Date __/__/____

Saathi

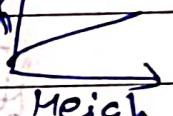
Machine learning

- Supervised
  - Regression
    - ① Linear ② Polynomial
  - Decision Tree
  - Random Forest
  - Classification
    - ① KNN
    - ② Trees
    - ③ Logistic Regression
    - ④ Naive Bayes
    - ⑤ SVM
    - ⑥ Apriori
- UnSupervised
  - Clustering
    - ① SVD
    - ② PCA
    - ③ K-means
- Rainforcement
  - ① Continuous
  - ② Categorical

Date ___ / ___ / _____

Saathi

(1) Linear Regression :- It shows the linear relationship between the dependent and independent variables and shows how the dependent variable (y) changes according to the independent variable (x). It tries to best fit a line between a dependent and independent variables, and this best fit line is known as the regression line.



(2) Logistic Regression :-

(i) This is used to predict the categorical variables or discreate values. It can be used for the classification problems in ml and the output of the logistic regression algorithm can be either Yes or No or 0 or 1, Red or Blue etc.

(ii) Logistic Regression is similar to the linear regression except how they are used such as Linear Regression is used to solve the regression problem and predict continuous values and

Saathi

Date __/__/____

whereas Logistic regression is used to solve the classification problem and used to predict the ~~continuous~~ discrete values.

(ii) Instead of fitting the best fit line, it forms as S-shaped curve that lies between 0 and 1. The S-shaped curve is also known as logistic function that uses the concept of the threshold. Any values above the threshold will tend to 1, and below the threshold will tend to 0.

(B) Decision Tree Algorithm :-

(i) This is a supervised algorithm that is mainly used to solve the classification problems but can also be used for solving the regression problems. It can work with both categorical variables and continuous variables. It

