	15	C1	C1
20	17	2	C2	C2
30	27	12	C2	C2
11	8	7	C2	C2
25	22	7	C2	C2

Current Centroids

M1: 4.75

M2: 19.6

Therefore

C1 = {2, 4, 10, 11, 12, 3}

C2 = {20, 30, 25}

New Centroids

M1: 7

M2: 25

Data Points	Distance to M1	Distance to M2	cluster	New cluster
2	2.75	17.6	C1	C1
4	0.75	15.6	C1	C1
10	5.25	9.6	C1	C1
12	7.25	7.6	C2	C1
3	1.75	16.6	C1	C1
20	15.25	6.4	C2	C2
30	25.25	10.4	C2	C2
11	6.25	8.6	C2	C1
25	20.25	5.4	C2	C2

Current centroids:

M1: 7
M2: 25

final cluster are:

C1 = {2, 4, 10, 11, 12, 3}
C2 = {20, 30, 25}

Data Points	Distance to		cluster	New cluster
	M1	M2		
2	5	23	C1	C1
4	3	21	C1	C1
10	3	15	C1	C1
12	5	13	C1	C1
3	4	22	C1	C1
20	13	5	C2	C2
30	23	5	C2	C2
11	4	14	C1	C1
25	18	0	C2	C2

2. Discuss about the Random forest Algorithm.

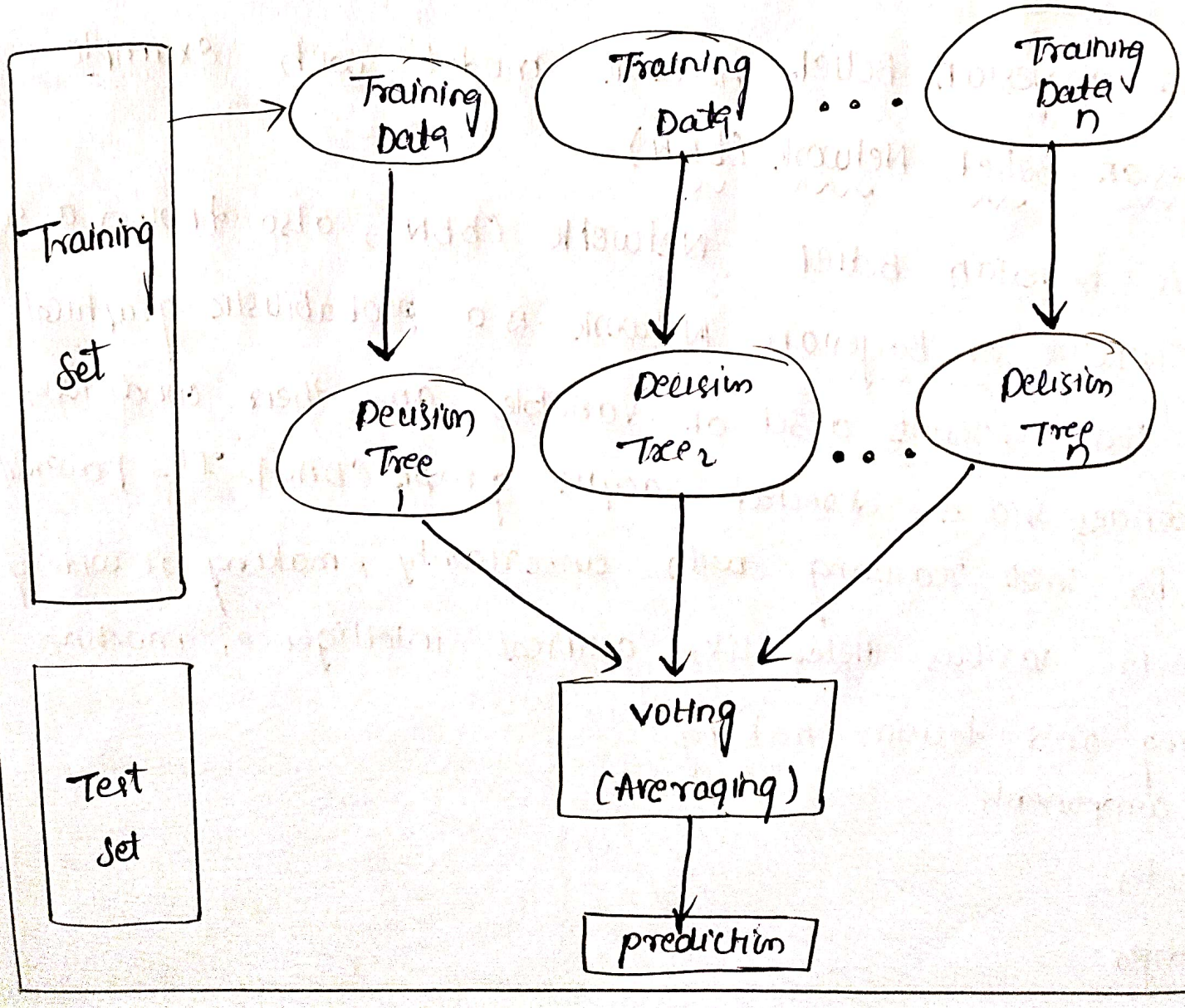
ns: we can Understand the working of Random forest algorithm. with the help of following steps-

