Hands-On-Lab Script

Vertex AI AutoML for Tabular Data

UnternehmerTUM - onsite workshop - 1. cohort

Author: Gabriel Werner
Last Updated: Sep 25, 2024 | Status: Approved

Create a tabular dataset

1. In the Google Cloud console, in the Vertex AI section, go to the **Datasets** page.
   Go to the Datasets page
2. Click **Create** in the button bar to create a new dataset.
3. Enter `Structured_AutoML_Tutorial` (or any other name you like) for the dataset name and select the **Tabular** tab.
4. Select the **Regression/Classification** objective.
5. Click **Create** to create the dataset.
   For this tutorial, you'll use a publicly available bank dataset hosted on Cloud Storage.
6. For **Select a data source**, click **Select CSV files from Cloud Storage**
7. In **Import file path**, enter `cloud-ml-tables-data/bank-marketing.csv`
8. Click **Continue**.

Analyze the dataset

The analyze section lets you view more information about the dataset, like missing or NULL values.

Because our dataset is formatted correctly for this tutorial, you don't need to do anything on this page and can skip this section.

1. **Optional**. Click **Generate statistics** to view the number of missing or NULL values in the dataset. This can take 10 minutes or longer.

2. **Optional**. Click on one of the feature columns to learn more about the data values.

Train an AutoML classification model

1. Click **Train new model**.
2. Select **Other**.
3. In the **Training method** pane, confirm that the dataset you created previously is selected for the **Dataset** field.
4. For the **Objective** field, select **Classification**.
5. Confirm that the AutoML training method is selected.
6. Click **Continue**.
7. In the **Model details** pane, select **Deposit** for the target column and click **Continue**.
   The target column is what we're training the model to predict. For the `bank-marketing.csv` dataset, the `Deposit` column indicates whether the client purchased a term deposit (2 = yes, 1 = no).
   The **Training options** pane gives you an opportunity to add features and transform column data. If no columns are selected, then by default all non- target columns will be used as features for training. This dataset is ready to use, so there's no need to apply any transformations.
8. Click **Continue**.
9. In the **Compute and pricing** pane, enter 1 for the training budget.
   The training budget determines actual training time, but the time to complete training includes other activities, so the entire process can take longer than one hour. When the model finishes training, it is displayed in the model tab as a live link, with a green checkmark status icon.
10. Click **Start training**.