**Healthcare ML Workshop**

Ryan Kather, Public Top Team, Data Science and IA

A close up of a logo

Description automatically generated

Lab: AutoML in Watson Studio

*August 18, 2019*

*Author: Elena Lowery elowery@us.ibm.com*

**Table of contents**

Contents

[Overview 1](#_Toc64466451)

[Required software, access, and files 2](#_Toc64466452)

[Required skills 2](#_Toc64466453)

[Cloud Pak for Data Access and Project Creation 3](#_Toc64466454)

[Create your Project 4](#_Toc64466455)

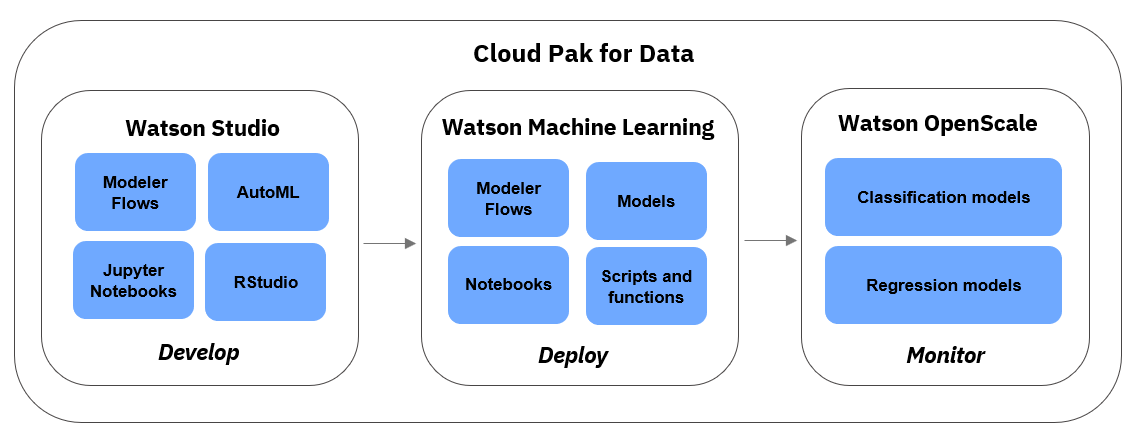
[Additional Resources 9](#_Toc64466456)

# Overview

In this lab you will learn how to use the various Machine Learning components in **Cloud Pak for Data as a Service (CPDaaS)**.

It is worth pointing out at this point that both Guided and Auto ML are concepts indicating the amount of technology support that the analyst receives when performing machine learning tasks, and the capabilities within the Cloud Pak for Data environment which support these concepts are AutoAI (Auto ML) and Modeler Flows (Guided ML).

Cloud Pak for Data includes over 30 services (applications) which can be used for data science, data management, and AI projects. Guided ML (Modeler Flows) and Auto ML (AutoAI) capabilities are a part of **Watson Studio** service. In addition to these tools, Watson Studio includes open source IDEs – Jupyter Notebook and RStudio. Models developed in **Watson Studio** can be deployed in **Watson Machine Learning** and monitored in **Watson OpenScale**.



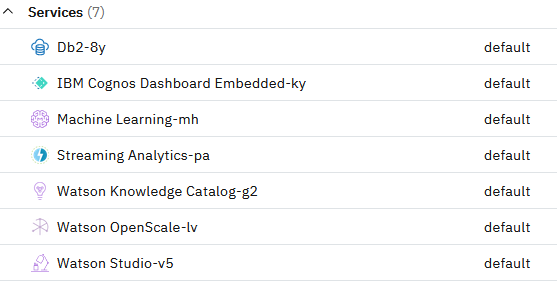
# Required software, access, and files

1. To complete this lab, you will need a **Cloud Pak for Data as a Service** (**CPDaaS**) account: <https://dataplatform.cloud.ibm.com>

* If you don’t have a CPDaaS account, use the same URL to sign up for a free trial. The account will be activated in approximately 5 minutes.

1. If you already have an **IBM Cloud** account, make sure that you provisioned the required services – **Watson Studio, Cloud Object Store, Watson OpenScale,** and **Watson Machine Learning.**

* Navigate to your *Services Dashboard* in your **IBM Cloud** dashboard: <https://cloud.ibm.com/login>
* Check if the mentioned services are displayed under **Services**. If not, search for the services in the **Catalog** and add them.



1. You will also need files from this *GitHub* page: <https://github.com/krondor/UWBazaarProject>

In the lab we will refer to this folder as the git repo folder.

# Required skills

We recommend that users who work through this lab:

* Understand data science lifecycle and steps in creating a model
* Have a working knowledge of Modeler flows.
* Have a working understanding of Python and SKLearn and XGBoost Frameworks.