Date __/__/____

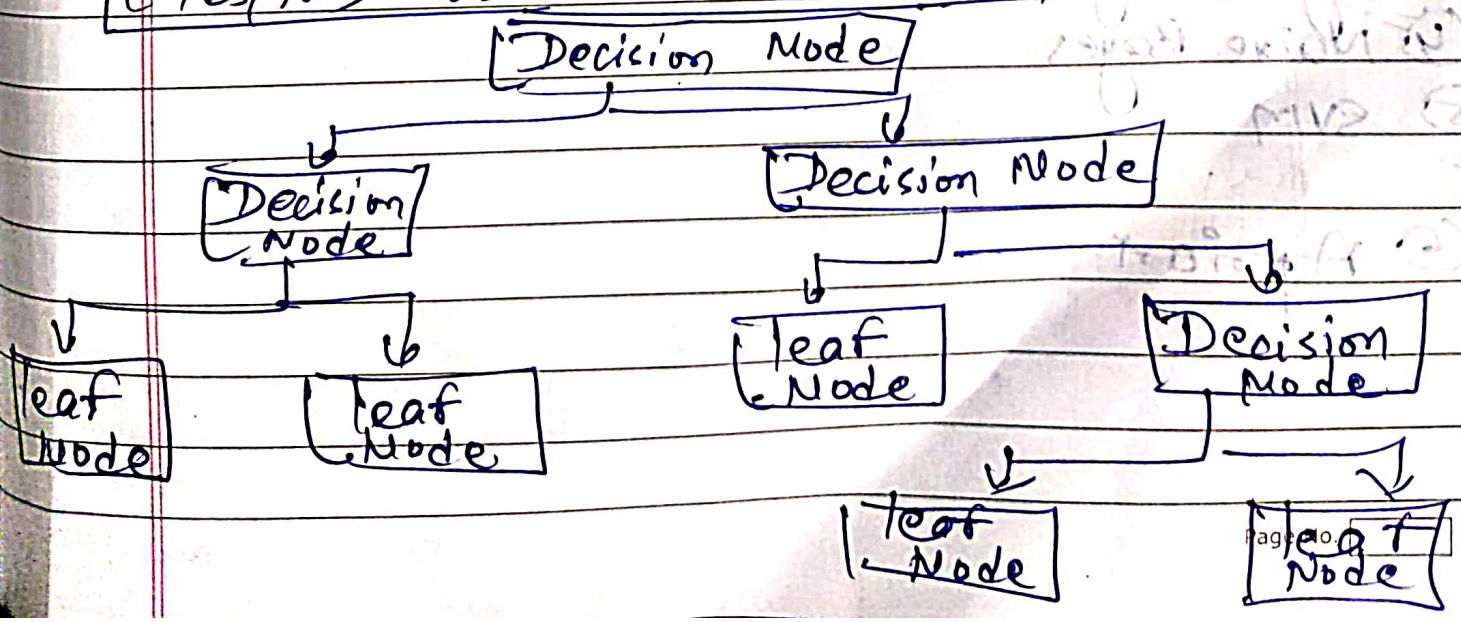
Saathi

shows a Tree-like structure that includes nodes and branches, and start with the root node that expand on further branches till the leaf node. The internal node is used to represent the features of the dataset, branches shows the decision rules, and leaf nodes represent the outcomes of the problem.

(ii) We use CART Algorithm which stands for Classification and Regression Tree.

(iii) A decision Tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

NOTE:- A decision Tree can contain categorical data (Yes/No) as well as numerical data.



Saathi

Date __/__/____

Q^Qu Why use Decision Trees?

Ans (1) Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.

(2) The logic behind the decision tree can be easily understood because it shows a tree-like structure.

⇒ Attribute Selection Measures :-

(i) Information Gain

(ii) Gini Index

(i) Information Gain :-

NOTE ⇒ Entropy ⇒ Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. Entropy can be calculated as :-

$$\text{Entropy}(S) = -P(\text{yes}) \log_2 P(\text{yes}) - P(\text{no}) \log_2 P(\text{no})$$

Page No.

Date __/__/____

Saathi

Where S = Total numbers of samples
P(yes) = Probability of yes
P(no) = ———— of No.

⇒(1) Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.
(2) It calculates how much information a feature provides us about a class.
(3) According to the value of information gain, we split the node and build the decision tree.
(4) A decision tree algorithm always tries to maximize the value of information gain, and a node/attribute having the highest information gain is split first. It can be calculated using by this:-

Information Gain = Entropy(S) - [(Weighted Avg)* Entropy(each feature)]

(ii) **Gini Index :-**

(i) Gini Index is a measure of impurity or purity used while creating a decision tree in the CART.

(ii) An attribute with the low Gini index should be preferred as compared to the high Gini Index.

(iii) It only creates binary splits, and the CART algorithm uses the Gini Index to create binary splits. Gini index can be calculated using the formula -

$$\text{Gini Index} = 1 - \Sigma_J P_J^2$$

⇒

Advantages of D.T.	Disadvantages of D.T.
(i) It is simply to understand as it follows the same process which a human follows while making any decision in real-life.	(i) The decision tree contains lots of layers, which makes it complex.
(ii) It can be very useful for solving decision-related problems.	(ii) It may have an overfitting issue, which can be resolved using the Random forest algorithm.

