Carolyn McAughren
558 755 906

CSCI 479 Advanced Topics in Artificial Intelligence
November 9th, 2022

Machine Learning Report

Microsoft Azure Machine Learning Platform

Table of Contents

Introduction

Microsoft's Azure Machine Learning platform is a cloud-based service which helps teams with their entire machine learning project lifecycle (Microsoft, 2022a). They have systems to build, train and deploy models (or use a model already built in an opensource platform) and to manage the machine learning pipeline, including monitoring, retraining, and redeploying of models. Azure ML platform helps teams collaborate in a shared environment to deploy ML models quickly and easily at scale, improving productivity (Microsoft, 2022a).

Azure ML is a "Machine Learning as a Service" (MLaaS) product. There are many other MLaaS platforms in competition with Azure ML, such as Amazon Machine Learning services, Google AI platform, and IBM Watson Machine Learning. These services cater to the many businesses who want to utilize ML without too much investment or advanced knowledge needed (AltexSoft, 2022). Azure ML has a leg up on the competition by drawing in a lot of enterprises who are already working with Microsoft Azure cloud for other services in their companies. Many companies already work with and are familiar with Microsoft suite tools, making Azure ML a natural choice.

Microsoft Azure ML services can be utilized through many interfaces, such as the popular Python SDK, but Microsoft created Azure Machine Learning Studio to act as UI for the Azure ML toolset. This ML platform is intimidatingly extensive, so I have focused my examination and review of this platform through the lens of the Microsoft Azure Machine Learning Studio specifically. The Azure ML Studio is web-based and has many low or no-code tools, so it is a convenient way to manage your ML projects with a low barrier for entry. The studio has many features, including "Azure Machine Learning designer" which has a drag and drop UI to help those who cannot code to modify their data, apply ML methods and to deploy ML solutions (AltexSoft, 2022).

Platform functionality

The Azure ML studio platform has a plethora of options available, some which are advanced and some that are low or no code. The studio has three main sections dividing its menu of options: Author, Assets, and Manage.

The Author tab handles the set-up work needed to design the pipeline which will be used to build and deploy the ML model (Marius, 2021). Under Author, you can build your workflow using “Notebooks”, “Automated ML” or “Designer”.

In Notebooks you can share Jupyter Notebooks or provide your own code files to use in experiments. In Automated ML (AutoML), you can find and train a model without any code – you select the ML “problem” (classification, forecast, regression etc.), specify the source of your data and configure parameters (number of iterations, what metrics should be used to decide on the best model etc.) but there is no need to select an algorithm. You simply submit the job to the AutoML tool and it will select the ideal algorithms for your task, iterate through multiple models, score them, and present you with the results (Microsoft, 2022b).

Under Author, there is also “Designer” mode, which is a drag-and-drop interface used to design a ML workflow. It has extensive features called ‘components’ which you drag onto a canvas and connect together with arrows to design your ML model pipeline. These components cover all aspects of the ML process, including Data Preparation steps (e.g. Data Input and Output methods and Transformation of Data steps such as cleaning missing data and joining datasets); Machine Learning Algorithms (19 options listed under the umbrellas of Regression, Clustering, or Classification categories, and more under other categories); and tasks related to the building and evaluation of the models (model training, scoring and evaluation).

In Author, the Notebook options allow experienced ML engineers to use advanced and highly customized techniques and AutoML and Designer can be used by those who are new to ML. They offer fast, easy ways to deploy a ML model without a lot of overhead.

The Assets section of the Azure ML studio holds resources previously built by the user during the Author stage (saved here so you may reuse them or tweak them as needed) and here the user will control the workflow for the project (Marius, 2021). There are sections for management of Datasets, Experiments (controlling metrics

and hyperparameters), pipelines (created via Designer), Models, and Endpoints (systems and application which can use the models after deployment).

