

Machine Learning for the Civic Tech

Introduction

Cities and mission-driven organizations have been using data and technology to solve some of society's most pressing social challenges for some time now. From performance measurement to program evaluation to more advanced analytics, civic stakeholders are demonstrating the value of data. And while there is no shortage of success stories and use cases, there is room to do more. In many cases, applying machine learning is a logical next step from the civic analytics work currently being done.

One thing we've learned about the civic space through our work in cities is that there are often barriers to using machine learning; but, it's not impossible. In fact, with [Azure Machine Learning](#), we're trying hard to make it more accessible, despite the capacity and tech barriers civic organizations often face.

Okay... great! But what exactly is machine learning?

Machine learning is all about patterns. While humans are not particularly adept at uncovering patterns in large amounts of data, it turns out that machines are extraordinarily good at it. When we understand patterns, it becomes easier to predict outcomes, which in turn helps us solve problems. In addition to predicting outcomes, machine learning can help us classify or group data in meaningful ways. It can also help us understand what, if any, anomalies may exist in our data.

How can machine learning be applied in civic tech? Oh... and what is civic tech?

We define civic tech as using technology for the civic good.¹ One way to do that is through data analytics, which includes machine learning. By using machine learning to uncover patterns related to complex social problems, we can inform public policy intervention or operations and create deployable, data-informed solutions.

What's an example of machine learning being applied to the civic tech space?

Microsoft recently partnered with [DataKind](#), a non-profit organization that provides pro-bono data science support for mission driven organizations, to use machine learning and data science to improve pedestrian safety in three US cities. Data scientists at DataKind developed models to examine street characteristics' impact on pedestrian injuries and estimate the volume of traffic on city streets, a key measure for safety and traffic planning. The results of this work are available for [further reading here](#).

Are you sure I don't need to know how to code or be a data scientist to use machine learning?

Yes! [Azure Machine Learning Studio](#) is Microsoft's machine learning service, which has an easy to use drag and drop interface, pre-developed algorithms, and [examples of experiments](#) that you can build off or adapt for your own work, all without having to use code. It's a great environment to learn in. And, if you have coding or data science skills, you can use those as well—Python and R scripts are easy to deploy (but not required!). [The full documentation is here](#). Note that it is helpful to have some working knowledge of statistical concepts, such as linear regression, before you deploy your work.

¹ There is no consensus on a definition of civic technology, even with the civic tech ecosystem itself. We take a broad definition to ensure that we capture all the good work being done by civic technologists everywhere.

Learning Resources

Here are several learning resources to help you get started:

The [Microsoft Professional Program for Data Science](#) is a great way to learn the basics of data science and Azure Machine Learning. The program is offered through [edEx](#), and you can audit any of the courses for free. If you would like the official certificate, there is a charge for each class, but it's not required. Here are three noteworthy machine learning courses from the program:

- [Data Science Essentials](#)
- [Principles of Machine Learning](#)
- [Applied Machine Learning](#)

If you prefer more of a self-paced, hands-on approach, try out the labs below from Microsoft's [Data Science and Machine Learning Essentials Course](#) on Github. Be sure to [download the zip file](#) with all of the data before you begin. You can also access these labs and the data [directly from GitHub](#).

| | |
|---------------------------|--|
| Lab Setup | Create Accounts and Prepare Your Workspace |
| Lab 1 | Getting Started with Azure Machine Learning |
| Lab 2A | Acquiring Data in Azure Machine Learning |
| Lab 2B | Transforming Data with Scripts |
| Lab 2C | Data Quantization |
| Lab 2D | Cleaning, Transforming, and Integrating Data |
| Lab 3A | Visualizing Data |
| Lab 3B | Building Models in Azure ML |
| Lab 3C | Evaluating Models in Azure ML |
| Lab 4A | Working with Regression Models |
| Lab 4B | Working with Classification Models |
| Lab 4C | Working with Unsupervised ML Models |

For those of you who prefer to explore on your own, be sure to check out our [Algorithm Cheat Sheet](#) to help you choose the right algorithm for your data science project.

The [Cortana Intelligence Gallery](#) has a wide range of pre-designed experiments that make for excellent training exercises and use cases. You can also use them as a starting point for your own projects and adapt them to fit your data and desired outcomes.

Lastly, if you're thinking about how you might apply your new machine learning skills for social good in the real world, look no further. Our wonderful friends at Harvard's [DataSmart City Solutions](#) have put together this robust [catalogue of civic data use cases](#) to give you ideas for improving your community.

Getting Started

If you're new to Azure Machine Learning, a good way to get started is by following the *Lab Setup* instructions from the training menu above. If you prefer a more independent approach, [sign up for free access](#) to Azure ML Studio (select the free option), or [begin via the Azure portal](#) for \$200 in Azure credit, which you can use for Azure ML or other select Azure services. Most importantly, have fun!

- The Microsoft Cities Team