Date ___/___/_____

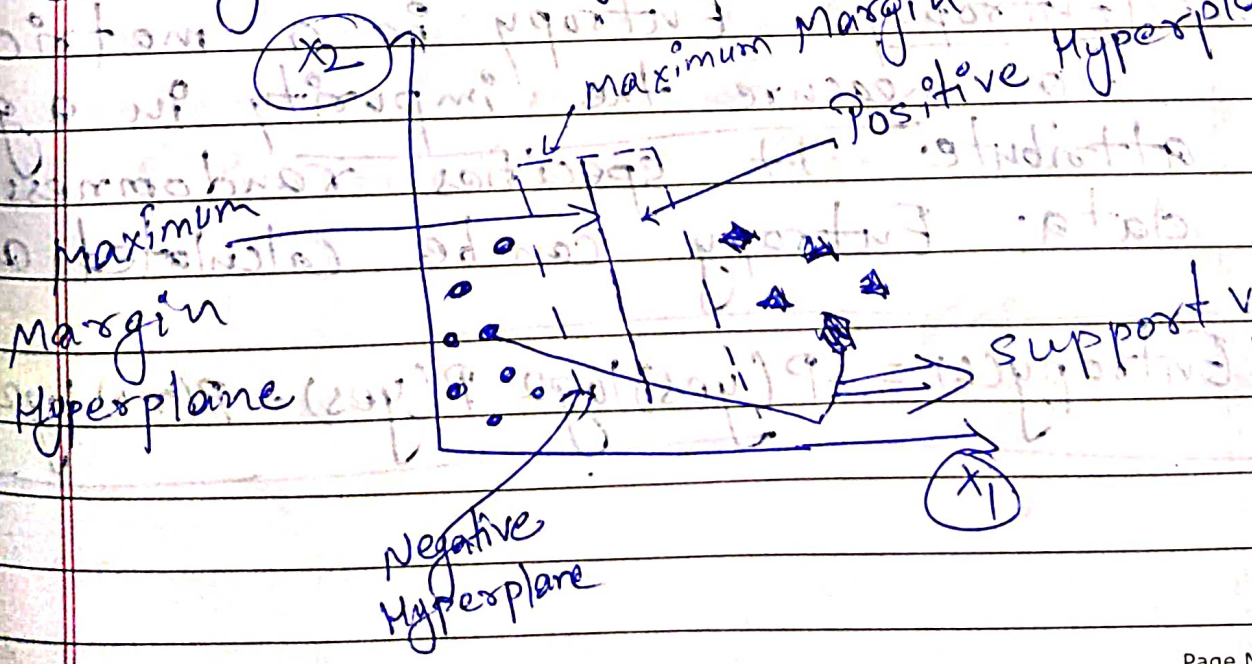
Saathi

(4) Support Vector Machine Algorithms :-

This is Supervised M.L. Algorithm that can also be used for classification and regression problems. However, it is primarily used for classification problems. The goal of SVM is to create a hyperplane or decision boundrey that can segregate datasets into different classes.

The data points that help to define the hyperplane are known as support vectors, and hence it is named as support vector machine algorithm.

Ex- Some real life applications of SVM are face detection, image classification, Drug discovery.



Page No. ____

Buti — Pr — (AI) 25.30

Saathi

Date __/__/____

⇒ Types of SVM :-

① **Linear SVM** :- It is used for linearly separable data, which means if a dataset can be classified into two classes by using a straight line, then such data is termed as linearly Separable data, and classifier is used called as Linear SVM classifier.

② **Non-linear SVM** :- It is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

⇒ Hyperplane and support vectors in SVM :-

Hyperplane :- There can be multiple line/decision boundries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary

Date __/__/____

Saathi

that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are two features, then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimensional plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

3 **Support Vectors** ⇒ The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since, these vectors support the hyperplane, hence called a Support vector.

Page No.

Saathi

Date ___/___/_____

(5) Naive Bayes Algorithm :-

(i) Naive Bayes classifier is a supervised learning algorithm, which is used to make predictions based on the probability of the object. The algorithm named as Naive Bayes as it is based on Bayes theorem, and follows the naive assumption that says variables are independent of each other.

The Equation for Bayes theorem is given as:-

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

(ii) It is mainly used in text classification that includes a high-dimensional training dataset.

(iii) ~~Naive~~ This is one of the simple and most effective classification algorithms ~~based~~ which helps in building the.

Page No.

Date ___/___/_____ (Saathi)

fast machine learning models that can make quick predictions.

Why is it called Naive Bayes?

(i) Naive ⇒ It is called Naive because it assumes that the occurances of ~~this~~ a certain feature is independent of the occurance of the other features. If the fruit is identified on the bases of color, shape and taste, then red, spherical and sweet fruit is recognized as an apple. Hence each ~~fruit~~ feature individually contributes to identify that it is an apple without depending on each other.

(ii) Bayes ⇒ It is called bayes because it depends on the principal of Bayes' theorem.

