MLOps
Whiteboard design session student guide
November 2019

Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, e-mail address, logo, person, place or event is intended or should be inferred. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The names of manufacturers, products, or URLs are provided for informational purposes only and Microsoft makes no representations and warranties, either expressed, implied, or statutory, regarding these manufacturers or the use of the products with any Microsoft technologies. The inclusion of a manufacturer or product does not imply endorsement of Microsoft of the manufacturer or product. Links may be provided to third party sites. Such sites are not under the control of Microsoft and Microsoft is not responsible for the contents of any linked site or any link contained in a linked site, or any changes or updates to such sites. Microsoft is not responsible for webcasting or any other form of transmission received from any linked site. Microsoft is providing these links to you only as a convenience, and the inclusion of any link does not imply endorsement of Microsoft of the site or the products contained therein.

Contents

MLOps whiteboard design session student guide

Abstract and learning objectives

In this workshop, you will learn how Trey Research can leverage Deep Learning technologies to scan through their vehicle specification documents to find compliance issues with new regulations, and manage the classification thru their web application. The entire process from model creation, application packaging, model deployment and application deployment needs to occur as one unified repeatable, pipeline.

At the end of this workshop, you will be better able to design and implement end-to-end solutions that fully operationalize deep learning models, inclusive of all application components that depend on the model.

Step 1: Review the customer case study

Outcome

Analyze your customer's needs.

Timeframe: 15 minutes

Directions: With all participants in the session, the facilitator/SME presents an overview of the customer case study along with technical tips.

1. Meet your table participants and trainer.

2. Read all of the directions for steps 1-3 in the student guide.

3. As a table team, review the following customer case study.

Customer situation

Trey Research Inc. delivers innovative solutions for manufacturers. They specialize in identifying and solving problems for manufacturers that can run the range from automating away mundane but time-intensive processes to delivering cutting edge approaches that provide new opportunities for their manufacturing clients. Trey Research has decades specializing in data science and application development that until now were separate units. They have seen the value created by the ad-hoc synergies between data science and app development, but they would like to unlock the greater, long term value as they formalize their approach by combining the two units into one, and follow one standardized process for operationalizing their innovations.

As their first effort of this combined initiative, they would to define a process for operationalizing deep learning that encompasses all phases of the application life cycle along with model creation and deployment of a deep learning model. For this first proof of concept, they would like to focus on component compliance.

Specifically they are looking to leverage Deep Learning technologies with Natural Language Processing techniques to scan through vehicle specification documents to find compliance issues with new regulations. Even though this first scenario is focused on vehicle components, they believe this approach will generalize to any scenario involving an inventory of components (which all of their manufacturing customers deal with). The component descriptions (which are free form text) are entered and managed via a web application. This web application take new component descriptions entered by authorized technicians and labels the component as compliant or non-compliant based on the text. For the PoC, they have exported all of their labeled component descriptions as flat files (CSV format).

According to Francine Fischer, CIO of Trey Research, they want to ensure the overall process they create enables them to update both the underlying model and the web app in one using a standardized approach. They also want to be able to monitor the model's performance after it is deployed so they can be proactive with performance issues. They believe they can accomplish most, if not all, of this using Azure Machine Learning but are wanting to be certain.

Finally, Trey would like to ensure that the process they establish is not littered with username and password credentials used to access the component used during build and release.

Customer needs

1. Want to understand the best practice process they should follow for end-to-end deployment of deep learning models.

2. Need a solution that addresses the management of the entire model lifecycle, inclusive of monitoring the model in production and being able re-train and re-deploy when a model needs updating.

3. A process that avoids checking credentials into source control.

Customer objections

1. We are not clear about the benefits that using ONNX might bring to our current scenario and future scenario. Also, we are confused about the various choices that seem to be available like Python vs. ML.NET, or Keras/Tensorflow vs. PyTorch. We are concerned about making the wrong technological choices.

