



Microsoft Cloud Workshop

Machine Learning
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.

© 2019 Microsoft Corporation. All rights reserved.

Microsoft and the trademarks listed at <https://www.microsoft.com/en-us/legal/intellectualproperty/Trademarks/Usage/General.aspx> are trademarks of the Microsoft group of companies. All other trademarks are property of their respective owners.

Contents

- [Machine Learning whiteboard design session student guide](#)
 - [Abstract and learning objectives](#)
 - [Step 1: Review the customer case study](#)
 - [Customer situation](#)
 - [Customer needs](#)
 - [Customer objections](#)
 - [Infographic for common scenarios](#)
 - [Step 2: Design a proof of concept solution](#)
 - [Step 3: Present the solution](#)
 - [Wrap-up](#)
 - [Additional references](#)

Machine Learning whiteboard design session student guide

Abstract and learning objectives

In this whiteboard design session, you will work with a group to design and implement a solution that combines Azure Databricks with Azure Machine Learning service to build, train and deploy the machine learning and deep learning models. You will learn how to use automated machine learning, model lifecycle management from training to deployment, in batch and real-time inferencing scenarios, and construct deep learning models for Natural Language Processing (NLP) in text classification and forecasting against time-series data. Finally, you'll learn to compare data with PyTorch and Keras for deep learning.

At the end of this workshop, you will have a deeper understanding of the capabilities and implementation solutions when leveraging the Azure Machine Learning service and Azure Databricks.

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.

According to Francine Fischer, CIO of Trey Research, Trey Research is looking to provide the next generation experience for connected car manufacturers by enabling them to utilize AI to decide when to pro-actively reach out to the customer thru alerts delivered directly to the car's in-dash information and entertainment head unit. For their proof of concept (PoC), they would like to focus on two maintenance related scenarios.

In the first scenario, Trey Research recently instituted new regulations defining what parts are compliant or out of compliance. Rather than rely on their technicians to assess compliance, they would like to automatically assess the compliance based on component notes already entered by authorized technicians. 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. They envision that with an approach like this to detect out of compliance parts, they can then process all of the components of each car (according to the specifications) to identify if any compliance issues should be flagged with the customer.

Each document in the component data has a short text description of the component as documented by an authorized technician. The contents include:

- Manufacture year of the component (e.g. 1985, 2010)
- Condition of the component (poor, fair, good, new)
- Materials used in the component (plastic, carbon fiber, steel, iron)

The compliance regulations dictate: *Any component manufactured before 1995 or in fair or poor condition or made with plastic or iron is out of compliance.*

For example:

- Manufactured in 1985 made of steel in fair condition -> **Non-compliant**
- Good condition carbon fiber component manufactured in 2010 -> **Compliant**
- Steel component manufactured in 1995 in fair condition -> **Non-Compliant**

The labels present in this data are 0 for compliant, 1 for non-compliant. Trey has already provisioned an Azure SQL Database containing their entire catalog of component descriptions, as well as labelled component data they expect to use in training the model.

In the second scenario, Trey Research would like to predict the likelihood of battery failure based on the telemetry stream of time series data that the car provides about how the battery performs when the car is started, how it is charging while running and how well it is holding its charge, among other factors. If they detect a battery failure is imminent within the next 30 days, they would like to send an alert.

With regards to the battery telemetry, the subject matter experts at Trey have explained that batteries are manufactured to a specification indicating how many cycles (that is complete charge, discharge cycles) they can handle before their performance starts to degrade (which indicates a good point to suggest replacing the battery). Effectively, the daily cycles consumed is like a clock counting up towards battery's rated lifecycle, approximating the battery's lifespan. In their solution, they can collect telemetry from a car on a daily basis, collecting data about the duration (in minutes) of the trips the car took during the previous day. For their prototype, they have provided historical data for batteries that have reached their replacement point. For a single battery, the telemetry looks similar to the following:

Date	Battery_ID	Battery_Age_Days	Number_Of_Trips	Daily_Trip_Duration	Daily_Cycles_Used	Lifetime_Cycles_Used	Battery_Rated_Cycles
1/1/2013	0	0	5	67.84560758	0.166920318	0.166920318	200
1/2/2013	0	1	5	53.45079793	0.131504817	0.298425135	200
1/3/2013	0	2	5	58.84143294	0.144767378	0.443192512	200
1/4/2013	0	3	5	60.63840306	0.149188457	0.59238097	200
1/5/2013	0	4	5	62.64690998	0.15412998	0.74651095	200

Upon detection of an out of compliance component or a battery at risk of failure, they would like to be able to send an alert directly to the customer inviting them to schedule a service appointment to replace the part.