Saathi

Date ___/___/_____

⇒ **Bayes' Theorem** :- This theoram is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

$P(A|B)$ ⇒ is Posterior probability ⇒ Probability of hypothesis A on the observed event B.

$P(B|A)$ ⇒ Probability of the evidance given that the probability of hypothesis is true.

$P(A)$ ⇒ Prior Probability ⇒ Probability of hypothesis before observing the evidance.

$P(B)$ ⇒ Marginal Probability ⇒ Probability of Evidance.

Date ___/___/_____

Saathi

Working of Naive Bayes' Classifier :-

⇒ We have a dataset of weather conditions and corresponding ~~that~~ target variable 'Play'. So, using this dataset we need to decide that whether we should play or not on a perticular day according to the weather conditions. So, to solve this problem, we need to follow the steps :-

1. Convert the given datasets into frequency tables.
2. Generate Likelihood table by finding the probabilities of given features.
3. Now, use Bayes theorem to calculate the posterior probability.

⇒

Outlook	Play
Rainy	Yes
Sunny	Yes
Overcast	Yes
Sunny	No
Rainy	Yes
Rainy	No
Over Cast	Yes

Saathi

Date ___/___/_____

Step-1 Frequency table for Weather Conditions:-

Weather	Yes	NO
Overcast	5	0
Rainy	2	2
Sunny	3	2
Total	10	5

Step-2 Likelihood table weather Conditions:-

Weather	NO	Yes	
Overcast	0	5	5/14 = 0.35
Rainy	2	2	4/14 = 0.29
Sunny	2	3	5/14 = 0.35
All	4/14 = 0.29	10/14 = 0.71	

Applying Bayes' Theorem :-

$$P(Yes|Sunny) = P(Sunny|Yes) * \frac{P(Yes|Sunny)}{P(Yes) / P(Sunny)}$$

$\Rightarrow$ $P(Sunny|Yes) = 3/10 = 0.3$

$P(Sunny) = 0.35$

$P(yes) = 0.71$

Page No.

Saathi

Date __/__/____

So, $P(Yes|Sunny) = 0.3 * 0.71 / 0.35 = 0.60$

$P(No|Sunny) = P(Sunny|No) * P(No) / P(Sunny)$

$P(Sunny|No) = 2/4 = 0.5$

$P(No) = 0.29$

$P(Sunny) = 0.35$

So, $P(No|Sunny) = 0.5 * 0.29 / 0.35 = 0.41$

So, as we can see from the above calculation that $P(Yes|sunny) > P(No|Sunny)$

Hence, on a sunny day, Player can play the game.

Advantages of Naive Bayes Classifier	Disadvantages
It can be used for Binary as well as Multi-class Classification.	① Naive Bayes assumes that all the features are independent or unrelated, so it cannot learn the relationship between features.
It performs well in Multi-class predictions as compared to the other Algorithms.	
It is most popular choice for text classification problems.	

Page No. ____

Saathi

Date __/__/____

⇒ Applications of Naive Bayes Classifier :-

① It is used for credit scoring.
② It is used in medical data classification.
③ It can be used in real-time predictions because Naive Bayes Classifier is an eager learner.

⇒ Types of Naive Bayes Model :-

① **Gaussian** ⇒ The Gaussian model assumes that features follow a normal distribution. This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.

② **Multinomial** ⇒ The Multinomial Naive Bayes classifier is used when the data is multinomial distributed. It is primarily used for document classification problems, it means a

Saathi

Date __/__/____

particular document belongs to which category such as sports, Politics, eduction. This classifier uses the frequency of words for the predictors.

Bernoulli :→ This classifier works similar to the Multinomial classifier, but the predictor variables are the independent Boolean variables. Such as if a particular word is present or not in a document. This model is also famous for document classification tasks.

K-Means Clustring :-

This is unsupervised algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be 2 clusters and for K=3, there will be 3 clusters and so on.

Page No. ____

Saathi

Date ___/___/___

NOTE ⇒ It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.

(ii) It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in unlabeled dataset on its own without the need for any training

most

(iii) It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of the distances between the data point and their corresponding cluster.

