Machine Learning on Cloud

According to Microsoft:

Machine learning is a data science technique that allows computers to use existing data to forecast future behaviors, outcomes, and trends.

*As you move forward, you will learn how to use****Azure Machine Learning Studio****as an integrated, end-to-end data science and advanced analytics solution, enabling data scientists to prepare data, develop experiments, and deploy models at cloud scale.*

##### How Does it Work?

(1) **Azure Machine Learning Studio** is a cloud service in Azure that enables you to experiment with data and create predictive models.

(2) An **Experimental Dataset** is taken and a workflow of transformations to apply to the data are defined.

(3) **Different models** are trained and tested to create a predictive experiment model with required accuracy.

(4) The **predictive experiment** encapsulates the machine learning model and the associated data transformations, ready to be used with new data.

(5) Your predictive experiment can then be **published as a web service**, which client applications and processes can call to generate predicted values.

##### What is it Based on?

Azure Machine Learning is built on top of:

* Jupyter Notebook
* Apache Spark
* Docker
* Kubernetes
* Python
* Conda
* Microsoft Machine Learning Library for Apache Spark
* Microsoft Cognitive Toolkit.

Azure ML is compatible with **scikit-learn, TensorFlow and Spark ML**

##### How the Components Work Together?

* Python and Jupyter Notebook together provide the basic scripting capabilities in the Azure ML Studio.
* Conda acts as a package manager for compatible python packages like sckit-learn, matplotlib, TensorFlow, etc...
* Docker and Kubernetes provide the underlying environment for the webservice deployment.
* Apache Spark and the Dependant Libraries help in Data gathering from Big Data Jobs which can be linked with the Azure ML Studio.
* Apart from these underlying components Azure ML Studio can further be connected to services such as Azure Storage, Azure SQL Storage, etc.. facilitating a single environment for the whole Data Science life-cycle.

##### Other ML Tools

In addition to Azure Machine Learning Studio, there are a variety of options at Microsoft to build, deploy, and manage your machine learning models.

* Machine Learning Service
* SQL Server Machine Learning Services
* Microsoft Machine Learning Server
* Azure Data Science Virtual Machine
* Spark MLLib in HDInsight
* Batch AI Training Service
* Microsoft Cognitive Toolkit (CNTK)
* Azure Cognitive Services

##### Machine Learning Service

Closely related to Machine Learning Studio, Machine Learning Service is an offering in Azure from Microsoft which provides the utility similar to Azure ML Studio.

It is currently in preview and provides the environment with much better support to open frameworks in python such as Tensorflow, sci-kit learn etc..

It supports Jupyter Notebooks, Visual Studio Code Tools for AI, Azure Batch AI and Containerised Deployment.

##### ML Service vs Azure ML Studio

When compared with the Machine Learning Service, Azure Machine Learning Studio provides pre-configured algorithms and a drag-drop interface for ML.

Azure ML Studio is recommended for beginners to **Experiment with ML modelling and deployment and for small-scale, low complexity business problems**.

However, Machine Learning Service provides a professional and full-fledged environment for solving complex business problems and handle global scale cloud deployment of services.

##### Moving Forward

As you move forward with the course, you will learn

* How to interact with Azure ML on the web.
* Use Azure ML functionality to clean data, create ML models and deploy them as web services.
* Consume the web services in real-time client applications.
* Re-train models using Azure Big data factory.
* Using Azure ML for IoT and Big data Applications.

Azure ML Platform

This helps understand using Azure ML platform on the web to generate predictions by creating and deploying ML models.

The workflow followed to generate web services from datasets is as follows:

1. **Train a Machine Learning Model** with training Experiment.
2. **Create a Predictive experiment** from training experiment encapsulating data transformations and training model.
3. Use the Predictive experiment to **create a web service** to generate predictions using API-endpoint and API key.

Workspace

Workspace is provision given on Azure Subscription which can be referred to as a playground for Machine Learning Experimentation.

A Workspace contains

* Training Experiment
* Predictive Experiment and
* Web Service collection for a user.

To explore the Azure ML Platform, we must have an Azure subscription and must create a workspace.

##### Notebooks

Azure ML platform has built-in support to execute R and Python scripts using Jupyter Notebook.

Notebooks can be used to transform, clean, visualise data and train models according to requirements.

They can be placed as part of an experiment or can be used for visualisation/transformation separately for particular datasets.

Azure Notebooks support R, Python 2 and Python 3.

##### Projects

Projects further enable us to easily manage various resources.

Different datasets, notebooks and resources used in experiments can be managed/organised easily by adding them to a single project.

The Projects tab gives summaries of all the assets i.e experiments, notebooks, and datasets used by us and added to the project.

Project layout enables easy management of various assets related to our experiments.

##### Cortana Intelligence Gallery

The Cortana Intelligence Gallery is a collection of resources people have shared on the Azure ML platform.

