"AutoML" Module Transcript

Chapter 1

Intro, Topics Covered, & Learning Outcomes

Hi. My name is Josh Lemoine, and I'm a product manager here at One Model, specifically focused on One AI.

In the "What is One AI?" module, you learned that One AI machine learning is an automated machine learning, or AutoML, pipeline that automates the process for building models. Get ready to delve deeper into the fascinating world of AutoML and explore how it simplifies model creation in One AI.

In this module, we will cover the concept of AutoML, the importance and goals of AutoML, the key components and process of AutoML, and how One AI works as an AutoML system.

Upon completion of this module, you will gain an understanding of what AutoML is and its significance in simplifying the process of building machine learning models. You will recognize the importance of AutoML in democratizing machine learning and its goals of improving efficiency, scalability, consistency, and fostering innovation in model development.

You'll understand the typical steps involved in the AutoML process from data preprocessing and model training to model evaluation and deployment. And you will gain insight into how One AI is an AutoML system that provides the option of manual configuration.

Chapter 2

AutoML Overview

AutoML is the process of automating the time-consuming and iterative tasks of machine learning model development, simplifying the end-to-end process of building and applying models to real world problems. This includes tasks such as data preprocessing, feature engineering, model selection, and hyperparameter tuning.

AutoML empowers users, even those not possessing machine learning expertise, to efficiently and productively build machine learning models at scale while maintaining high model quality and performance.

Chapter 3

AutoML Importance & Goals

The goal of AutoML is to simplify the process of creating machine learning models for individuals who are not experts in the field by providing a straightforward user friendly interface for training and deploying models. This is important because machine learning has the potential to solve a wide range of problems, and bringing in people of all different backgrounds can accelerate progress. Beyond getting machine learning in more folks' hands, AutoML is important for several reasons.

The first is efficiency. Building and deploying models manually is time consuming and resource intensive. AutoML streamlines the process by automating many of the tedious tasks, such as data preprocessing, model selection, and hyperparameter tuning, saving time and effort. Scalability.

AutoML can scale to handle large and complex datasets, allowing organizations to analyze and derive insights from massive amounts of data more efficiently. If you work with large datasets manually, the benefit here is obvious. Consistency.

AutoML ensures consistency in model development and deployment, reducing the risk of human error and ensuring that best practices are followed consistently across different projects and teams. The best AutoML systems, One AI included, enable traceability of the tasks they perform so this consistency can be validated.

And finally, innovation.By automating repetitive tasks, AutoML frees up model builders to focus on more creative aspects aspects of model development, such as feature engineering or effectively modeling the data into storyboards so its insights can be easily understood by others.

Chapter 4

How AutoML Works

The AutoML process typically involves the following steps. The user provides the data that is used to train the model. This is the input to the AutoML system. The AutoML

system preprocesses the data. This involves cleaning and transforming the data into a format that is suitable for training the model. The AutoML system trains multiple machine learning models on the preprocessed data using a variety of algorithms and hyperparameters.

Hyperparameters are configuration settings. The AutoML system evaluates the performance of the various trained models and selects the one that performed the best.

This model is then used as the output of the AutoML system. The user can then use the trained model to make predictions on new data that the model hasn't seen yet and/or deploy the model for others to use.

Chapter 5

One AI as an AutoML System

The end-to-end machine learning process in One AI looks like this. The dataset used to train the model comes from your data in One Model. Data from multiple sources can be combined and used in the same model. Since One AI is part of One Model, the data does not leak the system. The training dataset along with the dataset on which predictions will be made are defined in One AI recipes using a simple question and answer interface. Recipes include context about common HR predictive scenarios.

Engineered features called generative attributes can also be created in the recipes flow. The AutoML system within One AI then preprocesses the data to transform it into a suitable format. This includes cleaning and transforming the data by means of scaling, one hot encoding, null and constant handling, data leakage detection, correlation reduction, and dimensionality reduction.

We will dig deeper into data preprocessing in a later module.

One AI then trains several ML models on the data and selects the one that performed the best in terms of performance and fit. The model is then run on the prediction dataset, which is structurally the same as the training dataset, but is related to the individuals or groups for which outcomes are unknown.

An example here is a list of currently active employees. Nobody knows yet whether they will leave the company in the next year. Once the model has been created and run, exploratory data analysis and model summary reports provide insight into the data used to train the model as well as model configuration and performance. Reviewing these

reports informs whether the data or model can benefit from fine-tuning before the predictions are deployed and shared via One Model storyboards.

Chapter 6

AutoML as Optional in One AI

If AutoML is so great, then why would anyone not want it? One AI includes functionality to mitigate many of AutoML's shortcomings, but there are still cases where bypassing some or all of the automation makes sense. On the topic of mitigating AutoML shortcomings, One AI is transparent.

Many AutoML systems are black boxes that hide settings, while One AI is not. Experienced machine learning experts may prefer manual configuration over AutoML for the flexibility and granularity of options that it provides.

If you are reproducing a locally created model that you're satisfied with, manual configuration will be necessary and sometimes fine tuning an AutoML generated model can improve results. This is an example of a combination of AutoML and manual configuration. AutoML is the default behavior in One AI, but it can be either incrementally or completely overwritten.

Chapter 7

Conclusion & Thanks

In conclusion, this module has provided an exploration of AutoML and its significance in streamlining the process of building and applying machine learning models. We've covered the fundamental concepts, goals, and key components of AutoML along with an understanding of how One AI serves as an AutoML system. By automating time consuming tasks like data preprocessing, model selection, and hyperparameter tuning, AutoML enhances efficiency, scalability, consistency, and fosters innovation in model development. With this knowledge, learners are empowered to leverage AutoML tools like One AI to democratize machine learning, making it more accessible and impactful across various domains. Happy modeling!