Session: Build, Deploy and Run a Model with AutoAI

Goals:

- Understand AutoAI and its key features
- Learn how to use AutoAI to
  - Train a model to predict which applicants qualify for mortgages
  - Deploy a model to a deployment space
  - Make a prediction request to see if an applicant is qualified for a mortgage

Required services:

- Watson Studio
- Machine Learning
- Cloud Object Storage

What is AutoAI

What is AutoML?

- Automated machine learning (AutoML) is **the process of automating the manual tasks** that data scientists must complete as they build and train machine learning models (ML models). These tasks include feature engineering and selection, choosing the type of machine learning algorithm; building an analytical model based on the algorithm; hyperparameter optimization, training the model on tested data sets and running the model to generate scores and findings. Researchers developed AutoML to help data scientists build predictive models without having deep ML model expertise. AutoML also frees data scientists from the rote tasks involved in building a machine learning pipeline, allowing them to focus on extracting the insights needed to solve important business problems.

What is AutoAI?

- AutoAI is a variation of AutoML. It **extends the automation of model building to the entire AI lifecycle**. Like AutoML, AutoAI applies intelligent automation to the steps of building predictive machine learning models. These steps include preparing data sets for training; identifying the best type of model for the given data, such as a classification or regression model; and choosing the columns of data that best support the problem the model is solving, known as feature selection. Automation then tests a variety of hyperparameter tuning options to reach the best result as it generates, and then **ranks model-candidate pipelines based on metrics such as accuracy and precision**. The best performing pipelines can be put into production to process new data and deliver predictions based on the model training.

Features of AutoAI

- Data pre-processing

  Apply various algorithms, or estimators, to analyze, clean and prepare raw data for machine learning. Automatically detect and categorize features based on data type, such as categorical or numerical. Use hyperparameter optimization to determine the best strategies for missing value imputation, feature encoding and feature scaling.

- Automated model selection

  Select models through candidate algorithm testing and ranking against small subsets of the data. Gradually increase the size of the subset for the most promising algorithms. Enable ranking of a large number of candidate algorithms for model selection with the best match for the data.

- Feature engineering

  Transform raw data into the combination of features that best represents the problem to achieve the most accurate prediction. Explore various feature construction choices in a structured, non-exhaustive manner, while progressively maximizing model accuracy using reinforcement learning.

- Hyperparameter optimization

  Refine and optimize model pipelines using model training and scoring typical in machine learning. Choose the best model to put into production based on performance.

Features of AutoAI

- Model monitoring integration

  Integrate monitoring on model drift, fairness and quality though model input and output details, training data and payload logging. Implement passive or active debiasing, while analyzing direct and indirect bias.

- Model validation support

  Extend with model and data insights and validate if your models meet your expected performance. Continuously improve your models by measuring model quality and comparing model performance.

IBM’s AutoAI

As part of the IBM Cloud Pak for Data end-to-end data and AI platform, IBM Watson Studio features the AutoAI toolkit that **automatically prepares data, applies machine learning algorithms and builds model pipelines** that are best suited for your data sets and predictive modeling use cases.

- The AutoAI toolkit supports data format in tabular CSV files, with comma (,) delimiter for all types of AutoAI experiments.
- When you load data to train an AutoAI experiment, you can load a single data file, or you can join multiple data files that share common keys into a single training data set. For data gathered over a specified date/time range (such as stock prices or temperatures), you can create a time series experiment to predict future activity.
- Using AutoAI, you can build and deploy a machine learning model with sophisticated training features and no coding. The process includes data pre-processing, automated model selection, automated feature engineering, hyperparameter optimization, generation and evaluation of candidate model pipelines, as illustrated in the image below.
- The AutoAI graphical tool displays the results as model candidate pipelines ranked on a leaderboard for you to choose to save and deploy to production.