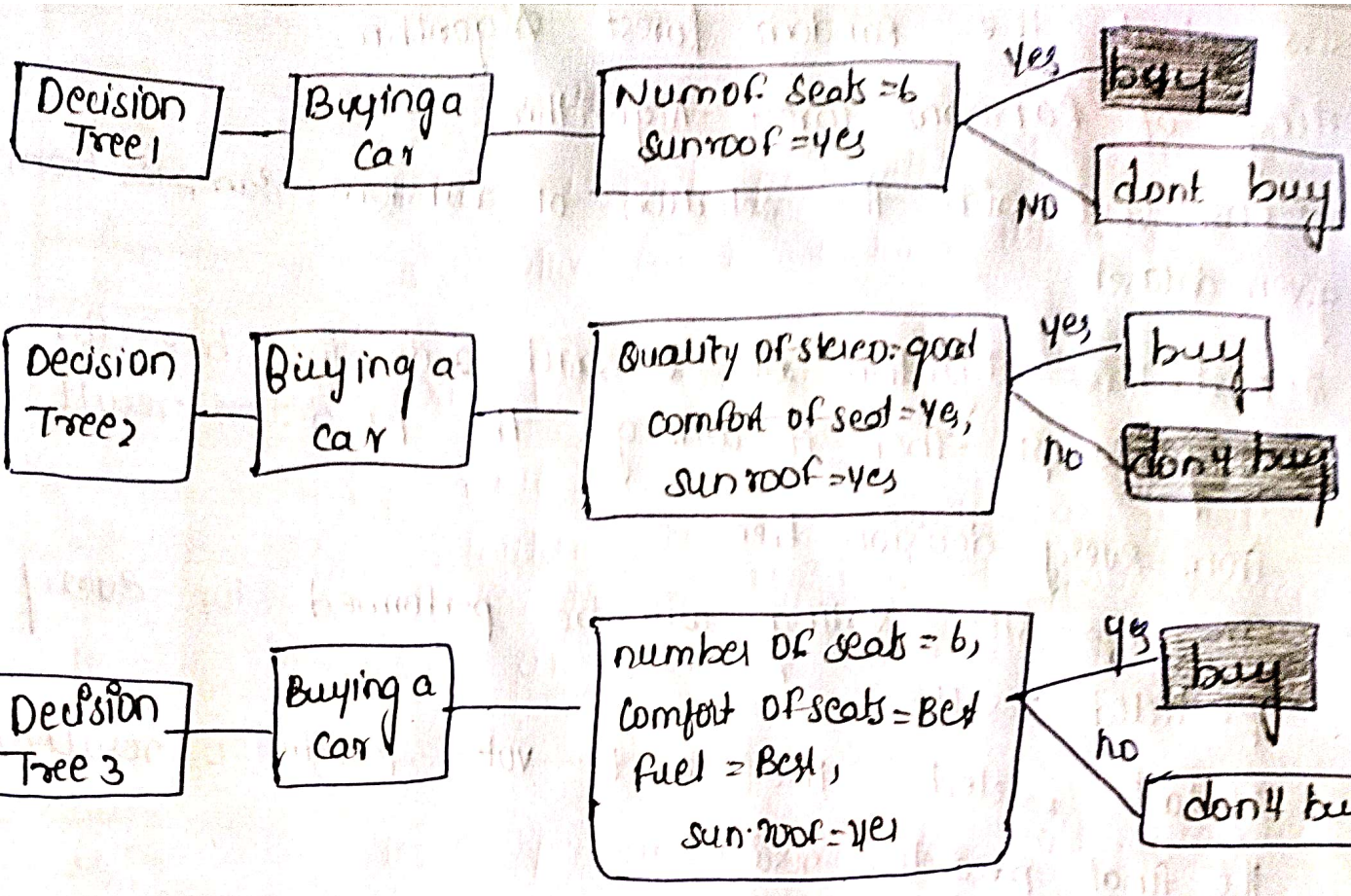
Step 1 - first, start with the selection of random samples from a given dataset.

Step 2 - Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

Step 3 - In this step, voting will be performed for every predicted result.

Step 4 - At last, select the most voted prediction result as the final prediction result.