It includes various tutorials, datasets and experiments.

Experiments can be copied from the gallery into a workspace and can be explored/modified for a better understanding.

We can also share our work on experiments and web services for others to learn and explore.

Move to the next topic to start creating and experimenting on ML models.

Datasets

An Azure ML experiment requires at least one dataset on which the model is created. The data can be imported directly or into Azure Storage and used for creating a model.

Generally, in the case of Big Data Sources:

* Hive or U-SQL jobs are used to clean and prepare for analysis.
* A business process stores large volumes of data in a database or data warehouse.

This data can be stored on Azure Storage and can be easily imported to Azure ML Workspace. Data can also be imported from Hive or Azure SQL database.

*Further in the topic you will learn different types of data used in experiments and ways to use them*

Big Data Sources - Considerations

When working with Big Data sources the following points must be considered:

* Multiple files can be uploaded to the Azure ML workspace but maximum size is less than 2 GBs.
* However, import up to 10 GB of data from other sources is possible.
* If you need to work with even larger volumes of data, **statistical sampling** techniques must be used to sample ten gigabytes of data for training.

Training Data

Datasets are broadly classified into two types in Azure ML perspective.

* **Training Data** to train the Experiment
* **Reference Data** providing additional information.

**TRAINING DATA :** The Data to train the Experiment is mandatory data and acts as the start of Dataflow.

*Move to the next card to learn about reference data.*

##### Reference Data

**REFERENCE DATA** is not directly used in the experiment or for training a model.It is only used to provide additional information.

For an example when a web service is published based on the ML model, a service endpoint is created to consume the web service.

Based on the client using the service, the reference data to be used can be varied accordingly to provide customisation and flexibility to the web service.

Data Import - Trivia

In this topic, you have learnt ways to import data from various Azure Data stores and the types of data used. Few other notable points are:

* Data can also be imported from On-premises SQL Server or several other online sources using Import module.
* Multiple Data formats such as .txt, .csv, .nh.csv, .tsv, .nh.tsv, Excel file, Azure table, Hive table, SQL database table, .RData etc.. are supported.
* Data types recognised by ML Studio are String, Integer, Double, Boolean, DateTime and TimeSpan.

*For more details refer the following:*

* [Import Data](https://docs.microsoft.com/en-gb/azure/machine-learning/studio/import-data)
* [Prepare Data](https://docs.microsoft.com/en-gb/azure/machine-learning/team-data-science-process/prepare-data)
* [Scenarios for advanced analytics](https://docs.microsoft.com/en-gb/azure/machine-learning/team-data-science-process/plan-sample-scenarios).

##### The Lifecycle

A Data Science solution involves:

- Identifying a Business Problem

- Data Extraction

- Data Cleaning

- Data Transformation

- Exploratory Data Analysis

- Data Visualisation

- Machine Learning Modeling

- Publishing the Prediction Service

##### Adopting ML Studio

**Azure Machine Learning Studio** helps through the entire lifecycle of a Data Science solution.

We already had a look at Data Import from various sources, which can be considered as Data Extraction.

Data Cleaning and Transformation are done based on our business problem and the approach we take to provide a solution for it. You can see this in the videos over the following cards.

Data Cleaning and Transformation

Data Cleaning refers to detecting and correcting missing, inaccurate and irrelevant parts of data while Data transformation refers to converting data from one format to another.

Azure ML Studio provides these features through Filters, Scale and Reduce, Manipulation and Sample and Split.

* Filters : Transforms and cleans digital data and can help in Speech Processing
* Manipulation : Cleans missing values, Meta data editing, SQL transforms, etc..
* Scale and Reduce : Normalization, Grouping, Clipping, etc..
* Sample and Split : Partition and Sample data.

Check [this](https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/data-transformation), for more info on Data Cleaning and Transformation.

##### Data Visualisation

Exploratory Data Analysis and Data Visualisation are facilitated by Notebooks in the Azure ML Studio.Data Visualisation is readily available in Azure ML Studio over the right-click of uploaded datasets.

However, Notebooks can be used to further visualise data in a required manner with Python/R scripts adding flexibility and functionality.

Watch the following video to learn using Notebooks for data visualisation.

##### Machine Learning Modelling

Coming over to devising a Machine Learning Model, Azure ML supports **25 Algorithms**.

They are divided into 4 major classes for both Structured and unstructured data.

- Regression

- Classification

- Anomaly Detection

- Clustering

Go through Machine Learning Axioms and other ML courses for model selection and feature selection.

##### Supervised Learning

Azure ML Platform is equipped with over 20 types of supervised learning methods in-built and ready to use.

Users can also write their own python scripts and embed them into the ML workflow for customised and optimised models using Notebooks and supported ML libraries like sci-kit learn, Tensorflow, etc.

