AI For Good

Curriculum

This curriculum is designed to take students through a series of exercises that can be applied to solving real world problems using Data Science, cloud compute, Machine Learning (A.I.). The following is specific but is missing the actual Data Science (A.I.) Solution as one has not been chosen yet. These specifics are part of the creation of the curriculum that will generally follow the guidelines set here.

# Part One: Problem Analysis and Solution Planning

In this section, a scenario will be outlined for the students. The scenario will be like the following:

1. A problem is defined, for example, climate change, or some form of symptom of climate change, to narrow the focus a bit. The problem may be bigger than the proposed project, so if climate change is the problem, the solution may be something like reporting rainforest loss, to determine how much regrowth is needed. Or to count the number of polar bears to see if numbers are declining, or to recognize diseases in forests that accelerate tree loss. Or track and predicts forest fires.
2. The problems should be easy to define using known data sets.
3. The solution(s) will have known data sets that students will ultimately be guided to.

## Cloud usage

The students will learn how to run scripts that load data sets into the appropriate Data store. This can be Synapse (Data Lake), or blob storage to be run by other processes – but the focus here is getting the data (once identified) into the cloud using the appropriate tools. Students will be able to work in a Azure Labs environment that will provide a log-in for the azure account for them to

## Learning objectives

1. Learn how to break a larger problem down into smaller pieces.
2. Learn how to think about the problem and figure out what kind of data might help improve or solve the problem in some way – through some sort of intervention or action taken etc. Even if the only result is to publish the data for others to act on.
3. After determining the data type, students learn where to find this data.
4. Learn how to aggregate data – is it a stream of telemetry or batch files to download and analyze?
5. Learn the data store types in Azure and decide which one is appropriate for the rest of the project.
6. Learn how to log into azure and create the needed resources.
7. Learn how to import data for later use.

# Part Two: Principle Component Analysis

In this section student learn to shape and clean the data sets and put together the notebooks or ML services / Databricks code to run analysis on the data. Essentially this should be a walk through of the process without turning it into a statistics course. The solution will already be known so it will be some studying and experimentations on the data and then implementation of the code needed to output the data in sets that can be used in the next section. In section one we determined what kind of data we would need, and what kind of ML model we will need to build – this step is about making sure we are creating models with only the needed data and not wasting compute on passing through ancillary and unneeded values.

## Cloud usage

Students will use cloud technologies like Databricks, Synapse, or others to experiment with data analysis. The students will store the outcome of the analysis for use in the next section.

## Learning objectives

1. Learn how to run analytics on data in Azure
2. Learn how to store transformed data for use in Machine learning
3. Understand the reason behind the actions they take to get to the data sets (what they are going to be doing with it later)

# Part Three: Build a Machine Learning Model

In the first section we learned how to approach a problem and imagine how AI could be used to help. Then we put that into action by aggregating the data. In section two, we learned how to munge the data; eliminating bulk in files to get clean data sets with the data we need.

In this section we will run the data through ML algorithms to create a model that will allow us to input new information and get predictive or classification results. Depending on which AI solution is used these paths will vary in what is used.

## Cloud usage

Students will access the cleaned data in the cloud and use ML resources in Azure to train a ML model for use in the real world. This can be DataBricks, ML Studios, Jupyter notebooks etc.

## Learning objectives

1. Learn how to create ML resources in azure and access the clean datasets.
2. Learn how to train a ML model using the tools
3. Learn how to output the model for use in the real world

# Part Four: Publish a Machine Learning Model

In this section we will have a working model and will create some form of API access or run the model on an edge device (like a Windows 10 machine) where we can run the model with live data and get results back. The results should be displayed in some form of graphic (Power BI) or web page

## Cloud usage

In this section the students might use a web application to upload images or data sets. Or just to display the results of data sets (powerbi.com) that are being run from an automated pipeline in Azure. This will be dependent on which AI project is used. Could also involve writing results to Cosmos DB and then consuming those results in a web app.

## Learning objectives

1. Learn how to set up a repeatable pipeline for ML model
2. Learn how to set up a repeatable pipeline for new data
3. Learn how to display the outcomes and keep them separate from other outcomes.

# Summary

In summary, the students will go through a complete Data Science cycle from analyzing the problem and settling on a hypothesis for a solution or something to help understand the problem better, to gathering and cleaning data sets. From there, the data sets are used to train a machine learning model that can take the data in and output a response that is helpful (i.e. is a polar bear or is not a polar bear). Finally, the students will ensure that a pipeline is available for the data to go through the ML pipeline and output a readable result. This can be text and graphics or even images. The results will be able to determine important details from the data it receives.

The Students will use Azure Cloud throughout the course and learn about data ingestion, data storage, machine learning and cognitive services and web apps / power BI as well. Students will learn how to monitor the ML process as well as monitor Azure resources.