2. It seems like data scientists deploy their models as web services from their own python scripts, where as our developers are accustomed to using Azure DevOps to deploy their web services. Can we really have one tool that provides us build and deployment pipelines irrespective of whether we are deploying a model or web application code? Also, can we use the same tool to have a unified approach for real-time and batch scoring scenarios?

3. Obviously, we can't just have new models automatically deployed into production. What kind of safeguards can we put in place? What kind of support does Azure Machine Learning service have for enterprise security?

4. In the future, we envision providing some of the trained models in devices that are embedded into the manufacturer's assembly lines. Will we need a completely different solution and process for that?

5. The business processes of our customers are subject to extensive scrutiny from auditors and regulators. Can we really have traceability through the entire process, including explaining the results produced?

On the same note, will be aligned with the principles of responsible AI?

6. We have heard Azure Machine Learning service supports automated machine learning; can we use automated machine learning to create models using deep learning? Can we incorporate automated ML in an end-to-end DevOps pipeline?

7. Some of our team has worked with Azure Databricks (trained a few models in Spark MLlib), and they are confused by the overlap with Azure Machine Learning service. How should we be thinking about when to use which?

Infographic for common scenarios
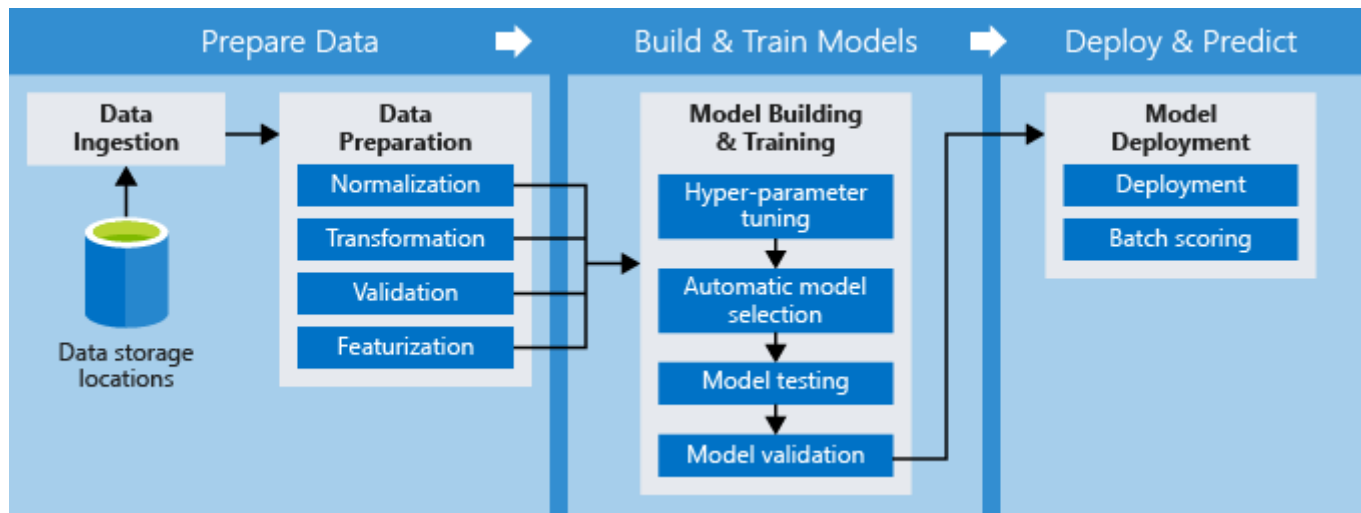
Step 2: Design a proof of concept solution

Outcome

Design a solution and prepare to present the solution to the target customer audience in a 15-minute chalk-talk format.

Timeframe: 60 minutes

Business needs

Directions: With all participants at your table, answer the following questions and list the answers on a flip chart:

1. Who should you present this solution to? Who is your target customer audience? Who are the decision makers?

2. What customer business needs do you need to address with your solution?

Design

Directions: With all participants at your table, respond to the following questions on a flip chart:

High-level architecture

1. Without getting into the details (the following sections will address the details), diagram your initial vision for the solution. You will refine this diagram as you proceed.

