Machine Learning Assignment

# The Brief

Splitting the data into independent sets of data for training and testing is critical for the correct measurement of the performance of a machine learning prediction model. K-fold cross validation is a better data splitting approach than hold-out testing when there is less data available. The task is to come up with a way of testing this claim.

You should design and run an experiment (using scikit-learn) that shows that k-fold cross validation is better than hold-out because it provides a more stable, and better estimate of performance on small amounts of data. You can use the [wine dataset](https://archive-beta.ics.uci.edu/dataset/109/wine) or any other dataset you wish from the [UCI repository](https://archive-beta.ics.uci.edu/). Results should be output in an appropriate graphical format.

Ensure you include, in your jupyter notebook, markdown cells that clearly describe your experimental approach (with clear reasoning for each step in your approach) and your findings.

# Submission

Submissions should include one Jupyter notebook clearly annotated. Any required Python modules should be installed by the notebook. The notebook should expect to find the dataset at ../input/wine.csv (i.e. the input directory is located in the directory above the notebook).

Submit through Brightspace by Friday 21st April.

# Marking Scheme

The rubric below provides for a breakdown of how marks will be allocated for this assignment

Table

Description automatically generated