

Academic Report

Decision Tree (CS 539)

Ehu Shubham Shaw

Prof Yanhua Li

October 3, 2024

Task 2-1: Draw your decision tree and report it. You may use visualization tools (e.g., Graphviz) or use text. You might find it easier if you turn the decision tree on its side, and use indentation to show levels of the tree as it grows from the left.

```
|--- property <= 0.50
|   |--- income <= 0.50
|   |   |--- class: low
|   |--- income > 0.50
|   |   |--- debt <= 1.50
|   |   |   |--- marital_status <= 0.50
|   |   |   |   |--- class: low
|   |   |   |   |--- marital_status > 0.50
|   |   |   |   |--- class: high
|   |   |--- debt > 1.50
|   |   |--- class: high
|--- property > 0.50
|   |--- class: low
```

Tom's has a ['low'] risk
Ana's has a ['low'] risk

```
|--- property <= 0.50
|   |--- debt <= 0.50
|   |   |--- class: high
|   |--- debt > 0.50
|   |   |--- gender <= 0.50
|   |   |   |--- class: low
|   |   |   |--- gender > 0.50
|   |   |   |--- debt <= 1.50
|   |   |   |   |--- marital_status <= 0.50
|   |   |   |   |   |--- class: low
|   |   |   |   |   |--- marital_status > 0.50
|   |   |   |   |   |--- class: high
|   |   |   |--- debt > 1.50
|   |   |   |--- class: high
|--- property > 0.50
|   |--- class: low
```

Tom's has a ['low'] risk
Ana's has a ['low'] risk

```
|--- property <= 0.50
|   |--- debt <= 0.50
|   |   |--- class: high
|   |--- debt > 0.50
|   |   |--- gender <= 0.50
|   |   |   |--- class: low
|   |   |   |--- gender > 0.50
|   |   |   |--- marital_status <= 0.50
|   |   |   |   |--- debt <= 1.50
|   |   |   |   |   |--- class: low
|   |   |   |   |   |--- debt > 1.50
|   |   |   |   |   |--- class: high
|   |   |--- marital_status > 0.50
|   |   |--- class: high
|--- property > 0.50
|   |--- class: low
```

Tom's has a ['low'] risk
Ana's has a ['low'] risk
Fred has a ['high'] risk
Sofia has a ['low'] risk

```
|--- property <= 0.50
|   |--- income <= 0.50
|   |   |--- class: low
|   |--- income > 0.50
|   |   |--- marital_status <= 0.50
|   |   |   |--- debt <= 1.50
|   |   |   |   |--- class: low
|   |   |   |   |--- debt > 1.50
|   |   |   |   |--- class: high
|   |   |--- marital_status > 0.50
|   |   |--- class: high
|--- property > 0.50
|   |--- class: low
```

Tom's has a ['low'] risk
Ana's has a ['low'] risk
Fred has a ['low'] risk
Sofia has a ['low'] risk

	Tom's has a ['low'] risk	Tom's has a ['low'] risk
Tom's has a ['low'] risk	Ana's has a ['low'] risk	Ana's has a ['low'] risk
Ana's has a ['low'] risk	Fred has a ['high'] risk	Fred has a ['low'] risk
	Sofia has a ['low'] risk	Sofia has a ['low'] risk

Task 2-2: How does your decision tree change if Sofia's credit risk is high instead of low as recorded in the training data? Given the decision tree constructed from the original dataset, if existing, name any feature not playing a role in the decision tree.

After modifying Sofia's credit risk from "low" to "high," I reran the decision tree model with the updated dataset and compared the resulting tree to the original. Below are my key findings:

- Original Decision Tree (Sofia's Credit Risk: Low)

- First Split: Based on `property <= 0.50`.
- Second Split: For individuals with `property <= 0.50`, the next split occurs on `debt`.
 - `debt <= 0.50` leads to the class: high (credit risk is high).
- Subsequent splits further refine the classification based on `gender`, `debt`, and `marital_status`.

- Modified Decision Tree (Sofia's Credit Risk: High)

- First Split: Remains the same (`property <= 0.50`).
- Second Split: For individuals with `property <= 0.50`, the feature used for the second split changes to `income` instead of `debt`.
 - `income <= 0.50` predicts a low credit risk, while for higher income values, further splits occur on `gender`, `debt`, and `marital_status`.

- Analysis of Feature Usage and Importance

After constructing both trees, I assessed the significance of characteristics based on their appearance in splits. Any feature not included in the decision tree was deemed unimportant for distinguishing between high and low credit risk. Here is what I found.:

1. Original Tree:

- The second split for individuals with `property <= 0.50` is based on `debt`.
- `Income` is not used as a splitting criterion, indicating it wasn't important in the original model.

2. Modified Tree (Sofia's Credit Risk: High):

- The second split now occurs on `income`, rather than `debt`. This suggests that changing Sofia's credit risk significantly altered the importance of features in the decision-making process.
- `Income` becomes a key feature, whereas in the original tree, it wasn't used at all.

My Conclusion

In both the original and updated trees, features such as 'property', 'gender', 'debt', and 'marital_status' are used consistently in different splits. However, **income**, which was underutilized in the original tree, becomes a substantial element in the amended tree when Sofia's credit risk has been adjusted. This modification in Sofia's risk classification demonstrates how changing a single instance in a dataset can affect feature relevance and the whole structure of the decision tree.