Gerald K. White, M.S.

https://geraldkwhite.com | (732) 476-8915 | geraldkwhite10@gmail.com

Results-driven and highly adaptable data scientist with a strong background in applied statistics and machine learning and a passion for leveraging data-driven insights to drive product development and strategy. As a recent master's graduate with experience on cross-functional project teams at The Dow Chemical Company, I am now looking to apply demonstrated analytical skills, business acumen, and product intuition in my next role.

PROFESSIONAL EXPERIENCE

Research Assistant – Toyota Technical Institute of Chicago Chicago, IL   **02/2023 – Present**
- o Researching current Reinforcement Learning with Human Feedback (technology behind ChatGPT) algorithms.
- o Identified potential improvement to RLHF, designed alternative solution, and created experimentation code by modifying AllenAI's RL4LMs repository. Active codebase can be found here.
- o Produced metrics and visualizations to track project progress in weekly meetings.
- o Communicated insights and progress to both technical and non-technical project stakeholders.

Data Scientist Intern – Lubrizol Chicago, IL   **06/2022 – 10/2022**
- o Improved multiple internal models by conducting exploratory data analysis, proposing alternative solutions, creating metrics for success, and developing production-grade modeling code in both Python and R.
- o Implemented stacked random forest classification model for large material performance dataset, resulting in 60% accuracy increase and heightened model value.
- o Augmented existing time-series chemical concentration model by introducing Bayesian feature selection and ARIMA, resulting in 44% higher predictive accuracy and better model reliability.
- o Identified potential for and subsequently engineered data pipeline in SQL to join two internal datasets and produce highly utilized material performance model.
- o Provided consultation and drafted code in PyTorch for new deep learning representation model.
- o Created R-Shiny app and Power-BI dashboard to present insights to project stakeholders.

Process Automation Engineer – The Dow Chemical Company St. Charles, LA   **07/2019 – 07/2021**
- o Participated in cross-functional project teams which often required presenting complicated and technical concepts to stakeholders from both technical and non-technical backgrounds.
- o Managed plant's control code repository, involving interfacing with project teams, managing contributions to control system, and creating procedures to standardize work processes.
- o Developed production-grade code to control chemical production assets at one of the company's largest sites.

EDUCATION

M.S. in Statistics – University of Chicago   **09/2021 – 06/2023**
  Coursework in statistics, computer science, machine learning, and deep learning.

B.S. in Chemical Engineering – Villanova University   **09/2015 – 05/2019**
  Minor in Mathematics, coursework in chemical engineering, upper-level mathematics, and computer science.

RELEVANT COURSEWORK

Statistics & Data Analysis:
- o Advanced Statistical Theory I & II:
  - o Topics: Distribution theory, experiment design, ANOVA, A/B testing, Bayesian methods, time-series analysis.
- o Applied Linear Models, Generalized Linear Models, Modern Methods in Applied Statistics:
  - o Topics: Logistic Regression, GLMs, ARIMA, MCMC methods, applied Bayesian inference, clustering.

Machine Learning & Deep Learning:
- o Intro to ML, Foundations of Deep Learning, Discriminative Deep Models, Generative Deep Models
  - o Topics: Deep networks, clustering, classification, computer vision, LLMs.
- o Computational Aspects of Large Language Models
  - o Topics: Querying LLMs (Hugging Face, GitHub, etc.), prompt engineering, reinforcement learning with LLMs

RESEARCH PROJECTS

Reinforcement Learning with Human Feedback as a Tool for Knowledge Distillation   **02/2023 – Present**
- o Researching novel Reinforcement Learning (RL) for Large Language Models (LLMs) algorithm to improve model alignment and understand recent success of RLHF.

Exploring Denoising Autoencoder Architectures in Self-Supervised Learning (MS Thesis)   **08/2022 – 02/2023**
- o Researched and developed experimentation code for novel self-supervised learning mechanism which aims to reduce labeled data dependency and improve model transfer ability. Project repository here.

LANGUAGES & SKILLS:

Python (advanced) | R (advanced) | SQL (advanced) | Bash (intermediate) | Java (beginner) | Django (beginner)

Machine Learning | Data Analysis & Visualization | Cross-functional Communication | Product Sense