(iv) This algorithm takes the unlabelled dataset as input, divides the dataset into k-number of clusters, and repeats

Page No.

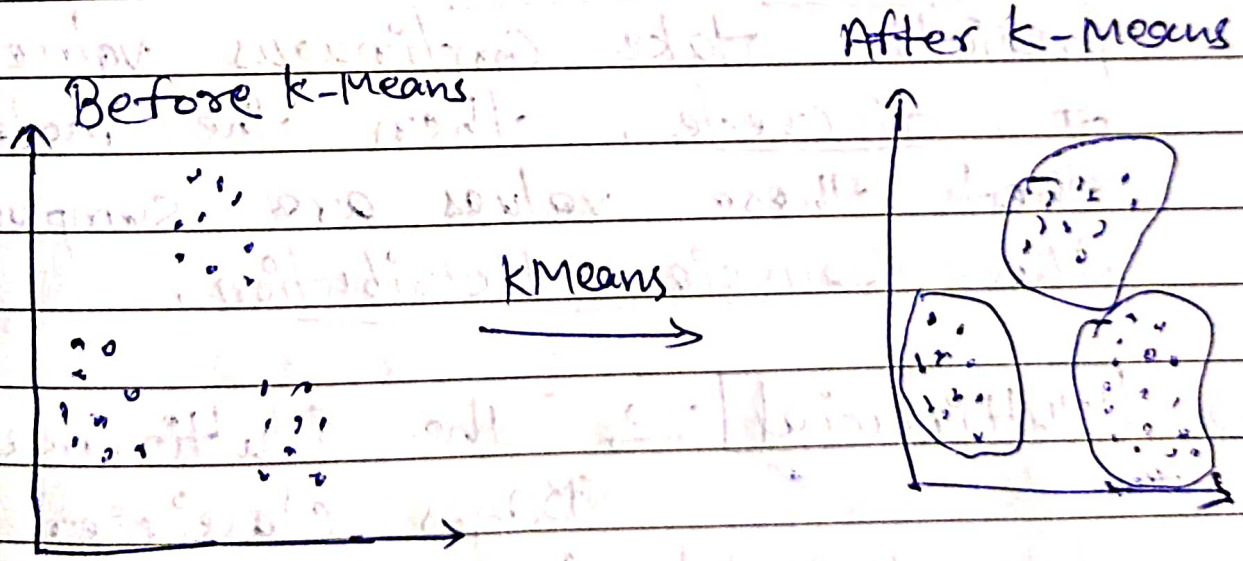
Saathi

the process until it does not find the best clusters. The value of k should be predetermined in the algorithm.

The k-Means clustering algorithm mainly performs 2 tasks:-

Determines the best value for k center points or centroids by an iterative process.

Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.



Saathi

Date __/__/____

⇒ How does the K-Means Algorithm Work?

Step-1 Select the number K to decide the numbe of clusters.

Step-2 Select random K points or centroids. (It can be other from the input dataset).

Step-3 Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4 Calculate the variance and place a new centroid of each cluster.

Step-5 Repeat the 3rd step, which means reassign each datapoint to the new closest centroid of each cluster.

Step-6 If any reassignment occurs, then go to Step-4th else go to finish.

Step-7 The model is ready.

Page No.

Date ___/___/_____

Saathi

Q5m. How to choose the value of "k number of clusters" in k-Means clustering?

Ans: ① **Elbow Method** :-

This is one of the most popular ways to find the optimal number of clusters. This method uses the concept of WCSS (Within Cluster Sum of Squares), which defines the total variations within a cluster.

The formula to calculate the value of WCSS (for 3 clusters) is given below:-

$$WCSS = \sum_{P_i \text{ in Cluster1}} distance(P_i C_1)^2 + \sum_{P_i \text{ in Cluster2}} distance(P_i C_2)^2 + \sum_{P_i \text{ in cluster}_3} distance(P_i C_3)^2$$

⇒ $\sum_{P_i \text{ in cluster1}} distance(P_i C_1)^2$. It is the sum of the square of the distances between each data points and its centroid with in a cluster1 and the same for the other $\sum P_i (P_i C_2)^2$ and $\sum P_i (P_i C_3)^2$.

