

Homework Assignment 3 – Decision Trees

Dr. Gail Gilboa-Freedman

Dr. Naveh Eskinazi

Mr. Asaf Lev

Submission: 1/6/2023

GENERAL INSTRUCTIONS

In the current assignment, you will classify bank customers with the decision tree algorithm.

The work will be based on a CSV named “**bank_marketing.csv**” located on the course’s Moodle site.

SUBMISSION:

Through the assignment box within the course Moodle, submit a **Jupyter Notebook file named HWA3_<student name>.ipynb** (e.g. HWA3_avia_malka.ipynb)

Should include all the relevant code needed to perform the assignment’s tasks along with the code's output.

(Recommendation: Add headers and sub-headers using the Markdown option)

Good Luck!

PART 1: PREREQUISITES**TASK 1: SETTING THE FOLDER**

1. Create a Jupyter Notebook named **HWA3_<student name>.ipynb**.
2. Download from the CSV file named **"bank_marketing.csv"** from Moodle.
3. Upload the CSV file to Jupyter (Note: make sure the file is placed in the same location as your Jupyter Notebook)

TASK 2: IMPORT LIBRARIES & MODULES

4. Import the following libraries and modules within your notebook: **pandas, numpy, DecisionTreeClassifier (from sklearn.tree), tree (from sklearn), and metrics (from sklearn)**

TASK 3: EXPLORE THE DATA

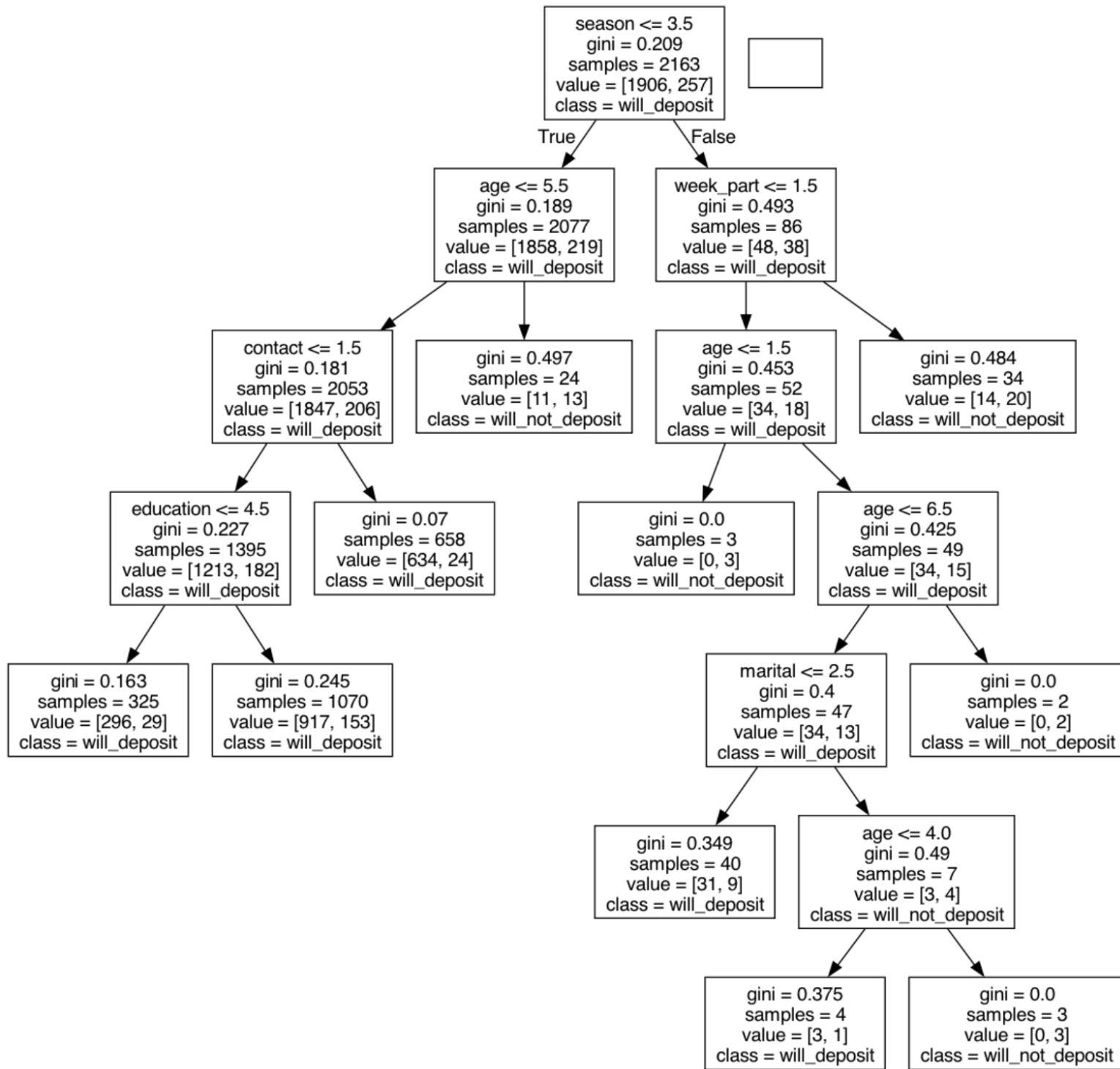
Use Python commands (e.g., head, columns, and shape) to plot the answers to the following questions:

5. Based on how many **customers** the algorithm will be trained and tested? (Provide a separate answer for both train and test).
6. How many **features** each customer is represented by? (Provide a numerical answer)
7. Which features are categorized as **predictors**? (Provide a verbal answer)
8. Which feature is categorized as **the label (target)**? (Provide a verbal answer)

PART 2: BUILDING A DECISION TREE MODEL**TASK 4: BUILDING THE MODEL**

Use Python commands (i.e., DecisionTreeClassifier and fit) to build a Decision Tree model (**Important Note! set the min_impurity_decrease property to 0.00065**)

9. In the next page, you will see visual representation of the tree.



10. Part of the predictors that were included in the data are not presented in the decision tree. Explain why (Provide a verbal answer)
11. The Gini value shown in the first tree node of the decision tree is **0.209**, explain what represents this value (Provide a verbal answer)
12. The attribute that was selected for the root of the tree is '**season**'. Provide a detailed calculation of the Gini index for this attribute (Hint: you can use Excel, Python or perform a manual calculation)
13. Below are two new customers that will be reached by the bank's campaign. Based on the decision tree, determine for each of them whether it will deposit money or not (Provide a verbal and detailed answer):
 - A customer that is older than 65, will be reached by cellular phone, at the end of a week during the Winter or Spring.
 - A customer that is older than 25, will be reached by telephone, at the beginning of a week during the Summer or Autumn.

TASK 5: EVALUATE THE MODEL

14. Use the Accuracy measure to evaluate the mode you have created in Task 4.
15. Describe, in the context of the model you have built, **the meaning of the numerical result** you have received for the Accuracy measure (Provide a verbal answer using terms such as classification matrix, True Positive, etc.)
16. Describe **one prominent disadvantage** of the Accuracy measure you have calculated.
17. **(Bonus Question)** Explain how measures of Sensitivity (TPR) and Specificity (TNR) can be used practically and wisely in the context of the bank's marketing campaign (Provide a verbal answer, no need to calculate the measures).

Good Luck!