Workspace for Summer 2025 `Ethics and Bias in AI` course

Week 1 (6/9, 6/11)

1. Setup Python 3.10+ and create Virtual Environment named `venv-week1`
2. For compute engine use Kaggle notebook, or Google colab, or Deepnote

Activities involving Exploratory Data Analysis (EDA)

1. How many rows are in dataset: `week-01/datasets/A.csv`? How about in `week-01/datasets/B.txt` and in `week-01/datasets/C.csv`?
2. Repeat the previous task to find out how many `samples` (i.e., instances) are there in each of the dataset?
3. How many columns are there in each of the 3 datasets?
4. Compute the mean of the last (i.e., rightmost) column of `week-01/datasets/C.csv`?
5. Where (i.e, in which sample) the two datasets: `week-01/datasets/B.txt` and `week-01/datasets/D.txt` differ?
6. What is the average credit score among the samples found in `week-01/datasets/A.csv`?
7. How many different countries are listed in `week-01/datasts/A.csv`?
8. Please briefly describe each of the 3 datasets (i.e., what the datasets are about)
9. Care to explore more of the datasets?

Activities with an online dataset

1. Get the dataset: i) Go to https://www.fueleconomy.gov, ii) Scroll down the page to find link to "Download EPA's MPG Ratings" and click on it. iii) Locate the section titled Datasets for All Model Years (1984–2025)., iv) Click on Zipped CSV File to download the dataset. How many samples are there in the dataset? Also, how many features are there per sample
2. Let's filter the dataset. We are interested only with vehicle's engine displacement (`displ`), model year (`year`), unadjusted mpg on highway (`UHighway`), unadjusted mpg on city (`UCity`) and the fuel type (`fuelType`). Create a dataframe that contains the filtered dataset.
3. Do the following:
   - Plot a histogram of engine displacement data.
   - Plot histograms for each of the attributes present in the filtered dataset.
   - Plot a scatter plot between unadjusted highway mpg (UHighway) and engine displacement (displ).
   - Plot a scatter plot between unadjusted city mpg (UCity) and engine displacement (displ).
   - Plot a scatter-matrix between variables present in the filtered dataset.
4. Find outlier samples that do not have value for engine displacement. Please report how many such samples did you find?
5. We now have several option to amend our dataset for further processing. Option 1: Remove the samples, or, Option 2: fill the value with any of the central tendency. For this question, please choose option 2 and prepare the dataset.
6. Shuffle the dataset

7. Take one half of the dataset as training set, and the remaining half as test set. Please separate independent (displ) and dependent variable (UHighway)
8. Build a linear regression model based on the training dataset.
9. Use the trained model to predict Uhighway of the test dataset
10. Evaluate the model performance (in terms of Root Mean Squared Error, and R-squared score)
11. pick one sample from the test set, and estimate MPG. How bad was it?

Activities with a classification problem

1. A classification task for you: Get the dataset first: go to this link at https://www.muratkoklu.com/datasets/, and download the `Date Fruit Datasets`. You will get the `Date Fruit Datasets.zip`. Unzip it to get the `Date Fruit Datasets.xlsx` file among other files. We will be working on the `xlsx` file today.
2. Load the dataset and tell i) How many samples are there in the dataset? ii) how many features are there per sample, excluding the class/type label? iii) Print the number of unique classes (i.e., fruit types), list the class names, and their frequency distribution.
3. List mean, stdev, min, max of each of the features in the dataset
4. Shuffle the data samples.
5. Split the dataset into training (80%) and test (20%)
6. Scale all independent features.
7. Build a classifier with the training set.
8. Predict class label of just one sample picked from the test set.
9. Evaluate the model performance based on the test set in terms of accuracy, precision, recall and F1.

Machine Learning Pipeline

1. Please establish a `scikit-learn Pipeline` for any of the above 2 activities.
   - Reference 1: https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html
   - Reference 2: Example code 1

A short orientation on bias detection and mitigation with `aif360` package

1. Here is the link to retrieve the German Credit dataset: https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data

- Easy way to import into Python workspace:

```
!pip install ucimlrepo
from ucimlrepo import fetch_ucirepo

# fetch dataset
statlog_german_credit_data = fetch_ucirepo(id=144)

# data (as pandas dataframes)
X = statlog_german_credit_data.data.features
y = statlog_german_credit_data.data.targets

# metadata
print(statlog_german_credit_data.metadata)
```

```
# variable information
print(statlog_german_credit_data.variables)
```

2. Let's use the German credit dataset, splitting it into a training and test dataset.
3. We will look for bias in the creation of a machine learning model to predict if an applicant should be given credit based on various features from a typical credit application. The protected attribute will be "Age", with "1" (older than or equal to 25) and "0" (younger than 25) being the values for the privileged and unprivileged groups, respectively.
4. Firstly, we will check for bias in the initial training data. Then, we apply a `Reweighing` based bias mitigation strategy from the `aif360` package and re-assess the bias. Here are the complete steps involved. But, here we will mainly focus on detecting biases.

- **Step 1**: Write import statements
- **Step 2**: Set bias detection options, load dataset, and split between train and test
- **Step 3**: Compute fairness metric on original training dataset. Please refer to the documentation: `from aif360.metrics import BinaryLabelDatasetMetric`
- **Step 4**: Mitigate bias by transforming the original dataset. Please refer to the documentation: `from aif360.algorithms.preprocessing import Reweighing`.
- **Step 5**: Compute fairness metric on transformed training dataset