Michael Knight

Data Scientist | Machine Learning Engineer | Data Analyst

Philadelphia, PA - OPEN TO RELOCATION | (202) 747-4509 | mknight4714@gmail.com | https://www.themichaelknight.com/ | https://www.linkedin.com/in/michael-knight-910b6b83/ | Github: @MichaelKnight4714

Leveraging Machine Learning, Data Science, and Advanced Mathematical practices to solve complex puzzles.

SKILLS

Languages: Python, R, Java, JavaScript, SQL, HTML 5, CSS, Scala
Libraries/Packages/Frameworks: Pandas, Numpy, SciPy, SciKit-Learn, PyTorch, Matplotlib, Seaborn, Keras, Tensorflow, NLTK, Gensim, EDL
Databases: AWS (EC2, S3, Lambda, Redshift, Sagemaker), DBeaver, MySQL, PostgreSQL

IDEs: Jupyter (Lab, Notebook), Visual Studio Code, RStudio, NetBeans
Statistics: Bayesian Statistics, Regression/Classification Modeling, Predictive Analytics
ML: Neural Networks, NLP, OCR, Unsupervised Learning, Deep Learning, Algorithm Scripting, Data Wrangling, Data Mining

Data Visualization: Tableau, PowerPoint, Excel, Google Slides, Google Docs, LaTeX
Version control: Git, Github, Gitlab, Bash
Operating Systems: Windows, Linux, OS
Methodologies: Agile, Scrum, Scrum of Scrums, JIRA, Confluence, Kanban boards
AI: OpenAI API, LLM, Llama2, GPT-4-turbo, Recommendation Systems, GAN

EXPERIENCE

Vidoori, Data Scientist/Machine Learning, Hyattsville, MD, February 2024 - September 2024

- Designed and built an AI driven resume grader (written in Python and deployed in AWS Lambda) that evaluates (on a scale of 0-5) candidates based on 2 pass-fail disqualifying criteria, 3 weighted scoring criteria, and 3 bonus criteria; with 90% similarity to human grading
- Designed 2 chatbots (HR Bot, Legal Bot) with distinct voices and exclusive access to their designated data using Llama2 via Llamaindex
- Developed cross-functional PowerPoint presentations to educate coworkers on the company's SOTA Transformer based Deep Neural Net used to link people across two different surveys for the Census Bureau, as well as other AI / ML /NLP models and techniques
- Precisely tracked and maintained project timelines cross functionally in JIRA tickets through Agile methodology and Scrum practices

Cherry Street Energy, Data Scientist (Contractor), Atlanta, GA (Remote), July 2023 - February 2024

- Within a two week deadline, created linear, ridge, and LASSO regression models that predict the Monthly Energy Usage (kWh) and Intensity (kWh/mo.) of a building in the U.S Southeastern Region within 90% accuracy, based on six inputs (square footage, stories, building profile, year constructed, weekly operating hours, and month)
- Optimized the best performing model (ridge) using GridSearchCV to tune hyperparameters
- Incorporated the regression model into a SEED calculator, which, in tandem with a function that calculates Billing Demand, determines how much a client would save on energy cost by switching to Cherry Street Energy over the next 20+ years
- Developed and designed data pipelines to support an end-to-end solution for accessing Georgia Power's API to extract meaningful insights on commercial building data for the SEED calculator

CareForge AI, Data Scientist, McLean, VA, August 2023 - October 2023

- Built and maintained a secure and organized data repository using PostGreSQL 16, ensuring that data integrity and accessibility are maintained throughout
- Trained and optimized machine learning models enriched with key terms and search strings tailored to the platform's specific needs
- Worked with the advisory boards and used NLP techniques to extract insights, automate processes, and enhance user interactions on the platform, ensuring the most relevant and accurate responses to user input
- Set up local large language models (LLMs) development pipeline

American University, Graduate Research Assistant (Machine Learning Engineer), Washington, DC, September 2021 - July 2023

- Developed human assisted machine learning and natural language processing (NLP) approaches to infer information about chemical compounds from highly technical open literature sources
- Enhanced existing machine learning techniques to predict 9 different properties of CNOHF chemical molecules from their molecular structures for 439 unique chemical compounds
- Designed and implemented nested K-fold cross validation on a Kernel ridge regression (KRR) model using radial basis function (RBF) kernel mapping to find parameters that would give the best possible Mean Absolute Error (MAE) score for the model when using the 439 vectors of the 28 dimensional Sum Over Bonds featurization method as the feature matrix (X) and the 9 chemical properties as the target vectors (y_1-y_9)