Component Classification

1. What is the general pipeline for approaching the training of text analytic models such as this? What are the general steps you need to take to prepare the text data for performing tasks like classification?

2. Provided that Trey wants to build the PoC using Azure Machine Learning, what is the first item they would want to deploy in Azure?

3. Within the above deployed item, what component would they use to orchestrate the various **machine learning** phases (specifically data access, model training and model evaluation)? Be specific about which programming language and framework or SDK they would use.

4. Where would they author any scripts?

5. Where should Trey upload the component compliance data?

6. Describe at a high level the objects involved in the scripted pipeline you would create for training the compliance classification model using Azure Machine Learning.

7. Diagram what happens when you run a machine learning pipeline in Azure Machine Learning.

Enabling DevOps for AI with MLOps

1. Building on the approach you suggested to Trey for machine learning pipelines, what would you propose they use to fold them into a bigger DevOps pipeline for continuous integration and delivery that would result in a new scoring web service being deployed whenever there was a change to the code supporting model training.

2. In your Azure Pipelines design, give an example of what would trigger the execution of the pipeline?

3. What type of Azure Pipeline would be triggered first in response (Build or Release)?

4. What are the core steps in this first pipeline? What does the pipeline output?

5. After the first pipeline, what kind of Azure Pipeline would Trey define to deploy the scoring web service? What are the core steps in this pipeline? What does the pipeline output?

6. How would Trey modify the aforementioned deployment pipeline to enforce that a manual sign-off is performed before the web service could be deployed into the production environment?

7. How could Trey configure their Azure Pipelines so as to avoid hardcoding credentials to access Azure resources?

Monitoring in Production

1. How would you recommend Trey collect diagnostics of the scoring web service in production?

2. How can Trey collect the data input to the scoring web service and the outputs that result, such that they could monitor how the model is performing in production?

Prepare

Directions: With all participants at your table:

1. Identify any customer needs that are not addressed with the proposed solution.

2. Identify the benefits of your solution.

3. Determine how you will respond to the customer's objections.

Prepare a 15-minute chalk-talk style presentation to the customer.

Step 3: Present the solution

Outcome

Present a solution to the target customer audience in a 15-minute chalk-talk format.

Timeframe: 30 minutes

Presentation

Directions:

1. Pair with another table.

2. One table is the Microsoft team and the other table is the customer.

3. The Microsoft team presents their proposed solution to the customer.

4. The customer makes one of the objections from the list of objections.

5. The Microsoft team responds to the objection.

6. The customer team gives feedback to the Microsoft team.

7. Tables switch roles and repeat Steps 2-6.

Wrap-up

Timeframe: 15 minutes

Directions: Tables reconvene with the larger group to hear the facilitator/SME share the preferred solution for the case study.

Description	Links
Azure Machine Learning documentation	https://docs.microsoft.com/en-us/azure/machine-learning/service/
Azure Machine Learning - How to enable App Insights	https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-enable-app-insights
Azure Machine Learning - How to enable data collection	https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-enable-data-collection

Machine Learning Pipelines	https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-create-your-first-pipeline?view=azure-devops
Azure Pipelines - Train and deploy machine learning models	https://docs.microsoft.com/en-us/azure/devops/pipelines/targets/azure-machine-learning?context=azure%2Fmachine-learning%2Fservice%2Fcontext%2Fml-context&view=azure-devops
Azure DevOps Release Approvals and Gates	https://docs.microsoft.com/en-us/azure/devops/pipelines/release/approvals/index?view=azure-devops
Enterprise security for Azure Machine Learning service	https://docs.microsoft.com/en-us/azure/machine-learning/service/concept-enterprise-security
PyTorch	https://pytorch.org
Keras	https://keras.io/
ML .NET	https://dotnet.microsoft.com/apps/machinelearning-ai/ml-dotnet
Azure Machine Learning - deploy trained models to IoT Edge	https://docs.microsoft.com/en-us/azure/iot-edge/tutorial-deploy-machine-learning
Microsoft's approach to AI	https://www.microsoft.com/en-us/ai/our-approach-to-ai