Assignment 2

Dinesh J Kodwani

Task 2-1

1) Creating the data frame and cleaning and encoding the data.

Original data is added *(see table 1)*. Since the data contains categorical data, we encode it into numerical format *(see table 2)*.

	Name	Debt	Income	Married?	Owns_Property	Gender	Risk
0	Tim	low	low	no	no	male	low
1	Joe	high	high	yes	yes	male	low
2	Sue	low	high	yes	no	female	low
3	John	medium	low	no	no	male	high
4	Mary	high	low	yes	no	female	high
5	Fred	low	low	yes	no	male	high
6	Pete	low	medium	no	yes	male	low
7	Jacob	high	medium	yes	yes	male	low
8	Sofia	medium	low	no	no	female	low

Table 1. Original data

	Name	Debt	Income	Married?	Owns_Property	Gender	Risk
0	Tim	0	0	0	0	1	0
1	Joe	2	2	1	1	1	0
2	Sue	0	2	1	0	0	0
3	John	1	0	0	0	1	1
4	Mary	2	0	1	0	0	1
5	Fred	0	0	1	0	1	1
6	Pete	0	1	0	1	1	0
7	Jacob	2	1	1	1	1	0
8	Sofia	1	0	0	0	0	0

Table 2. Encoded data

The data is encoded such that for:
'Married?', 'Owns_Property': *yes=1*, *no=0*,
Gender: *'male'=1*, *'female'=0*, 'Risk': *high=1*, *low=0*,
'Debt' and 'Income': *'high'=2*, *'medium'=1*, *'low'=0*.

Decision Tree: We remove the 'Name' column, and add all the features in the DataFrame 'X' and we store the Target labels 'Risk' in DataFrame 'Y'.

Since this whole dataset is meant for training, we don't split the dataset with train and test.

We train the Decision Tree Classifier with X and Y. See trained decision tree *(see figure 1 and 2)*.

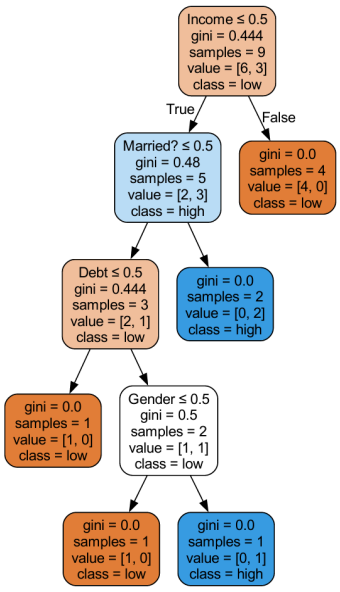

figure 1. Built Decision Tree for Credit Prediction

```
Income <= 0.50
|  Married? <= 0.50
|  |  Debt <= 0.50
|  |  |  Predict: Low Risk
|  |  Debt > 0.50
|  |  |  Gender <= 0.50
|  |  |  |  Predict: Low Risk
|  |  |  Gender > 0.50
|  |  |  |  Predict: High Risk
|  Married? > 0.50
|  |  Predict: High Risk
Income > 0.50
|  Predict: Low Risk
```

figure 2. Decision tree with indentation.

Using this decision tree, we predict credit risk for two data instances. *(see table 3)*

Name	Debt	Income	Married?	Owns Property	Gender
Tom	low	low	no	Yes	Male
Ana	low	medium	yes	Yes	female

figure. Table 3

Decision Sequences:

Tom's decision sequence: *Income = low -> Married = no -> Debt -> low -> Credit risk = low*	Ana's decision sequence *Income = medium -> Married = yes -> Debt -> low -> Credit risk = low*

Task 2-2: Sofia's risk = high

1) How does the decision tree change, with Sofia's risk as high? *(see figure 3)*

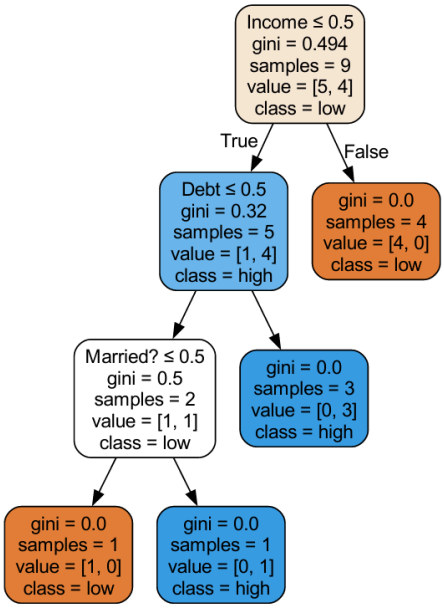

figure. 3 New Decision Tree

We can observe that 'Debt' is given more importance than 'Married?' now as compared to the decision tree earlier *(figure 1)*.

Further, it can be observed that now 'Gender' does not play a role in identifying credit risk.