- Designed and implemented convolutional neural networks within PyTorch and PyTorch_Geometric to create neural fingerprints for these compounds based on their graphical representations (as generated using RDKit)

Hunger Free America, Data Analyst, New York, NY, February 2020 - September 2021
- Analyzed meal site data using Python to observe the biggest increase and decrease in numbers of lunches and breakfasts served between 3 years, by all 5 boroughs, and compared to census tract by poverty rate, to find underserved neighborhoods
- Mapped analytic discoveries for 1,312 meal sites over the span of 3 years using Tableau
- Created complex, cross-object and cross-platform reports and dashboards in Tableau
- Designed and implemented Python modules to conduct ETL and EDA processes on 3 sets of yearly data using NumPy and Pandas

General Assembly, Data Science Fellow, Washington, DC, June - August 2019
Applied data analysis and data visualization skills in a 500-hour immersive course:
- Completed 35 labs and 5 projects, leveraging data science best practices to solve real-world problems, using programming (Python {NumPy, SciPy, Pandas, SciKit-Learn}, SQL), data visualization (Matplotlib, Tableau), and version control (Git)
- Developed a Natural Language Processing model to recommend 10 cheeses (given an input cheese name and/or up to 13 cheese features), by binarizing a dataset of 1,827 cheeses scraped off of cheese.com and calculating the cosine similarities within this binary matrix using the pairwise distance function, and made a user friendly interface using ipywidgets
- Developed a Natural Language Processing model to predict which of two subreddits a post originated from, using statistical methods and modeling, including data collection (via Reddit's API), TFIDF Vectorization, and GridSearch
- Modeled large and disparate data sets using a variety of machine learning techniques

Weber, Gallagher, Simpson, Stapleton, Fires & Newby, LLP, File Clerk, Philadelphia, PA, November 2018 - February 2019
- Organized and maintained legal documents and case archives for over 300 different workers' compensation cases
- Cleaned up and updated over 150 files for cases that had been in complete disarray for upwards of 15 years
- Scanned 130 files (which were comprised of usually over 1000 documents of 26 different filing classifications) into Aderant Total Office legal software and shredded physical files, freeing up significant space

Kane, Pugh, Knoell Troy & Kramer, LLP, File Clerk, Norristown, PA, February 2015 - November 2018
- Organized and maintained legal documents and case archives for 3 different partners with 3 unique filing styles, and created 15-page documents outlining specifically how to file for each partner as an aid for other clerks and paralegals
- Created an organization system and protocol for maintenance, to support 6 attorneys and 7 paralegals

Catering by Design, Prep Cook/Caterer, Philadelphia, PA, June 2013 - February 2015
- Kept the walk-in refrigerator and walk-in freezer organized by food type and date for over 300 dishes and ingredients
- Prepared and cooked multiple dishes in a short time frame for high profile clientele
- Plated and served appetizers, entrees, and desserts for various events

Nueva Esperanza Academy Public Charter High School, In-School Math Tutor, Philadelphia, PA, September 2011 - June 2013
- Tutored 30 students in 3 different fields of math, individually and in small groups, during and after school hours
- Composed graphic organizers to assist learning
- Collaborated with the math teachers in lesson planning and assessment

EDUCATION

- **American University, Master of Science in Data Science**, Washington, DC, 2021-2022
- **General Assembly, Data Science Fellowship**, Washington, DC, 2019
- **Community College of Philadelphia**, Philadelphia, PA, 2016 - 2017 additional CS courses (Java)
- **University of Maryland**, College Park, MD, 2009 - 2010 additional Math and CS courses (JavaScript, Advanced Calc, Linear Alg)
- **Bard College, Bachelor of Arts in Mathematics**, Annandale-on-Hudson NY, 2001-2006

VOLUNTEER WORK

Books Through Bars, Letter Logger/Packager, Philadelphia, PA, June 2012 - February 2015
- Received, processed and cataloged educational print donations and individual materials requests
- Wrapping up and helping mail out packages of books to the prisons that allowed such programs