**Assignment -02 – Decision Tree**

**Part 2**

Task 2-1. Decision Tree:

The decision tree was built using the ID3 algorithm implemented in Part 1. Below is the text representation of the decision tree with the predicted credit risk for Tom and Ana:

|  |
| --- |
| A computer screen shot of a black screen  AI-generated content may be incorrect. |

The decision tree was used to predict the credit risk for the following individuals:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Debt** | **Income** | **Married?** | **Owns Property** | **Gender** | **Predicted Risk** |
| Tom | Low | Low | No | Yes | Male | Low |
| Ana | Low | Medium | Yes | Yes | Female | Low |

Task 2.2. After Changing Sofia’s Credit Risk and seeing any unused feature

Before: Decision Tree and prediction with Sofia’s credit risk as low

|  |
| --- |
| A screenshot of a computer  AI-generated content may be incorrect. |

After: Decision tree after changing Sofia’s credit risk from low to high

|  |
| --- |
| A screen shot of a computer  AI-generated content may be incorrect. |

In the original tree, the split after Married? = 2 is on Owns\_Property

The tree first splits on Married?.

* If Married? = 0 or Married? = 1, the prediction is low.
* If Married? = 2, the tree further splits on Owns\_Property.
  + If Owns\_Property = 0, the prediction is high.
  + If Owns\_Property = 1, the tree further splits on Debt.
    - If Debt = 1 or Debt = 2, the prediction is low

In the modified tree, the split after Married? =2 is on Debt

The tree still first splits on Married?.

* If Married? = 0 or Married? = 1, the prediction is low.
* If Married? = 2, the tree now splits on Debt instead of Owns\_Property.
  + If Debt = 1, the tree further splits on Owns\_Property.
    - If Owns\_Property = 0, the prediction is high.
    - If Owns\_Property = 1, the prediction is low.
  + If Debt = 2, the prediction is high.

The modified tree introduces a new branch for Debt = 2 with a prediction of High

**Unused Feature** – in both trees, the features Income and Gender do not play a role in the decision tree. The Income feature does not provide significant information gain for splitting the dataset. As a result, it is not used in the decision tree. This could be because the distribution of Income values does not strongly correlate with the target variable (Risk).

The Gender feature also does not provide significant information gain for splitting the dataset. This could be because the distribution of Gender values is not strongly correlated with the target variable (Risk).