Move to the next cards to check out examples of various Supervised Learning Algorithms along with concepts of Data cleaning and Transformation.

##### Regression Model

Watch this video to understand how to train a regression model from a sample dataset using Azure ML Experimentation service.

##### Classification Model

Watch this video to understand how to train a classification model from a sample dataset using Azure ML Experimentation service.

##### Moving Further

Supervised learning is the widely used ML modelling technique for structured data.

Different algorithms have their own pros and cons. Our requirements must be decided on to select the optimal algorithm.

Refer the following links for a better understanding on Supervised learning methods w.r.t Azure ML platform: [*Feature Engineering*](https://docs.microsoft.com/en-gb/azure/machine-learning/team-data-science-process/create-features), [*Algorithm Selection*](https://docs.microsoft.com/en-gb/azure/machine-learning/studio/algorithm-choice) and [*Evaluating ML Model*](https://docs.microsoft.com/en-gb/azure/machine-learning/studio/evaluate-model-performance).

##### Unsupervised Learning

Unsupervised Learning is used to work on unstructured and un-labelled data.

Clustering is the most commonly used method where similar data is grouped by finding features and grouping them based on their feature set.

Azure ML Platform supports K-Means Clustering for Unsupervised Learning.

##### K-Means Clustering

Watch the video to learn using K-Means Clustering Algorithm and visualise data after clustering.

##### Recommenders

Recommenders, as their name suggests, are used in sectors of e-commerce, ads and social platforms etc. They recommend related items of interest based on users previous interaction.

Watch the following video to learn to model a recommender on Azure ML platform.

##### Using the Trained Models

In the previous topic, training different ML models using Azure ML Platform is discussed.

Once the trained models are tested for accuracy and optimised, they need to be deployed for consumption through API.

This topic helps to learn to deploy the trained model as a web service using a predictive experiment.

Move through the cards to find out..

##### Predictive Experiment

Predictive Experiment is created from successful Training Experiment.

* This defines the workflow of processing web requests to generate predictions.
* It is similarly structured to training experiment but the data transformations like normalisation, are encapsulated into a single block in order to not modify the training experiment.
* Webservice input and output will be the new blocks added to the experiment.

##### Webservice

A Webservice is to be created from a Predictive Experiment to **consume the generated ML model**.

* **Creating a Webservice** generates API end-points which can be accessed through Primary and Secondary API-keys.
* The **API end-points** take inputs required by the ML model and given JSON output with predictions.

Webservice Workflow

The following workflow is adopted for building a web service with predictive experiment:

1. Create a training experiment and optimize it.
2. Generate a predictive experiment from the trained model along with operations like data normalization, encapsulation and transformation.
3. Include definitions for web service inputs and outputs that define how client applications will consume the model.
4. Modify predictive experiment without affecting training experiment in accordance with client requirements.
5. Finally, publish the predictive experiment as a web service that can be accessed through a URL or endpoint.

Web service usage

* Client applications can call the web service, passing in the features and getting the output according to the schema.
* When new training data is gathered, the training experiment can be re-opened and improved by retraining.
* Predictive experiment can be updated accordingly while taking consideration of transforms that encapsulate the trained model and other operations required while preserving other changes of the predictive experiment.
* With updated predictive model the web service can be re-deployed and client applications can continue to access through the same endpoint as before.

Diving Deep - Web services

This topic covers the

* Management
* Consumption and
* Customizing of Web services according to the client requirements.

You will also learn about metrics and logging regarding the usage of Web services.

##### Webservices - Factors to consider

After models are created and published as API, we need to make them consumable by various client applications.

Web services from Azure ML Platform are flexible.

* Web services can be made to retrieve reference data from a database.
* Parameters can be passed to retrieve reference data from the required database to service different clients.
* Can be managed and monitored with respect to utilization and activity through logging.

Consumption

Web service can be consumed from the published API either in

* Request-Response mode or
* Batch Mode in asynchronous way.

API endpoints and API Keys are used according to the requirement.

* The APIs are built as REST APIs and can be consumed by required client application by passing required parameters as an HTTPS request.
* Microsoft Excel along with Azure ML Plug-in can also be used to consume the service.

##### Parameters

Web Service can be customised to take required parameters to access additional data.

For example, you can specify a database to fetch the data. It will be provided as an additional parameter apart from the required parameters for a prediction.

This helps in customised client consumption such that each client will use the same service for prediction but data output/retrieval is different for each of them.

##### Monitoring

Azure ML Platform enables Web service management through the Usage Statistics and Logging.

Dashboards provide an overview of the total no.of requests made to the API and their success/fail rate over a selected period of time. They also give the average compute time and latency associated with the API.

Logging can be enabled for a more detailed JSON response files stored automatically on Azure storage providing a detailed report of the request and response.