Machine learning Classification

Lab 10

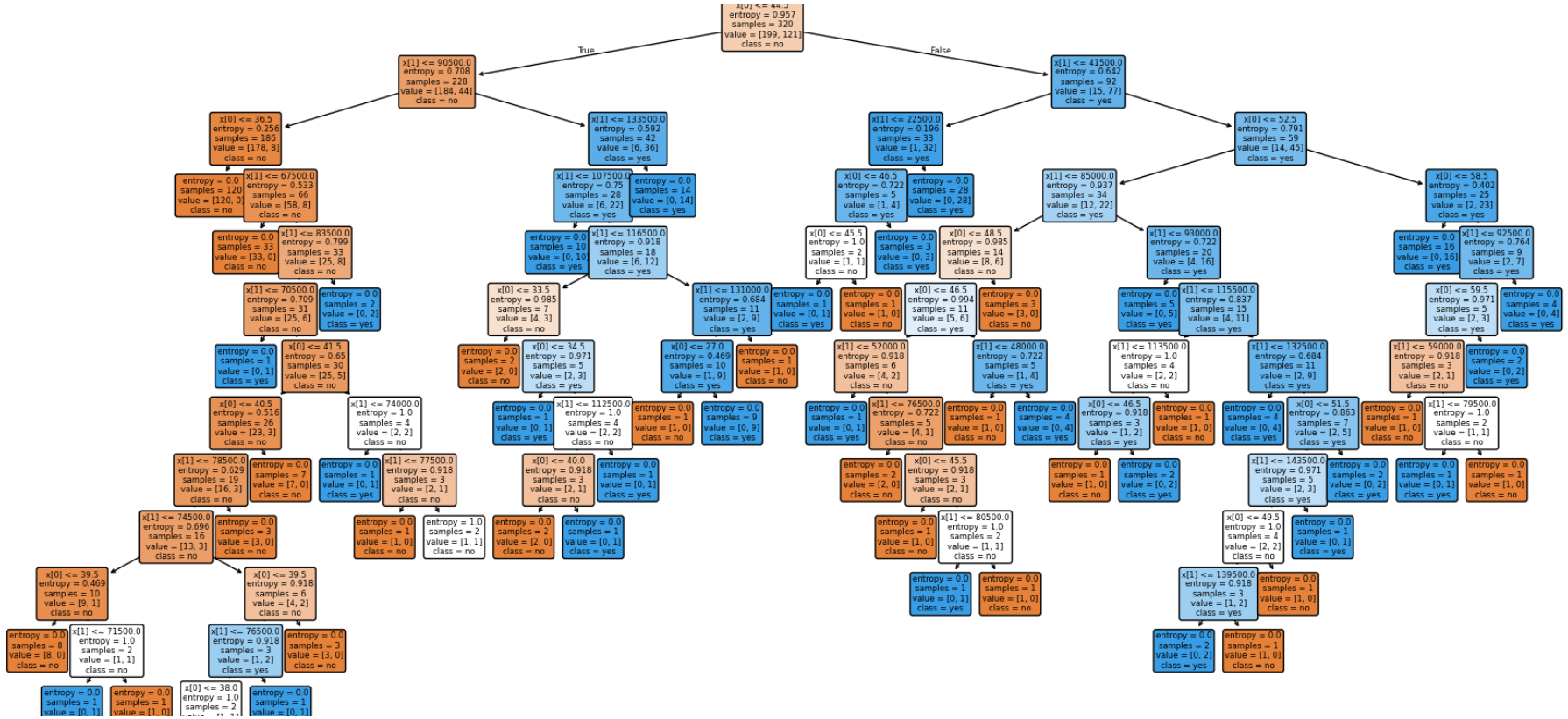
Exercise 1: Model training and checking overfitting

1. Explore the code in `decision_tree_classification.py`. Use `social_network_ads.csv` dataset to develop a decision tree model. “*purchased*” column is the output class, demonstrating whether a person bought a product based on advertisement.

2. Write code scripts to visualize the dataset features to better understand the data.

3. Implement the code scripts for checking the overfitting and fix potential overfitting, if there is one.

Visual Representation of an example Decision Tree



Exercise 2 – Classification on Adult Income Data

1. Explore the `adult_income.csv` dataset. "*income_high*" is the output class. Visualize the dataset features to better understand the data.

2. Create a decision tree based classification model and evaluate its performance.

3. Do you need to use all features in the predictions? Try to improve your base model's performance by changing the model parameters or dataset features.

4. Check if your model overfits. If so, apply some strategies to fix overfitting.