

Objective: Write a python program to perform a binary classification task on the dataset: “usa_income_data.csv”.

Problem statement: All columns in the data, except for the last column, are the predictors. The last column: “income_class” is the target variable. Create a classification model to predict the target column.

Tools you can use: The constraints of this test are that the modeling has to be done in Python. You are free to choose any library of your choice such as sklearn, etc.

Deliverables: 1) Python code to train and validate the model. 2) A brief explanation about choice of model trained and the results and findings. While this should be a pretty short task, we encourage you to use object-oriented design for your code. In short, use classes, functions, etc. where appropriate although it is not mandatory. Write a short paragraph in your submission file explaining your methodology, and findings. The findings should include comments on what accuracy (or recall, precision, f1 score, etc.) is to be expected on unseen data.

Allotted time: There is no hard deadline, but if you think you will need more than 2 days, please reach out to us and let us know with a brief explanation about why you are requesting more time.

Submission instructions (PLEASE READ): Please fork the repo, and commit the changes to your own forked repo. Then submit a pull request once you are done. Please reach out if you have any questions.

Important Note: Please note that this is not a machine learning (ML) / data science (DS) challenge by any means where the model with the best prediction performance wins. You are mainly being evaluated on your understanding of DS and ML and ability to perform standard DS/ML tasks in Python.