Training Set
Training Data 1
Training Data 2
Training Data n
Decision Tree 1
Decision Tree 2
Decision Tree n
Voting (Averaging)
prediction
Test Set



3) Explain Bayesian Belief network model with example

* Bayesian Belief Network (BBN)

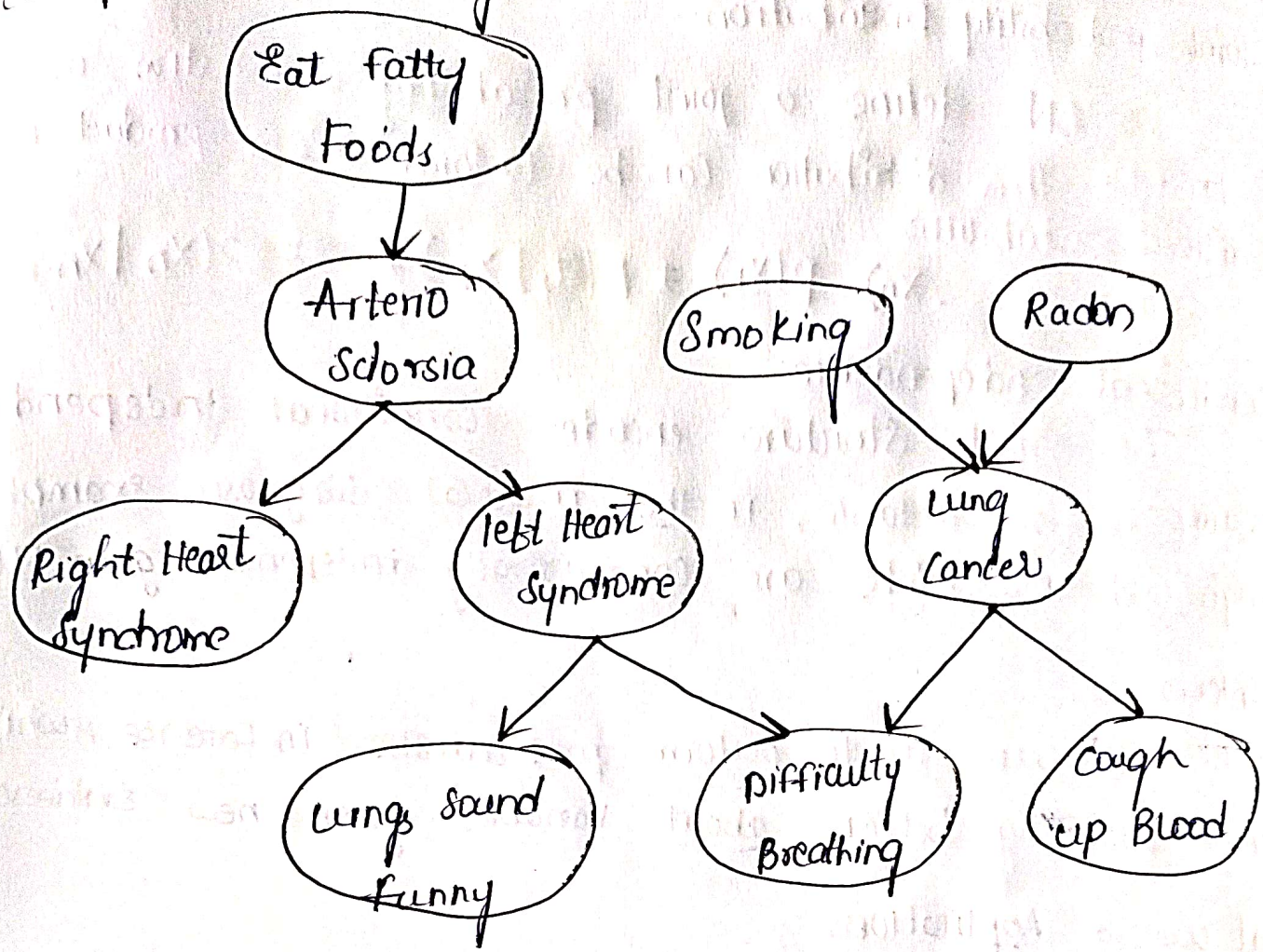
A Bayesian Belief Network (BBN), also known as a Bayes Network or Bayesian Network, is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). It's powerful tool for intell reasoning with uncertainty, making it widely used in various fields like artifical intelligence, machine learning and decision making.

key components

1) Nodes:

2) Edges:

Example: Medical Diagnosis



In this network:

Nodes:

* Fever
* FLU
* malaria
* Cough
* MuscleAche

Edges:

* Flu directly influnces Fever, cough and MuscleAche.
* Malaria directly influces fever and muscle-Ache.

How it works:

1) Joint probability Distribution:

A BBN defines a joint probability distribution overall all variables. This distribution can be factored into a product of conditional probabilities

$$P(X_1, X_2 \ldots X_n) = P(X_1) * P(X_2 | X_1) * \ldots * P(X_n | X_{n-1} \ldots X_1)$$

2) Conditional Independence:

The BBN structure encodes conditional Independence assumptions. For instance, in the medical diagnosis example, Cough and muscleAche are conditionally independent given FLU.

3) Interface:

BBNs allows us to perform probabilistic inference, which involves updating beliefs about variables given new evidence.

Real world Applications:

→ medical Diagnosis

→ Fault Diagnosis

→ Information Retrieval

→ Decision making