In building this PoC, Trey Research wants to understand how they might use machine learning or deep learning in both scenarios, and standardize the platform that would support the data processing, model management and inferencing aspects of each.

They are also interested to learn what new capabilities Azure provides that might help them to document and explain the models that are created to non-data scientists or might accelerate their time to creating production ready, performant models.

Finally, they would like to be able to easily create dashboards that summarize the alerts generated so they can observe the solution in operation.

Customer needs

1. We would like to minimize the number of distinct technologies we need to apply across the data science process, from data collection, exploratory data analysis, data preparation, model training, model management and model deployment. How can Azure help us in this?

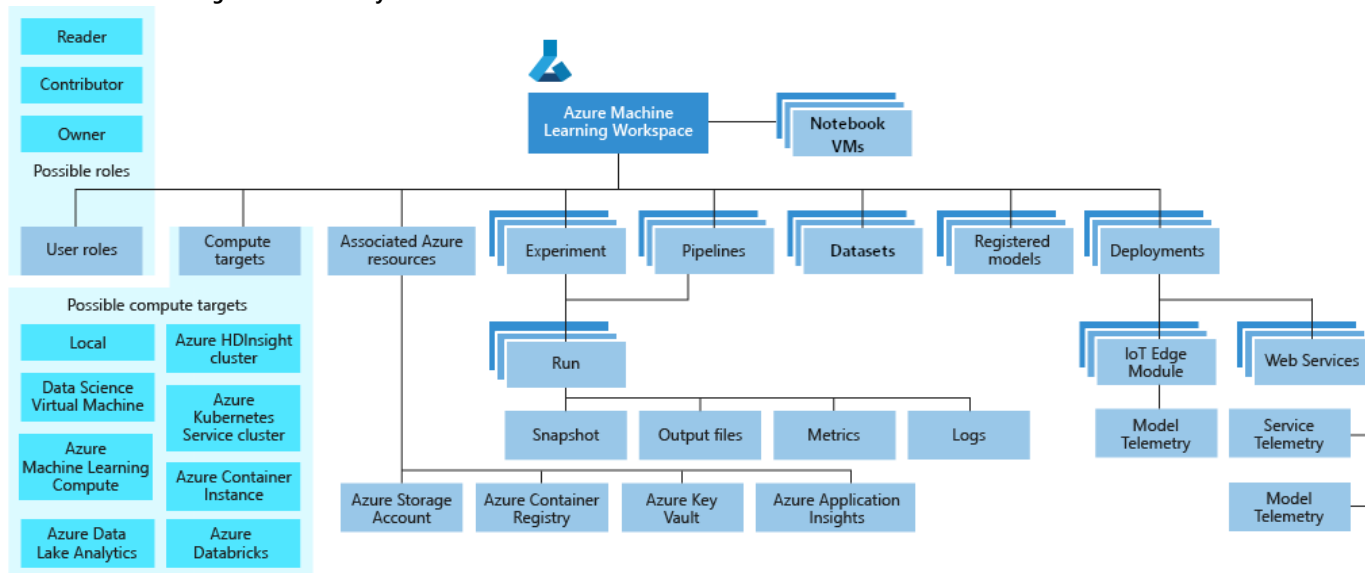
2. Our board is concerned about the liability of not being able to justify and explain the function of the models we create. We recognize many models are not easily explained in layman's terms, but how can we go about documenting how our models make predictions generally or how why they made specific predictions for particular input?
3. Understand how they should create the NLP based model for compliance and the battery life-span forecasting model.
4. Know the data pipeline they need to build in Azure, from ingesting telemetry, to storing both the compliance text and battery telemetry, to visualizing the result.

