Personal Background, Current Technology Trends, Interests

Indiana University Bloomington: Big Data Applications

Naimesh Chaudhari: naichaud@iu.edu

Assignment 1 – September 12, 2019

1. Abstract

In this paper, I provide a brief introduction about myself and the data science concepts I have utilized. Additionally, I give examples of significant projects that I have worked on in my current role as a Data Scientist with Uline. I also discuss some of the top technology trends that I am interested in learning more about through this class. The trends that I discuss primarily focus on deep learning & AI. My goal is to give a brief overview of deep learning & AI and then discuss the PaaS services offered by one of the most prominent stakeholders in the cloud computing stage, Microsoft.

2. Introduction

My interest in the data science field started in 2017. Up until then, I was exploring concepts in SQL & VBA. During my time at Sears Holdings, I developed countless reports and optimized many processes using VBA and SQL. Once I was comfortable with both languages, my interests changed to understanding advances analytics. I joined Uline in 2017 and started exploring options to pursue a master's in data science. I joined the IU master's program in 2018 and am planning on earning my degree in December 2019.

I am familiar with the two major programming languages used within the data science field, Python & R. I prefer Python for almost every task other than graphing. I believe the ggplot library in R is much better than the ones I have encountered in Python.

My focus during the master's program has primarily been on understanding machine learning concept and algorithms. I am very familiar with the scikit learn library in Python and the numerous data processing and modeling objects within the library.

In my current role with Uline, I have completed many projects through the use of machine learning that have impacted the bottom line directly. Some of my accomplishments are customer comments classification/sentiment analysis, competitor analysis, & the development of a Python/Machine Learning course for interested parties within Uline.

3. Deep Learning

Although I have in-depth knowledge of many machine learning algorithms, deep learning and cloud computing are fields I have never explored. To close this gap, I have done two things. I have taken a course in deep learning for this semester and I am setting my primary focus in this class to understanding deep learning /cloud computing concepts and applying them in my current work environment.

Wu states, "*Deep Learning is the next generation of machine learning algorithms that use multiple layers to progressively extract higher level features (or understanding) from raw input*" (Wu, Jun).

This, to me, speaks volume, my understanding is that in the next 5-10 years, deep learning shall be the primary ML process for many corporations. Understanding and applying these techniques within the field can be an invaluable skill set.

Most of the AI systems in today's world incorporate some deep learning process within them. Seven key examples are self-driving cars, text classification, text translations, language recognition, computer vision, and self-learning robots. There are three deep learning branches I am interested in: Convolutional Neural Networks, Artificial Neural Networks, and Feed Forward Neural Networks.

One of the significant advantages data scientists have with deep learning libraries is that they can utilize GPU's to do their computing tasks. This allows for much faster processing. GPU computing is also a disadvantage for corporations like Uline, whose local employee computer systems do not have a strong GPU attached to them. This is where the benefits of cloud computing can be realized.

4. Microsoft Azure

Microsoft Azure offers many services to its clients through the cloud. Some examples are: "Deploy Infostructure", "Secure and Manage Resources", "Develop Apps", and "Manage Data & AI". I focus on the PaaS branch "Manage Data & AI", more specifically, the "AI & Cognitive Services" area.

Microsoft states, "*Azure Machine Learning offers web interfaces & SDKs so you can quickly train and deploy your machine learning models and pipelines at scale. Use these capabilities with open-source Python frameworks, such as PyTorch, TensorFlow, and scikit-learn.* " (Microsoft)

Microsoft offers a full-stack service for Data Scientist to build and deploy models quickly. User can have online storage, and cloud computers, that can scale at demand. This allows employees to train complex deep learning models that require very costly hardware at a marginal cost. They work with all the major deep learning frameworks that are out there. Also, not only are you able to train models quickly but also deploy onto the cloud. This makes the integration of different technologies within a corporation extremely easy. For example, a Data Scientist can train a model using Python. A Data Engineer can work on deploying the model to the cloud. Finally, a Software Engineer can make API calls to the deployed model that return JSON data and use that information in a production Java application. This allows for rapid deployment into production.

Slowly as Uline advance into the cloud computing stage, Azure shall play a big role for us. Most of the technologies we use are through Microsoft and having this sort of integration for the Analytics team shall be priceless. My goal is to understand the full azure stack. In the future, this shall allow me to make recommendations and be part of a team that can lead Uline's transformation into the cloud!

5. References

1. Wu, Jun. "AI, Machine Learning, Deep Learning Explained Simply." Medium, Towards Data Science, 12 July 2019, https://towardsdatascience.com/ai-machine-learning-deep-learning-explained-simply-7b553da5b960.

2. Squillace. "Microsoft Azure Documentation." Microsoft Docs, https://docs.microsoft.com/en-us/azure/.