Weekly Assignment - III

1. Using the general structure of Hunt's algorithm, design a decision tree using the dataset illustrated via Figure 1.

ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Figure 1: Borrower dataset

2. Construct a Decision Tree classifier from scratch, using the Gini Index as the splitting criterion. The program should be able to take a dataset (Figure 1) as input, calculate the Gini Index for each possible split, and choose the best split based on the lowest Gini value. The program should recursively build the decision tree until all data is classified or no further splits are possible.

3. Using the dataset illustrated via Figure 2, perform the following tasks:

   a. Generate the C4.5 rules.

   b. Compute the coverage and accuracy of each designed rule.

   c. Compute the overall model performance in terms of accuracy after designing (actual class Vs predicted class) dataframe.

Day	outlook	temp	humidity	windy	play
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes

Figure 2: Play Dataset

4. Predict whether the bank note is authentic or not using K-Nearest Neighbor (KNN) Classifier. Implement the KNN on the following Bank Authentication dataset.
   Dataset Link: `https://archive.ics.uci.edu/dataset/267/banknote+authentication`
   Table 1 describes the Bank authentication dataset

Variable Name	Role	Type
variance	Feature	Continuous
skewness	Feature	Continuous
curtosis	Feature	Continuous
entropy	Feature	Continuous
class	Target	Integer

Table 1: Bank Authentication Dataset

5. Implement a Naive Bayes Classifier to calculate the probability of playing when the weather is overcast, and the temperature is mild. The program should calculate the conditional probabilities using the Naive Bayes assumption (independence between features), and it should classify test instances based on the calculated probabilities.

Whether	Temperature	Play
Sunny	Hot	No
Sunny	Hot	No
Overcast	Hot	Yes
Rainy	Mild	Yes
Rainy	Cool	Yes
Rainy	Cool	No
Overcast	Cool	Yes
Sunny	Mild	No
Sunny	Cool	Yes
Rainy	Mild	Yes
Sunny	Mild	Yes
Overcast	Mild	Yes
Overcast	Hot	Yes
Rainy	Mild	No

Figure 3: Caption