Microsoft

Azure Machine Learning Free Lab

Contents

Overview 3

Terms of Use..... 5

Overview

Prerequisites

In this lab you can choose a “recipe” to further explore Azure Machine Learning. However, the recipes are light on detail to aid your exploration. Therefore, before embarking on this lab it is a good idea to have completed the following labs:

- Introduction to Azure Machine Learning
- Deploying a Predictive Model with Azure Machine Learning
- Text Analytics with R and Azure Machine Learning

Option 1 – Any Data!

Choose any dataset you like either:

- In the Samples provided by Azure Machine Learning or
- Upload a publically available dataset

and using Azure Machine Learning to model the data (including any cleaning/transformation)

Can you operationalize both the scoring and re-training of the model you create?

Option 2 – Restaurant Recommendation Engine

Using the following three Azure Machine Learning Sample datasets:

- Restaurant ratings
- Restaurant customer data
- Restaurant feature data

Build a recommendation system that is able to recommend restaurants to customers. You can use the Train Matchbox Recommender module to train a recommendation model based on the Matchbox recommender engine. The recommendation algorithm is based on the Matchbox model. To complete this lab, you will need to:

1. Select salient columns
2. Split data
3. Matchbox Recommender & scoring (see here for more details).
4. Evaluation

Option 3 – Twitter Sentiment Modelling

In this example, you want to find the sentiment (good/bad) of individual tweets. The dataset is available at this public blob store:

http://azuremlsampleexperiments.blob.core.windows.net/datasets/Sentiment140.tenPercent.sample.tweets.tsv

You can use the Feature Hashing module to transform a stream of English text into a set of features represented as integers. You can then pass this hashed feature set to a machine learning algorithm to train a text analysis model. The feature hashing functionality provided in this module is based on the Vowpal Wabbit framework. For more information, see here.

Guidelines:

1. You will need a reader module to read the above file held in blob storage.

2. An R script to remove punctuation, special characters, digits with space and convert to lower case. **Hint:**

```
# Separate the label and tweet text
sentiment_label <- dataset[[1]]
tweet_text      <- dataset[[2]]

# Replace punctuation, special characters and
digits with space
tweet_text <- gsub("[^a-z]", " ", tweet_text,
ignore.case = TRUE)

# Convert to lowercase
tweet_text <- sapply(tweet_text, tolower)

data.set <-
as.data.frame(cbind(sentiment_label,tweet_text)
,
stringsAsFactors=FALSE)
```

3. Feature Hashing – see here for more details on this technique.

4. Split the data

5. Filter-based filter selection (Chi-squared)

6. Two-Class binary classification

7. Compare scored dataset vs training dataset

Can you operationalize this model?

Terms of Use

By using this Hands-on Lab, you agree to the following terms:

The technology/functionality described in this Hands-on Lab is provided by Microsoft Corporation in a “sandbox” testing environment for purposes of obtaining your feedback and to provide you with a learning experience. You may only use the Hands-on Lab to evaluate such technology features and functionality and provide feedback to Microsoft. You may not use it for any other purpose. You may not modify copy, distribute, transmit, display, perform, reproduce, publish, license, create derivative works from, transfer, or sell this Hands-on Lab or any portion thereof.

COPYING OR REPRODUCTION OF THE HANDS-ON LAB (OR ANY PORTION OF IT) TO ANY OTHER SERVER OR LOCATION FOR FURTHER REPRODUCTION OR REDISTRIBUTION IS EXPRESSLY PROHIBITED.

THIS HANDS-ON LAB PROVIDES CERTAIN SOFTWARE TECHNOLOGY/PRODUCT FEATURES AND FUNCTIONALITY, INCLUDING POTENTIAL NEW FEATURES AND CONCEPTS, IN A SIMULATED ENVIRONMENT WITHOUT COMPLEX SET-UP OR INSTALLATION FOR THE PURPOSE DESCRIBED ABOVE. THE TECHNOLOGY/CONCEPTS REPRESENTED IN THIS HANDS-ON LAB MAY NOT REPRESENT FULL FEATURE FUNCTIONALITY AND MAY NOT WORK THE WAY A FINAL VERSION MAY WORK. WE ALSO MAY NOT RELEASE A FINAL VERSION OF SUCH FEATURES OR CONCEPTS. YOUR EXPERIENCE WITH USING SUCH FEATURES AND FUNCITONALITY IN A PHYSICAL ENVIRONMENT MAY ALSO BE DIFFERENT.

FEEDBACK. If you give feedback about the technology features, functionality and/or concepts described in this Hands-on Lab to Microsoft, you give to Microsoft, without charge, the right to use, share and commercialize your feedback in any way and for any purpose. You also give to third parties, without charge, any patent rights needed for their products, technologies and services to use or interface with any specific parts of a Microsoft software or service that includes the feedback. You will not give feedback that is subject to a license that requires Microsoft to license its software or documentation to third parties because we include your feedback in them. These rights survive this agreement.

MICROSOFT CORPORATION HEREBY DISCLAIMS ALL WARRANTIES AND CONDITIONS WITH REGARD TO THE HANDS-ON LAB, INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. MICROSOFT DOES NOT MAKE ANY ASSURANCES OR REPRESENTATIONS WITH REGARD TO THE ACCURACY OF THE RESULTS, OUTPUT THAT DERIVES FROM USE OF THE VIRTUAL LAB, OR SUITABILITY OF THE INFORMATION CONTAINED IN THE VIRTUAL LAB FOR ANY PURPOSE.

DISCLAIMER

This lab contains only a portion of the features and enhancements in Microsoft Azure. Some of the features might change in future releases of the product.