**Omaha Data Science Academy (OSDA) – Python and Advanced Modeling**

**Cohort 6 (Jun. 10 – Jul. 3, 2019)**

Instructor(s):  Jeremy Bergmann, BS/MS

Phone: 402-321-7530; Email:  [jeremy@jbanalyticsconsulting.com](mailto:jeremy@jbanalyticsconsulting.com); Office: By Appointment

Github:  <https://github.com/jbergmann56/ODSA-PythonAdvModels>

**COURSE DESCRIPTION:**

Data Science has gone from “needing to know how to code” to modeling techniques that have standardized libraries that can be pasted into a program. This means that one may do data science without understanding what the models mean or do. This class will drill into how to program the models the “traditional way”, by understanding the methods and programming python code to accomplish these tasks.

We will use various Python libraries to scrape, debug, and enhance data science models - leading to the creation of a Natural Language Processing (NLP) model using Word2vec. We will also show how to use Python to solve other gaps such as calculations, other data manipulation, and random number population.

**Overall Approach:**

The overall approach to the course is to provide the background needed for advanced modeling within python, including python coding basics, reviewing data science-related python libraries, and utilizing tools (Dataiku) to assist and automate the build and deployment of data-science models.

**Class Schedule:**

**Class 1 - Applying Model Building to Python Programming**

Hour 1 - Introductions, Class Overview, Data Science Philosophy

Hour 2 - Tool/Environment Setup (Sublime Text, Jupyter Notebook, Git Desktop, CLI & Repo),

            Python Basics + Working Examples

Hour 3 - Library review: Built-In/Standard, Data Science: NumPy, Pandas & MatPlotlab

**Class 2 - Python code for Data Science Projects**

Hour 1 - Class Project Overview, Pandas Library, Random Number Population (Benford’s Law)

Hour 2 - HTTP Web Scraping, Rest API Requests, JSON Parsing & Creation

Hour 3 - Obtain Data Sets - Text Analytics (Excel), Stock Market Data, Social Media/Screen Scraping

**Class 3 - Model Building Process (CRISP-DM) & Scikit-Learn Library**

Hour 1 - Week 1 Quiz, CRISP-DM Methodology

Hour 2 - SciKit-Learn - Overview

Hour 3 - Exercises - Data Cleaning & Prep, Data Imputation &  Model Building using Scikit-Learn

**Class 4 - Text Mining & NLP Introduction using Word2Vec**

Hour 1 - Text Mining & NLP - Lecture

Hour 2 - Word2Vec Intro - Library, Data + Working Examples

Hour 3 - Data Prep for Initial Model Build (Class 3), Exercises/Working Session

**Class 5 - Word2Vec Model Building**

Hour 1 - Week 2 Quiz, Initial Model Build - Python Working Example

Hour 2 - Model Evaluation/Modeling Approach

Hour3 -  Exercises/Working Session

**Class 6 - Model Building for Deployment**

Hour 1 - Model Deployment  & Change Management - Overview

Hour 2 - Python Code Refactoring - Generalize Word2Vec solution using Scikit-Learn

Hour 3 - Exercises/Working Session - Translation to Dataiku Python Code

**Class 7 - Data Engineering using Python & Dataiku**

Hour 1 - Week 3 Quiz, Data Engineering vs. Data Science

Hour 2 - Dataiku Overview, Workflow & Understanding Results

Hour 3 - Exercises/Working Session - Implementing/Operationalizing Results

**Class 8 - Class Project - NLP**

Hour 1 - Creating a Data Pipeline - Python

Hour 2 - Model Evaluation & Monitoring

Hour 3 - Data Product Creation - Python + Google Fusion/GCP NLP Application