

# Homework 1 - Programming

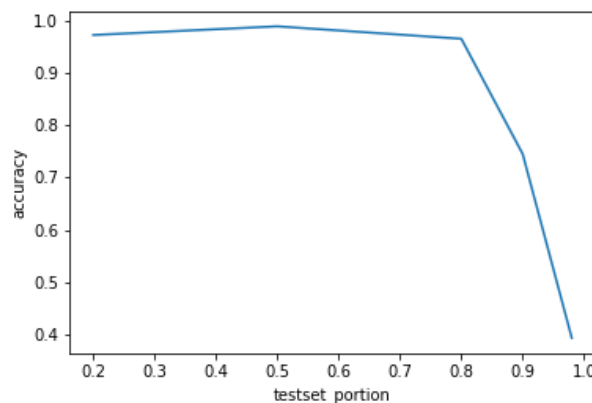
## Python Tutorial and Naive Bayes Classification

1.

- a. The parameter **random\_state** can control the randomness of the selection applied to data before applying the split.
- b. It does affect the accuracy. The parameter **random\_state** can determine how to split the training data and testing data. We can imagine the case if the training data doesn't contain the specific feature, the trained classifier won't recognize the feature in testing dataset, thus to induce wrong predictions.

2.

- a. There is a decreased trend of accuracies when **testset\_portion** becomes more and more large. The following shows the plot of testset\_portion vs. accuracy :



- b. If **test\_portion** becomes larger, it means that the amount of training data is small. Thus, we have insufficient data to train the model, resulting in declined accuracies. (i.e., underfitting problems)