Page No.

Saathi

Date ___/___/___

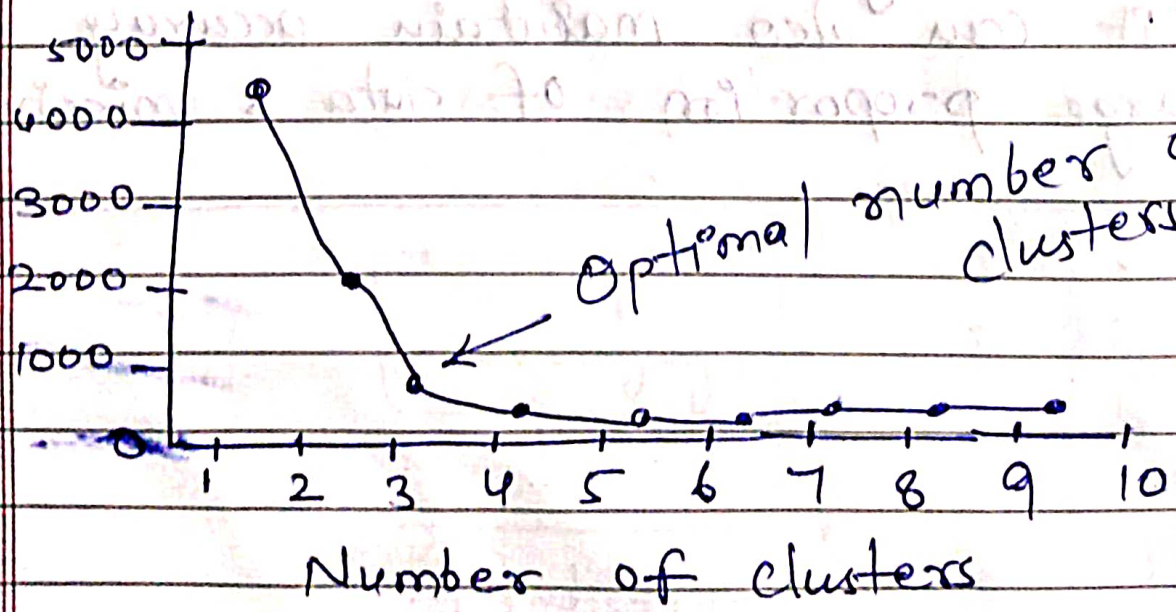
⇒ To find the optimal values of clusters, the elbow method follows these steps :-

(i) It executes the K-Means clustring on a given dataset for different K values (range from 1-10).

(ii) For each value of K, calculates the WCSS value.

(iii) Plot a curve between calculated WCSS values and the number of clusters K.

(iv) The sharp point of bend or a 'point of the plot looks like an arm, then that point is considered as the best values of K.



Date ___/___/_____

Saathi

NOTE ⇒ We can choose the number of clusters equal to the given data points. If we choose the number of clusters equal to the data points, then the value of WCSS become zero, and that will be the endpoint of the plot.

---

(7) Random Forest Algorithm :-

It can be used for both regression and classification problems in ML. It is based on the concept of ensemble learning, which is a process of combining a multiple classifier to solve a complex problem and to improve the performance of the model.

"Random Forest is a classifier that contains a number of decision trees on various subsets of the given datasets and take the average to improve the predictive accuracy of that dataset."
Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the

