229352 Statistical Learning for Data Science 2 **Fall 2021**

Lab 1: due July 4

Instructor: Donlapark Ponnoprat

Load the *mini* MNIST dataset using the following code:

```
from sklearn import datasets
mnist = datasets.load_digits()
```

With this dataset, perform the following experiment:

1. Split the data into training, validation and test sets.
2. Train a k-nearest neighbor (kNN) model with the training set. Then use the validation set to find the best value of k and the best distance metric. You can also use the cross-validation instead of the validation set. In that case, you might want to take a look at `sklearn.model_selection.GridSearchCV`.
3. Report the *test* accuracy of the kNN model with the value of k and the distance metric obtained above.
4. Show at least two images that the model mis-classifies. For both images, what are the actual numbers and what are the model's predictions?

The report must be turned in as a PDF file. If you are using the Python Notebook, go to:
`File>Download as>pdf(.tex)`.