

AWS Machine Learning Course – Data Setup

Chandra Lingam

Most of the data needed for the course is already included in the resources for the course.

There are two datasets that you would need to download directly from their respective owners.

1. Kaggle Bike Train Dataset

Location: <https://www.kaggle.com/c/bike-sharing-demand/data>

- a. Kaggle is a website that hosts competition and machine learning resources
- b. Registration is free and needed to download data
- c. From the above URL, please download train.csv, test.csv and sampleSubmission.csv files and save it to **C:\AWSMLCourse\Data\RegressionExamples\BikeTrain**

2. AWS Twitter Dataset

For this course, we need the following two files:

https://aml-sample-data.s3.amazonaws.com/social-media/aml_training_dataset.csv

https://aml-sample-data.s3.amazonaws.com/social-media/aml_training_dataset.csv.schema

Save the above two files to: **C:\AWSMLCourse\Data\ClassExamples\TwitterAWS**

Complete overview of the problem, solution developed by AWS is available in this location:

<https://github.com/awsmlabs/machine-learning-samples/tree/master/social-media>