# Experiential Learning Classification II Machine Learning Modeling

# Experiment A

## Download the Dataset

From the datasets folder download the dataset named **Dataset\_Classification\_II\_LAB\_WineQuality\_White.csv** and save it in your preferred folder on your computer.

**Citation and credit for the dataset**:

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Wine Quality [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C56S3T>.

## Sign in to the Azure ML studio

1. Sign in to Azure Machine Learning studio at ml.azure.com.
2. In the left pane, click on **Automated ML** under the **Authoring** section.
3. Click on **+New automated ML job**.

## Create and load dataset

1. In the **Basics settings** tab, enter the **Job** name and **Experiment** name. Enter the names like the ones below. (if they exceed 60 characters in length enter only your last name and first name initial)  
     
   **FirstName\_LastName\_Classification\_II\_LAB\_WineQualityWhite\_Job\_A**  
   **FirstName\_LastName\_Classification\_II\_LAB\_WineQualityWhite\_Exp\_A**
2. Click **Next**.

## Task Type and Data

1. In the **Select Task Type** drop down menu, select **Classification**.
2. For **Select Data** click **Create** to create a new data asset from the file on your local computer.
3. On the page **Set the name and type for your data asset**:
4. Enter the **Data asset** name as  
   **Dataset\_Classification\_II\_LAB\_WineQuality\_White**
5. For the **Type**, select **Tabular** from the dropdown list.
6. Click **Next**.

## Choose a source for your data asset

1. Select **From local files**
2. Click **Next**.

## Select a datastore

Here we specify the Azure Storage location to upload our data.

1. For the **Datastore type**, select Azure Blob Storage.
2. In the list of datastores, select *workspaceblobstore*.
3. Select **Next**.

## Choose a file or folder

1. Click on the **Upload files or folder** dropdown menu and select the **Upload files** option.
2. Browse to the location where you saved the “**Dataset\_Classification\_II\_LAB\_WineQuality\_White.csv**” file and select **Open**.
3. After the file is uploaded, select **Next**.

## Settings

1. On the **Settings** page browse your data to check the values and see if anything is unusual or out of order.
2. Click **Next**.

## Schema

For this lab, we will use all the features available in the dataset.

Click **Next**.

## Review

Click **Create** to create your dataset.

## Task type & data

When your dataset is ready, the Azure Machine Learning studio returns back to the Task type & data page.

1. In the **Select task type** dropdown menu the choice **Classification** should already be there. If not, select it.
2. Click on the radio button to the left of the dataset  **Dataset\_Classification\_II\_LAB\_WineQuality\_White**
3. Click **Next** to continue.

## Task Settings

In the **Target column** dropdown list, select the column to use for the model predictions, in this case **Quality**.

## Classification Settings

Click on “**View Additional Configuration Settings**”

1. For the **Primary Metric**, select **AUCWeighted**, if it is not already selected. AUC means Area Under the Curve (AUC).
2. For the **Explain Best Model**, make sure the checkbox is checked.
3. For the **Enable Ensemble Stacking**, make sure the checkbox is unchecked.
4. For the **Use all supported models**, make sure the checkbox is checked.
5. For the **Blocked Models**, drop-down menu, leave it blank, that is, do not exclude any models.
6. For the **Positive class label**, leave it blank.

## Limits

1. For the **Experiment timeout (minutes) enter 120**. We want the experiment to end in 120 minutes. It might take more than 120 minutes.
2. Leave the rest of the **limits settings** as they are.

## Validate and test

1. **Validation Type**: **Enter k-fold cross validation**. In k-fold cross validation, the dataset is split into multiple folds. Then, as the model goes through training iterations (cross validations), it is trained on the number of folds we have defined minus one. One fold is used as the test dataset (fold). For example, if we assign three (3) as the number of cross validations, two (2) will be used as the training data sets, while the last fold will be used as the test fold.
2. **Number of cross validations**: 3
3. Click **Next**.

## Compute

1. **Select compute type**: From the drop down menu select **Serverless.**
2. **Virtual Machine Size**: From the drop-down menu, select “Standard\_F8s\_v2 (8 core(s), 16GB RAM, 64GB storage” or another upper-end high configuration, or you can even leave it blank.
3. Click **Next**.

**ATTENTION**! Just in case the serverless virtual machine is not enough to carry out the experiment, use the compute instance you have setup.

1. **Select compute type**: From the drop down menu select **Compute Instance**.
2. **Select Azure ML compute instance**: Select the compute instance you have already created. If you do not have a compute instance setup, go to the lab " **Create an Azure Cloud Workspace and Compute instance**" and create one.

## Review

Click **Submit Training Job**. You might see an error that the page does not exist or it has been deleted. It is OK, your experiment is running.

## Start Exploring the Models

You do not need to wait for the final model recommendation from Azure ML Studio. It will take about 15 minutes for the first model to be generated. You can start exploring the generated models and their characteristics at this point.

# Experiment B

Create the previous Auto ML experiment from scratch. Use the dataset you have already uploaded. Name this job and the experiment as:  
  
**FirstName\_LastName\_Classification\_II\_LAB\_WineQualityWhite\_Job\_B  
FirstName\_LastName\_Classification\_II\_LAB\_WineQualityWhite\_Exp\_B**

Keep all the Hyperparameter settings the same, except for the following:

**Primary metric**: NormMacroRecall

**Enable Ensemble Stacking**: Check

**Validation Type**: Automatic

# Experiment C

Create the previous Auto ML experiment from scratch. No need to reupload the data of course. Name this job and the experiment as:  
  
**FirstName\_LastName\_Classification\_II\_LAB\_WineQualityWhite\_Job\_C  
FirstName\_LastName\_Classification\_II\_LAB\_WineQualityWhite\_Exp\_C**

Keep all the Hyperparameter settings the same, except for the following:

**Primary metric**: Accuracy

**Validation Type**: Monte Carlo cross validation

Number of Cross Validations: 3

**Percentage validation of data**: 10%

**Test Data**: None