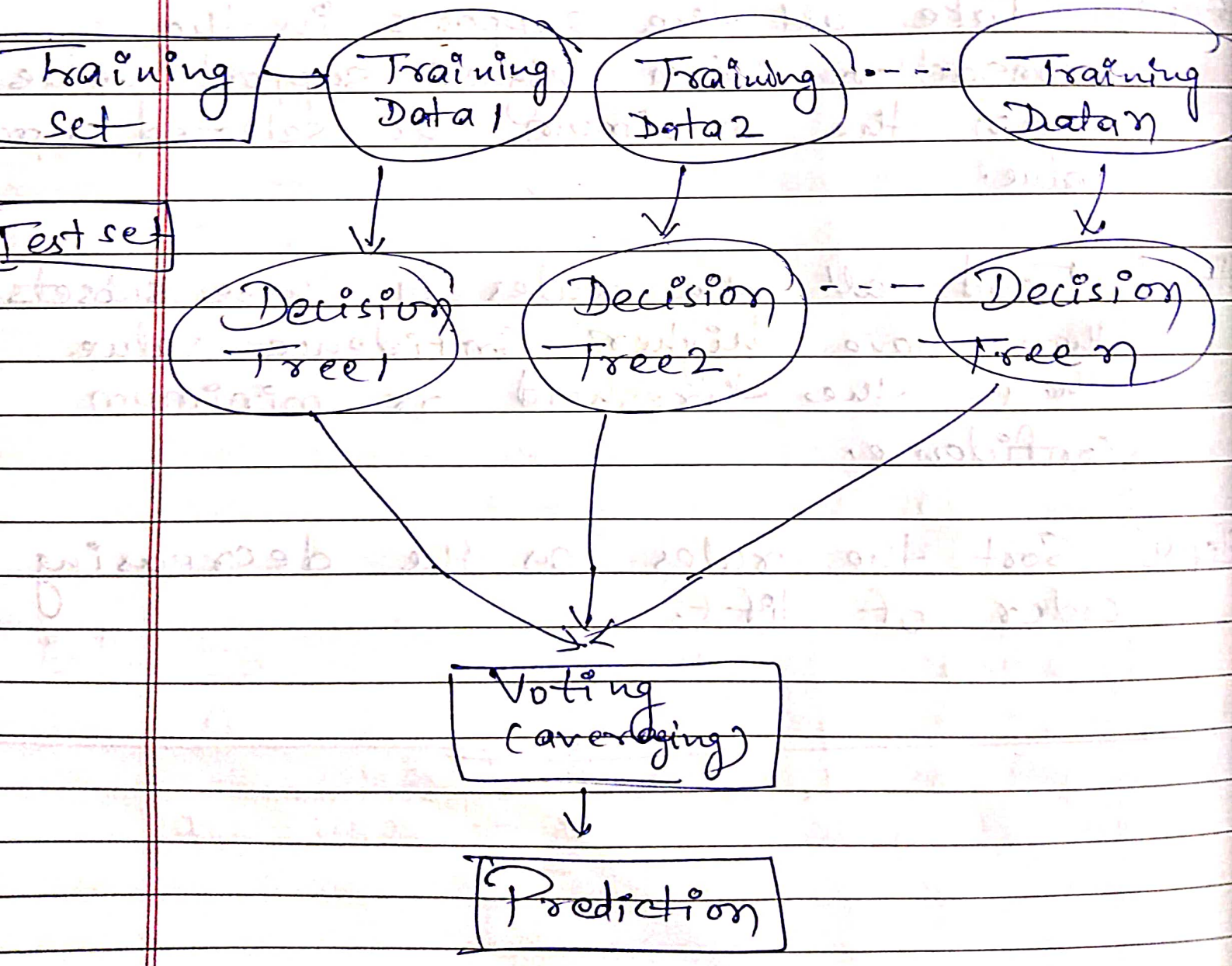
Page No. ____

Saathi

Date ___/___/_____

majority votes of predictions, and it predi
the final output.

The greater number of trees in the fores
leads to higher accuracy and prevents th
problem of overfitting.



Date ___/___/_____

Saathi

⇒ Assumptions for Random Forest :-

(i) There should be some actual values in the feature variables of the dataset so that the classifier can predict accurate results rather than a guessed results.

(ii) The predictions from each tree must have very low correlations.

⇒ Why use Random Forest ?

(i) It takes less training time as compared to other algorithms.

(ii) It predicts output with high accuracy, even for the large dataset it runs efficiently.

(iii) It can also maintain accuracy when a large proportion of data is missing.

