Metis Senior Data Scientist Passion Projects

Metis Senior Data Scientists have the opportunity to pursue a passion project of their design once every three quarters. Below are four stellar examples of what's been accomplished in the recent past.

ROBERTO REIF

Previous Employer:
Sensoria Inc.

Education:
Boston University
(Ph.D., Biomedical Engineering)

BUILDING AN IMAGE-BASED PRODUCT RECOMMENDATION SYSTEM

Roberto spent his passion quarter with more than 25,000 unique color images of shoes, using them to create a recommendation system that helps users find the right style and fit. All images were on a white background, presented in side view, and were pre-processed to fit into the pre-trained image classification Convolutional Neural Networks (CNN) architecture. Specifically, he used the VGG16 architecture pre-trained with the ImageNet dataset.

As a result of this work, he presented at the Global Artificial Intelligence Conference and hosted a well-attended Meetup event in Seattle to share findings with the local data science community. You can read about the project in further detail on **his blog**.

ENGAGING LEARNING EXPERIENCES WITH NLP & STORYTELLING

Alice used her passion quarter as an opportunity to teach. During the three month window, she gave a number of informative, instructional talks and tutorials tied to natural language processing.

1) She developed a Natural Language Processing tutorial based on analyzed transcripts of the most popular stand up comedy specials from the past 5 years, according to IMDB. Her goal was to nail down which topics and types of language translate to popularity while incorporating sentiment analysis, topic modeling, and text generation to showcase prominent NLP techniques. She found that the majority ended up falling into two main categories: family (parents, spouses, etc.) and profanity (bad words, guns, etc.).

As a result of the project, she published the tutorial on **Github**, as well as presented it at **PyOhio Conference** and the ODSC West Conference.

2) In an effort to deepen her own understanding of the basic architecture of deep learning, how it relates to NLP, and how it can be explained in a visual way to a beginner-level audience, Alice developed and gave talks on Recent Advancements in NLP at Big Data Innovation Summit and Predictive Analytics Summit.

ALICE ZHAO

Previous Employers:
Cars.com, Redfin

Education:
Northwestern University
(M.S. Analytics, B.S. Electrical Engineering)

CHAD SCHERRER

Previous Employer:
Pacific Northwest National Laboratory, Galois Inc.

Education:
Indiana University
(Ph.D. Mathematics)

EXPLORING JULIA'S SOSS FOR PROBABILISTIC PROGRAMMING

Much of data science involves reasoning about complex probability distributions. Just ten years ago, this process was slow and labor-intensive, mostly limiting benefits to experienced researchers. The advent of probabilistic programming has changed this dramatically. What once required advanced understanding of calculus and algorithm development can now be done automatically by a compiler, making powerful custom modeling approaches available to a mainstream data science community.

During his passion quarter, Chad continued his development of Soss, a new approach to probabilistic programming that represents a "model" as unexecuted Julia code. This greatly simplifies things "under the hood" while allowing flexible modeling and a potential for arbitrarily high-performance execution. Check out the project on **Github** and read about it in detail on Scherrer's **blog**. He also presented on the topic at **PyData Miami**.

MACHINE LEARNING FROM SCRATCH IN PYTHON

Zach wanted to both better understand machine learning algorithms and learn how to write them in ways that were pedagogically enlightening. He used his passion quarter to tackle this dual goal. As an example, on **his Github**, he rewrote many of the algorithms in Scikit-Learn, the leading Machine Learning module for Python, to make the abstract code easier to understand. Because his modules are much simpler in design, a student can read through and understand how the algorithm works.

Zach has continued his work beyond his passion quarter and has now written more than 50 unique machine learning tools and algorithms from scratch. These include gradient descent and regression techniques, tree-based techniques, neural nets, and even statistical tools like population estimation through kernel density estimation.

ZACH MILLER

Previous Employers:
Metis (Currently at CreditNinja)

Education:
University of Kentucky
(Ph.D., Nuclear Physics)