Then, the Manage section of the menu handles behind the scenes tasks such as computation clusters, instances, where datasets are stored and how they are integrated into other systems (Marius, 2021). In addition to the Compute (cluster management), Data Stores, and Linked Services sections, you can also handle Data Labeling under the Manage tab. There is extensive support to help you complete your dataset labelling, like help dividing the job out to multiple labelers who are working on your project in Azure ML platform, and automation tools which use ML to assist you in the labelling of your dataset.

The Azure Machine Learning Studio is expansive and exhaustive in its options for ML model building, at all stages of the process, and this is just Microsoft's flagship platform to access these tools. Azure ML services are also available for use on other popular platforms, and there is built in compatibility with opensource models which are built outside the studio. The Azure ML platform was built to be an all-encompassing, one stop location for all your ML needs and it is still ongoing in its development and refinement.

Algorithms used by the platform

Microsoft Azure ML platform has a plentitude of algorithms available for use in the Designer and AutoML tools which are available under the Author tab. The AutoML tool selects the algorithm for you based on your specified programs and goals for your model. Designer mode gives you more power over your models by allowing you to select your preference.

The AutoML tool selects an ideal algorithm for you based on your problem definition. It supports data-based tasks (classification, regression, forecasting), computer vision tasks (image classification, object detection), and natural language processing (text classification, recognition tasks). The list of algorithms it may use behind the scenes include 12 options for Classification (including Decision tree, KNN, Gradient Boosting, Naïve Bayes), 12 options for Regression, 4 for Time Series Forecasting, and several others listed under a 'K nearest Neighbors' category (Microsoft, 2022c).

There are also many algorithms available in the Designer workspace. Microsoft created a helpful 'cheat sheet' flow chart to help users identify the best algorithm for their task. There is also a lot of documentation under Microsoft's Learn website explaining how to evaluate your data and problem set to find the best algorithm fit, as well as documentation explaining all of these algorithms in detail (Microsoft, 2022b). There is a total of 31 algorithms outlined in the cheat-sheet. Those listed under Regression, Clustering, Multiclass Classification and Two-class Classification are listed in the "Algorithm & component reference for Azure Machine Learning designer" documentation for the Azure ML platform (Microsoft, 2022d) along with links to articles explaining each algorithm in detail. There is also documentation outlining how all the other algorithm's work under links for "Text analytics", "Computer Vision", "Recommendation" and "Anomaly Detection".

Notably, there is only one algorithm available which can handle clustering: the K-Means algorithm. The documentation outlining this algorithm describes it as "one of the simplest and best-known unsupervised learning algorithms" (Microsoft, 2022e). It is used for detecting abnormal data, clustering text documents, and analyzing datasets before the use of other classification or regression models, as a preprocessing step. The algorithm is unsupervised, so it does not need labelled data. If the data set you provide the K-means cluster algorithm is labelled though,

the algorithm will use these to help it guide selection of clusters and optimize the model (Microsoft, 2022e).

The K-means clustering algorithm uses iterative techniques to group elements into clusters which have similar qualities. The documentation reference recommends it be used to explore data, identify abnormalities in the data (like outliers) and for making predictions (Microsoft, 2022e). It is ideal for helping to "identify relationships in a dataset that you might not logically derive by browsing or simple observation" and is useful in the early ML phases to help discover unexpected correlations (Microsoft, 2022e).

To use the K-means clustering algorithm you will specify k, the number of centroids you want to use in your model. This point will be the reference for each cluster, and the algorithm will assign all data points to one of these k centroids by "minimizing the within-cluster sum of squares" (Microsoft, 2022e).

