UDACITY

PROJECT SPECIFICATION

Optimizing an ML Pipeline in Azure

Documentation

CRITERIA	MEETS SPECIFICATIONS
Explain the pipeline architecture.	The README contains an explanation of: • The pipeline architecture, including data, hyperparameter tuning, and classification algorithm. • The benefits of the chosen parameter sampler. • The benefits of the chosen early stopping policy.
Compare a provided model with one generated by AutoML.	The README contains: • One or more sentences describing the model and parameters generated by AutoML. • Two or more sentences comparing the two models and their performance.

CRITERIA	MEETS SPECIFICATIONS
Explain and justify ways to improve models.	The README contains two or more sentences explaining potential improvements for a future experiment and why these improvements might improve the model.

Training Pipeline and AutoML

CRITERIA	MEETS SPECIFICATIONS
Use HyperDrive to automatically find optimal parameters.	A hyperdrive config is used and includes: • A parameter sampler • A policy for early stopping
Pass parameters to training scripts.	All specifiable parameters of the training script are specified in the hyperdrive config.
Retrieve the best run using `.get_best_run_by_primary_metric()`.	`.get_best_run_by_primary_metric()` is used on the hyperdrive run to retrieve the best run.
Use the *RunDetails* widget to explore run metrics.	The hyperdrive run is passed to the *RunDetails* widget.

CRITERIA	MEETS SPECIFICATIONS
Create an AutoMLConfig for training.	The solution notebook includes an AutoML config, which contains the following parameters: - `task` - `primary_metric` - `experiment_timeout_minutes` - `training_data` - `label_column_name` - `n_cross_validations`

Infrastructure

CRITERIA	MEETS SPECIFICATIONS
Create a compute cluster using the SDK.	A compute cluster is created using the Azure SDK and the `ComputeTarget` and `AmlCompute` objects.
Import data to a Dataset using the SDK.	A `TabularDatasetFactory` is used to create a dataset from the provided link.
Clean up deployed	The `delete` method of the `AmlCompute` object is used

CRITERIA	MEETS SPECIFICATIONS
resources.	to remove the cluster following training. **OR** An image of the compute cluster being selected for deletion is included in the README.

Suggestions to Make Your Project Stand Out!

1. Include a diagram of your pipeline architecture.
2. Export your model and run it in Cloud Shell.
3. Extend your AutoML config to include more parameters.
4. Have your code check for existing compute clusters before creating a new one.