Submit your code and a separate pdf describing the results that Phoenix Solar should expect from the classifier, you constructed for them.
The pdf should at minimum include 1) a description of your approach and the classifier chosen , including Why you chose the selected classifier, and 2) the test metrics computed.

Results to expect:

1. I have trained two models viz SVM and Decision Tree.
2. I chose those two particular models because they possess the following advantages:
3. **SVM**
    a. SVMs are effective in high-dimensional spaces and are therefore a good choice for data sets with a variety of attributes.
    b. SVMs can avoid overfitting by finding the best-separating hyperplane that maximizes the margin between classes. They are robust to noisy data.
    c. SVMs can handle data that is not linearly separable by translating it into a higher-dimensional space where a hyperplane can be used to divide it.
    d. SVMs have a wide range of applications, including both classification and regression.
    e. SVMs are capable of locating the global minimum or maximum, guaranteeing that the ideal solution is discovered.
4. **Decision Tree**
    a. Decision trees are effective for simplifying the explanation of difficult ideas since they are simple to understand and intuitive.
    b. Interpretable: Decision trees are a great tool for understanding how the model generates decisions since they are simple to visualize.
    c. Strong: Decision trees can withstand missing values and noisy data.
    d. Decision trees can be used to discover the most informative characteristics for the classification task, making them useful for feature selection.
    e. Decision trees are nonparametric, therefore they don't make any assumptions about how the data are distributed.

I have achieved an accuracy of 78.965% with SVM and 83.82% with Decision Tree.

My approach to went was Data cleaning, data normalization, and dropping null and duplicate values. Once the data is cleaned, I replaced all the categorical values with numbers and the refined data was ready to be trained.

I chose two models, and performed to hyper-parameter tuning with different sets of parameters, two find the best estimator in both cases, I then calculated the parameters such as Recall, F1 score, and accuracy to determine which model will be the best fit to predict the income catagory of the individual.

The following are the test matrices for each of the model.

SVM:
Overall Accuracy: 0.7896545798247121
Precision: 0.7896545798247121
Recall: 0.789654579824712
F1 Score: 0.7896545798247121
Best hyperparameters: {'C': 100, 'kernel': 'poly'}

Decision Tree:
Best Parameters: {'criterion': 'gini', 'max_depth': 9}
Overall Accuracy: 83.82883656985737
Precision: 0.8382883656985737
Recall: 0.8382883656985737
F1 Score: 0.8382883656985736