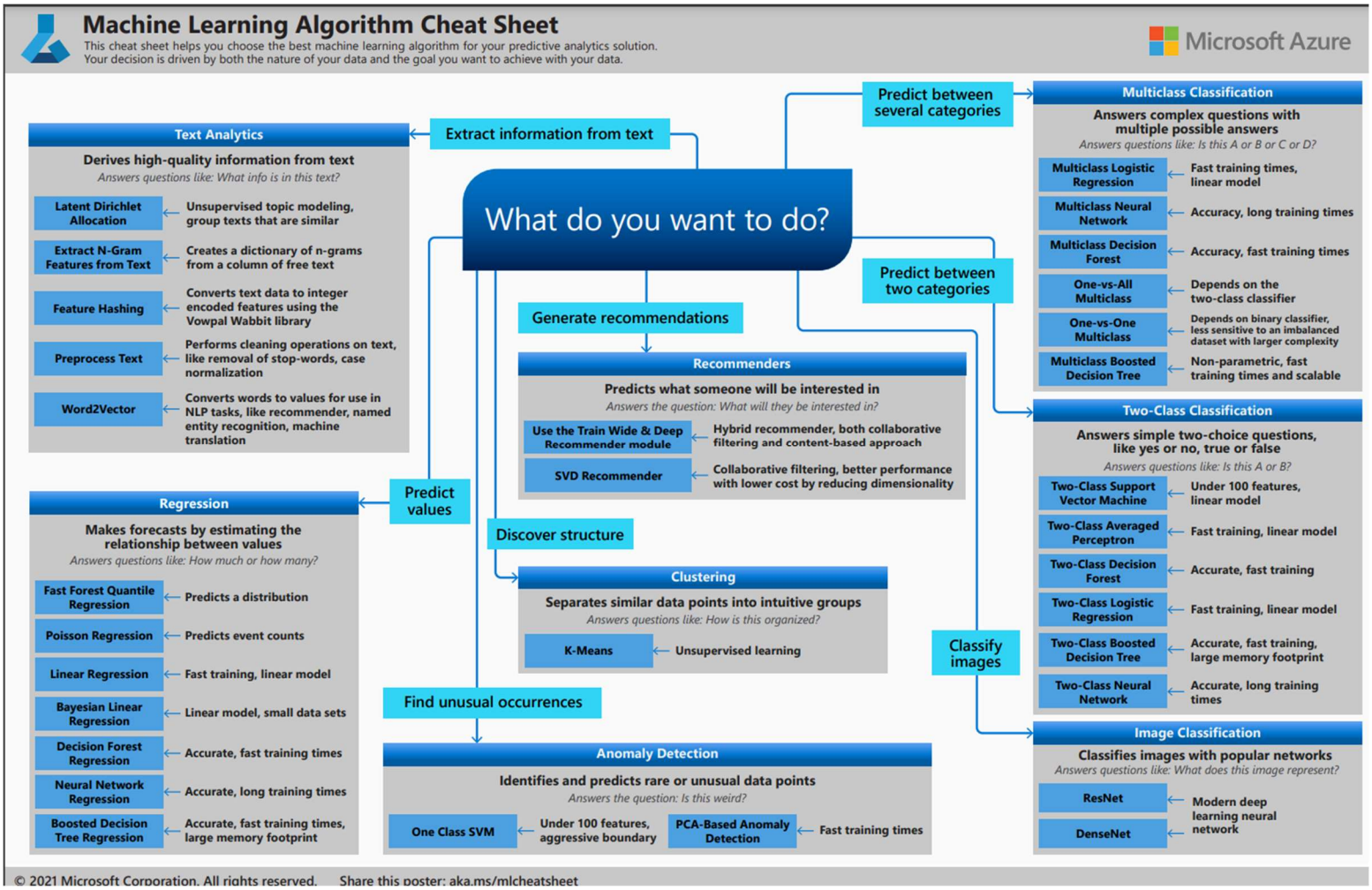
The algorithm begins with a set of randomly selected centroids, then uses "Lloyd's algorithm to iteratively refine their locations" (Microsoft, 2022e). It will stop iterations when one or more of these stopping conditions are met:

- Centroids stabilize: the cluster assignments for each point no longer change in new iterations.
- The specified max number of iterations has been reached.

After the training phase is complete, there are tools in Designer to help you use the model to assign new data items to the clusters. They do this by assigning the new data point to the cluster whose centroid is nearest (Microsoft, 2022e).

