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

**Cohort 5 (Jan. 7– Feb. 4)**

Instructor(s): Jeremy Bergmann, BS/MS (Class 1-5), Gordon Summers (Class 6-8)

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

**COURSE DESCRIPTION:**

Data Science has gone from “needing to know how to code” to most modeling techniques having 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 Word2Vec to scrape, debug, and enhance data science models. 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 automat 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 + Packages)

Python Basics + Working Examples

Library review: Built-In/Standard, API/HTTP Related, NumPy, Pandas+MatPlotlab, Word2Vec

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

Working Session - Examples & Libraries

**Class 2: Word2Vec**

Hour 1 - Text Mining - Lecture

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

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

**Class 3 - Word2Vec Model Building**

Hour 1 - Initial Model Build - Python Working Example

Hour 2 - Model Evaluation/Modeling Approach

Hour3 - Exercises/Working Session

**Class 4 - Model Building on Different Sources**

Hour 1 - Python Code Refactoring - Generalize Word2Vec solution to different data sources

Hour 2/3 - Exercises/Working Session

**Class 5 - Expanding the Scikit-Learn Library**

Hour 1 - SciKit Overview, Library + Exercises

Hour 2 - Dataiku Review, Tool/Environment Setup (Juypter Notebook, Git Desktop, CLI & Repo)

Hour 3 - SciKit Model Building, Word2Vec Additions, Translation to Dataiku Python Code

**Class 6 - Add Python to Dataiku flow**

Hour1 - Dataiku Overview, working example

Hour2 - Review, Understand Results

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

**Class 7 - Python code to solve other gaps**

Calculations

**Class 8 - Other Data Manipulation**

Random Number Population (Benford’s Law)