LABORATORY : Classification

1. If you forgot how to classification in Python, study about it from Data Camp's tutorial

    - Decision Tree : *https://www.datacamp.com/tutorial/decision-tree-classification-python*
    - KNN : https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn
    - Naïve Bayes : *https://www.datacamp.com/tutorial/naive-bayes-scikit-learn*
    - Logistic Regression : *https://www.datacamp.com/tutorial/understanding-logistic-regression-python*

2. Read about confusion matrix from *https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning*

3. Personal Equity Plan

    bank.csv contains data of 600 bank customers. There are 11 columns :- id, age, sex, region, income, married, children, car, save_act, current_act, mortgage, pep.

    The last column, pep, indicates whether a customer purchased a Personal Equity Plan (PEP). We are to predict if a customer would purchase the PEP.

    1. Perform the EDA to get to know the data.
    2. Clean the data, if necessary. Prepare the data for analysis.
    3. Create classification models :-
        - Decision Tree (entropy and gini)
        - KNN
        - Naïve Bayes
        - Logistic Regression
    4. For Decision Tree models, visualize the trees.
    5. Evaluate the models.
    6. Try vary hyperparameters.
    7. Discuss what you have learned at the end of the ipynb file.

What to submit :

1. the ipynb file(s) with markdowns explaining the process in detail.
2. the html version of the ipynb file(s) showing all outputs.