Figure 1

Machine Learning Algorithm Cheat Sheet



Note. Algorithms available for use in Azure ML platform are outlined in a flowchart. (Link for download available at: Microsoft, 2022d)

Strengths of the platform

- The Microsoft Azure Machine Learning platform works for both inexperienced and experienced data scientists. The platform has an extensive selection of algorithms prebuilt in it, on top of their Notebook options which leave the models open ended for custom algorithms and workflows.
- There is a plethora of documentation which covers all aspects of the ML process and each of the algorithms and data analytic tools which are available on the platform.
- The barrier for entry is quite low for those who have little or no programming experience – drag and drop UI make it accessible to those who cannot code and the thorough documentation available will walk them through every step of the process.

Weakness of the platforms

- This platform has a step learning curve due to its many complicated and advanced features.
- I also personally struggled with the sign-up process, which has to go through Azure Cloud to set up a 'workspace' before you can get into the Azure ML platform.
- The issue of price. Azure ML platform is Saas and as such has a steep price tag for anyone with a large amount of data or an enterprise level projects. While it may be ideal for small teams or businesses who wish to take advantage of ML technology but do not have the resources to hire specialized ML engineers, the cost of hosting in the cloud may not outweigh the convenience for larger companies.

How to use the platform practically

Microsoft offers a free tier Azure account which has many free services available for the first 12 months after signing up. This also provides you with around $200 in credits which you can use as you please. This is great way to test out the Azure ML platform without commitment.

You can sign up for the free trial here: https://azure.microsoft.com/en-gb/free/machine-learning/search/

The Azure ML Studio platform is a web application, so no software is required for download (but there are many other platform options where you can use the Azure ML services, such as VScode or Python SDK).

Once you have signed up for the Azure free trial, you can navigate to the Azure ML Studio platform here: ml.azure.com

However, from my experience, you can't build a workspace to get started from that link. You must create a workspace back in the Microsoft Azure portal: portal.azure.com

This was a confusing step, but I think the reasoning is that all ML workspaces must be associated with an Azure portal for billing purposes, so it's better that the portal account handles the creation of workspaces rather than the individuals who will be working in that space. This means the Azure account can have multiple ML workspaces associated with it and linked under one team umbrella, and multiple developers can be added onto a team under one workspace and centrally billed Azure cloud account.

Once I figured out how to create the workspace in Azure portal, I found the rest of the process very easy, I was quickly able to get into the Azure ML platform and explore the extensive options available for building ML models. Each section under the left-hand menu has links to documentation. There is help and tutorials available at every step of the way.

One helpful tutorial provided in the Microsoft Learn documentation for Azure ML helps new users to master the platform: https://learn.microsoft.com/en-us/azure/machine-learning/tutorial-azure-ml-in-a-day

Resources

AltexSoft. (2022, April 25). *Comparing Machine Learning as a Service: Amazon, Microsoft Azure, Google Cloud AI, IBM Watson.* **https://www.altexsoft.com/blog/datascience/comparing-machine-learning-as-a-service-amazon-microsoft-azure-google-cloud-ai-ibm-watson/**

Marius, H. (2021, October 8). *A Brief Introduction to Azure Machine Learning Studio.* Towards Data Science. **https://towardsdatascience.com/a-brief-introduction-to-azure-machine-learning-studio-9bbf41800a60**

Microsoft. (2022a, December 2). *What is Azure Machine Learning?* Microsoft Learn. **https://learn.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-machine-learning**

Microsoft. (2022b, October 31). *What is automated machine learning (AutoML)?* Microsoft Learn. **https://learn.microsoft.com/en-gb/azure/machine-learning/concept-automated-ml**

Microsoft. (2022c, February 12). *Set up AutoML training with the Azure ML Python SDK v2.* Microsoft Learn. **https://learn.microsoft.com/en-gb/azure/machine-learning/how-to-configure-auto-train#supported-algorithms**

Microsoft. (2022d, October 12). *Algorithm & component reference for Azure Machine Learning designer.* Microsoft Learn. **https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/component-reference**

Microsoft. (2022e, November 4). *Component: K-Means Clustering. Microsoft Learn.* **https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/k-means-clustering**