Parker J. May

February 10, 2018

McKesson | Final Project | Deep Azure

Topic: Azure Machine Learning

Problem Statement:

There are few problems facing modern society as pressing as predicting the quality of a bottle of wine. With the rising popularity of wine, prices rise as well. Finding a good bottle at a good price seems more like luck than anything else. This project will combine the Azure Machine Learning service and wine rating data to create a model for predicting wine quality based on several factors. It will use an Azure SQL Database to store the data and Azure API Management Services to make a REST API available to a browser based client.

High Level Overview of Steps:

1. Download dataset from Kaggle
2. Clean data
3. Create Azure SQL Database and populate with data
4. Create experiment in Azure Machine Learning Studio
5. Score and evaluate model
6. Publish model as web service
7. Create Azure API Management service
8. Implement browser based client using the Angular JavaScript framework.

Data Set:

https://www.kaggle.com/zynicide/wine-reviews

Before data cleansing: 48.6 MB, 150,936 rows | After data cleansing: 13.2 MB, 149,145 rows | Format: csv

Hardware:

Lenovo laptop running Windows 10 and Dell laptop running Windows 7

Software:

MS SQL Utilties bcp, sqlcmd and ODBC driver; Cygwin linux shell; node/npm; Angular JavaScript library.

Lessons Learned:

With the easy access to large data sets and the power of the built-in machine learning algorithms in Microsoft Azure, it's tempting to think it's trivial to throw a bunch of data at the tools and expect good information to fall out. This experience has shown me that it takes a lot more to achieve meaningful goals. Not that this is a bad thing, everything you need is out there, but learning to apply it is an odyssey one must be prepared to see through.

URLs:

- Two minute (short): https://youtu.be/Fi_lH-p_n4M
- 15 minutes (long): https://youtu.be/Y4gIIUNv0a0
- GitHub: https://github.com/parkermay/final-project