# Cloud Pak for Data Access and Project Creation

**Watson Studio** provides several AutoML capabilities. In this section you will learn how to use *AutoAI* to automatically build *classification* models. Classification models are used for data science use cases in which we need to predict the *class* of data. The class of data is a specific *value* that we want to predict.

Here are a few examples of use cases that can be solved with classification models:

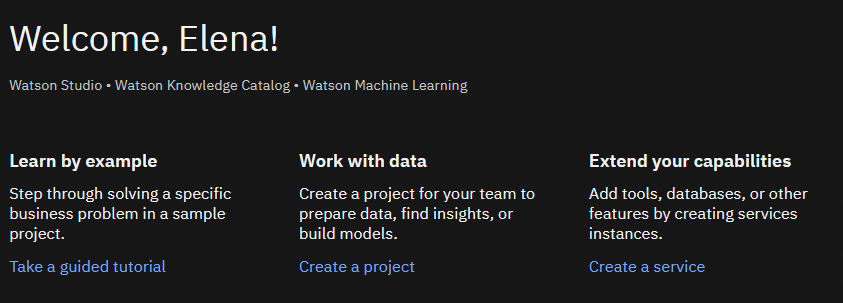
|  |  |  |
| --- | --- | --- |
| **Use case** | **What are we predicting?** | **Sample values (classes)** |
| Customer churn | Will the customer churn? | *Yes* or *no* |
| Credit card fraud detection | Is this a fraudulent transaction? | *Yes* or *no* |
| Predictive maintenance | What is the cause of mechanical failure? | *Overheating, pressure, friction* |
| Customer promotions | Which promotion should we send to the customer? | *Free shipping, buy 1 get 1 free, 20 percent discount* |
| Healthcare | Is this customer at risk for readmission? | *Yes* or *no* |

As you can see from the examples, classification models can be used for many use cases in various industries. In order to create a classification model, we need to have *historical* (*training*) data with the values we are trying to predict. For example, if we want to predict credit card fraud, we need to have records of transactions that were fraudulent and non-fraudulent. Models are created from *statistical algorithms*, and a data scientist can choose from over a dozen of available classification algorithms.

If you would like to learn more about classification models, we recommend that you take an **IBM Data Science Course** from Coursera. Please see the **Additional Information** section.

## Create your Project

1. Log in to **CPDaaS**.
2. On the *Welcome* screen click **Create a project**



1. Select **Create a New Project from a sample or file**.

Graphical user interface, text, application

Description automatically generated

1. If you do not already have a storage service, selected from the right pane, click **Add**.

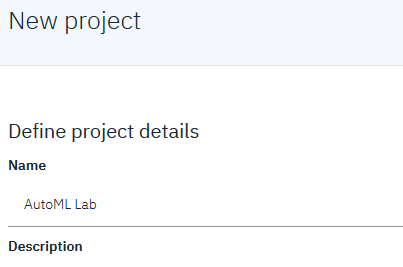
Graphical user interface, text, application, email

Description automatically generated

1. Select the **Lite** plan for a Cloud Object Storage Service and click Create.

Graphical user interface, application, Teams

Description automatically generated



1. Choose the storage option, if it is not visible click **Refresh.**

Graphical user interface, text, application, email

Description automatically generated

1. Continue with the project creation, and after the project completes you will see a view of objects. Browse to the project settings.

Graphical user interface, application, Teams

Description automatically generated

1. Associate WML Service with Project under Settings Tab select the Watson Associated Services Option.

Graphical user interface, application, Teams

Description automatically generated

1. Select your WML Service and apply that to your project.

Graphical user interface, application

Description automatically generated

Now let’s create a deployment space for the project.

1. From the left Hamburger menu, select **View All Spaces** under **Deployments**

Graphical user interface, application

Description automatically generated

1. Choose **New Deployment Space** and then name it your preference. **Record the Name** you created it as for use in the first notebook.

Graphical user interface, application, Teams

Description automatically generated

The project already has a number of data elements, as well as connections to an external database, data flows, and various notebooks. Browse to Assets and Open the **Use XGBoost to Classify Tumors notebook**.

Follow the steps in the notebook.

# Additional Resources

* **AutoAI documentation:** <https://www.ibm.com/support/knowledgecenter/en/SSQNUZ_3.5.0/wsj/analyze-data/autoai-overview.html>
* **Modeler Flows documentation:** <https://www.ibm.com/support/knowledgecenter/en/SSQNUZ_3.5.0/wsd/spss-modeler.html>
* **Modeler Flows Algorithms Guide:** <https://www.ibm.com/support/knowledgecenter/en/SSQNUZ_3.5.0/wsd/spss_algorithms.html>
* **Introduction to Data Science:**
  + IBM Coursera classes: <https://www.coursera.org/professional-certificates/ibm-data-science>