Customer objections

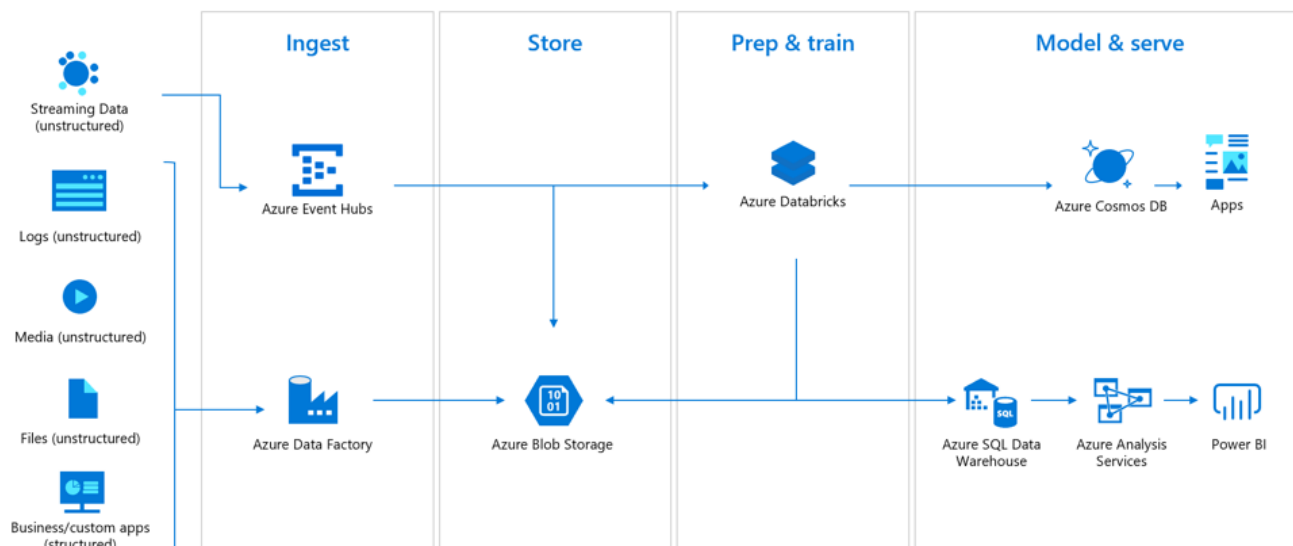
1. Should we use machine learning or deep learning approaches?
2. How should we choose between Keras and PyTorch for performing deep learning?
3. We have heard Azure Machine Learning service supports automated machine learning; can we use automated machine learning to create models using deep learning? Can we really expect a non-data scientist to create performant models using these tools?
4. Some of our team has worked with Azure Databricks, and they are confused by the overlap with Azure Machine Learning service. How should we be thinking about when to use which?
5. We are concerned about the capabilities around securing the solution. We are thinking mostly about security (including networking) and identity. How is Azure Machine Learning supporting enterprise security?
6. Are we fully aligned with the principles of responsible AI?

Infographic for common scenarios

Azure Machine Learning service taxonomy



Real-time analytics



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.

Classifying component descriptions text data

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. What data would they need to train the model?
3. How would Trey create a deep learning model to classify the component descriptions? Give an example of the architecture of the neural network.
4. Describe how Trey would use the model in the context of batch scoring the component text for compliance? What services and frameworks would you suggest?

Forecasting battery failure

1. At a high level, describe the steps to apply a forecasting model to predict battery failure pending within the next 30 days.
2. With regards to the model, what options does Trey have for creating the model against the time series data? Should they create a regression model, a forecasting model or a classifier? Why?
3. Can this model be built using machine learning or does it require deep learning?
4. If you were to suggest a deep learning model for forecasting against the time-series data, what architecture of neural network would you consider using first?
5. Describe how Trey would use the model in the context of scoring the streaming telemetry? What services and frameworks would you suggest?

Automated machine learning

1. For which scenario could Trey apply automated machine learning? Why?
2. How could they use automated machine learning with Azure Databricks?

Model Management

1. For both models, how should Trey track the performance of each of their training runs?
2. Can they quickly view this experiment data somewhere? Are there programmatic options to querying this data?
3. How would they manage versioning of each the models they have created and associate these models with the results of evaluating their performance?
4. In your solution, how will Trey retrieve previous versions of models and use them for further evaluation or scoring?

Model Explainability & Reproducibility

1. How would you suggest Trey programmatically create explanations of their models? What is the process? Explain how it works to improve the interpretability of the model generally as well as for the results of specific predictions.
2. Is the approach you suggest limited to working against machine learning models only (e.g., it does not work against deep learning models)?
3. Does your approach support explaining models created with automated machine learning?
4. How do machine learning pipelines help improve the reproducibility of model training and scoring? How could your solution leverage pipelines?

Enabling visualization

1. During model training and evaluation, what services and libraries would you recommend that Trey utilize for visualizing the data and performance results?
2. Would the approach you suggest extend to visualizing the streaming battery data?
3. What should Trey utilize for providing easy to customize visualizations to business analysts and stakeholders? Is it the same or different tool you suggested they use during modeling?

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.

Additional references

Description	Links
Azure Databricks documentation	https://docs.microsoft.com/en-us/azure/azure-databricks/
Azure Machine Learning service documentation	https://docs.microsoft.com/en-us/azure/machine-learning/service/
Microsoft Learn - Data Scientist	https://docs.microsoft.com/en-us/learn/browse/?roles=data-scientist
PyTorch	https://pytorch.org
Keras	https://keras.io/
Enterprise security for Azure Machine Learning service	https://docs.microsoft.com/en-us/azure/machine-learning/service/concept-enterprise-security
Microsoft's approach to AI	https://www.microsoft.com/en-us/ai/our-approach-to-ai