Saathi

Date __/__/____

(9) Apriori Algorithm :-

This algorithm uses frequent itemsets to generate association rules, and it is designed to work on the databases that contain transactions.

With the help of these association rule, it determines how strongly or how weakly two objects are connected.

Q. Sn. What is Frequent Itemset?

Ans. Frequent Itemset are those items whose support is greater than the threshold value or user-specified minimum support. It means if A & B are the frequent itemsets together, then individually A and B should also be the frequent itemset.

Ex. - If there are two transactions: A = {1, 2, 3, 4, 5} and B = {2, 3, 7} in these two transactions 2 and 3 are the frequent itemsets.

Date __/__/____

Saathi

⇒ **Steps of Apriori Algorithm :-**

Step 1 Determine the support of itemsets in the transactional database, and select the minimum support and confidence.

Step 2 Take all the supports in the transaction with higher support value than the minimum or selected support value.

Step 3 Find all the rules of these subsets that have higher confidence value than the threshold or minimum confidence.

Step 4 Sort the rules as the decreasing order of lift.

Saathi

Date __/__/____

Q5n: Suppose we have dataset that have various transactions, and from of this dataset, we need to find the frequent itemsets and generate the association rules using Apriori Algo?

TID	ITEMSETS
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

and Minimum Support = 2,
Minimum Confidence = 50%.

Sol^n:

Step1 Calculating C1 and L1 :-

Itemset	Support_Count
A	6
B	7
C	5
D	2
E	1

Page No.

Saathi

Date ___ / ___ / ____

Step 1 Now, we will take out all the itemsets that have greater support count that the Minimum Support (2). It will give us the table for the frequent itemset L1.

Since, all the itemsets have greater or equal support count than the minimum support, except the E. So, E Itemset will be removed.

Itemset	Support_Count
A	6
B	7
C	5
D	2

Step 2

Itemset	Support_Count
{A, B}	4
{A, C}	4
{A, D}	1
{B, C}	4
{B, D}	2
{C, D}	0

⇒

Itemset	Support_Count
{A, B}	4
{A, C}	4
{B, C}	4
{B, D}	2

Page No. ____

Step3 Candidate generation C3, and L3 :-

Itemset	Support_Count
{A, B, C}	2
{B, C, D}	1
{A, C, D}	0
{A, B, D}	0

Step4 Finding the association rules for the subsets :-

To generate the association rules, first, we will create a new table with the possible rules from the occurred combination {A, B, C}. For all the rules, we will calculate the Confidence using formula $Sup(A \cap B)/A$. After calculating the confidence values for all rules, we will exclude the rules that have less confidence that the minimum threshold (50%).

Date ___/___/___

Saathi

Rules	Support	Confidence
$A \wedge B \rightarrow C$	2	$Sup\{(A \wedge B) \wedge C\} / Sup(A \wedge B)$ $= 2/4 = 0.5 = 50\%$
$B \wedge C \rightarrow A$	2	$Sup\{(B \wedge C) \wedge A\} / Sup(B \wedge C)$ $= 2/4 = 0.5 = 50\%$
$A \wedge C \rightarrow B$	2	$Sup\{(A \wedge C) \wedge B\} / Sup(A \wedge C)$ $= 2/4 = 0.5 = 50\%$
$C \rightarrow A \wedge B$	2	$Sup\{(C \wedge (A \wedge B)\} / Sup(C)$ $= 2/5 = 0.4 = 40\%$
$A \rightarrow B \wedge C$	2	$Sup\{(A \wedge (B \wedge C)\} / Sup(A)$ $= 2/6 = 0.33 = 33.33\%$
$B \rightarrow B \wedge C$	2	$Sup\{(B \wedge (B \wedge C)\} / Sup(B)$ $= 2/7 = 0.28 = 28\%$

As the given threshold or minimum confidence is 50%, so the first 3 rules $A \wedge B \rightarrow C$, $B \wedge C \rightarrow A$, and $A \wedge C \rightarrow B$ can be considered as the strong association rules for the given problem.

Page No.