Lily Northcutt, M.P.S.

+1 (240) 671-3796 | northcuttlilya@gmail.com | Portfolio

Results-driven, master's level data scientist with expertise in machine learning (ML), artificial intelligence (AI), cloud computing, and descriptive visualization to derive actionable insights. Seeking to apply my analytical and problem-solving skills within a dynamic and collaborative team to tackle complex challenges and optimize results.

SKILLS

Language and Tools: Python, R, SQL, Git, Jira
Machine Learning: Deep Learning, Natural Language Processing, Time Series Analysis, Regression Models, Statistical Analysis
Data Visualization: Shiny, Tableau, Quarto
Data Management: ETL Pipelines, Data Cleaning, Cloud Computing and Big Data Systems (AWS, Spark, Hadoop)

EDUCATION: *University of Maryland, College Park*

Master of Professional Studies, Machine Learning August 2020 – May 2022
Relevant Coursework: Principals of Machine Learning, Natural Language Processing, Big Data Systems

Bachelor of Science, Mathematics August 2017 – May 2020
Relevant Coursework: Mathematical Modeling, Theory and Methods in Statistics, Computational Methods

WORK EXPERIENCE

Data Scientist **Las Cruces, New Mexico**
New Mexico State University – Plant Science Dept. January 2024 – Present

- Analyzed pepper speed breeding experiments, demonstrating potential for 12% acceleration in breeding time, with results presented at the American Society of Horticulture conference (Fall 2024).
- Created a pipeline using Nextflow and AWS to process millions of genetic sequences for efficient clustering and classification.
- Designed an interactive Quarto document incorporating Shiny to showcase the code, analysis, explanations, and interactive graphs, allowing users to explore the results dynamically.

Operations Research Analyst **Patuxent River, Maryland**
Naval Air Systems Command – Procurement Group February 2021 – June 2023

- Supported aircraft contract negotiations through ETL processes, regression analysis, and the application of ML algorithms such as LSTM on large amounts of historical aircraft cost and labor pricing data, reducing project costs by 10%.
- Transformed unstructured data into clean and usable formats, improving the efficiency of the data cleaning processes.
- Developed an R Shiny application to present interactive graphics, train ML models, and visualize outputs for specialists without technical programming knowledge.
- Collaborated in cross-functional environment with developers, negotiation specialists, to defense contractors, recognized with two awards for excellence in performance and contributions.

Research Fellow **Gaithersburg, Maryland**
National Institute of Standards and Technology (NIST) – Mathematics Group May 2019 – August 2019

- Modeled statistical properties of Brownian motion and fit the behavior to the negative binomial distribution. Project culminated in a presentation to senior NIST researchers and peers at summer symposium.

Research Intern **Gaithersburg, Maryland**
NIST – Polymer Group January 2017 – August 2017

- Cleaned, transformed, and analyzed data collected from Raman spectrometer to measure crystallinity of polymers.
- Publication: **Northcutt**, Orski, Migler, and Kotula. "Effect of processing conditions on crystallization kinetics during materials extrusion additive manufacturing", *Polymer*, vol. 154, 2018, pp. 182-187. https://doi.org/10.1016/j.polymer.2018.09.018

PROJECTS

Magma Migration Prediction with Machine Learning

- Developed novel methods for uniform geologic age prediction with ML algorithms in collaboration with USGS scientists. Used Python (pandas, NumPy, scikit-learn) to train and compare 5 ML models. Created data pipeline to integrate results into an R Shiny app for interactive visualizations. [GitHub] [Shiny Dashboard]

Heart Disease Indicators by Sex

- Conducted an analysis of heart disease data using decision trees in R (tidy, dplyr, ggplot2) highlighting disparities in identification and treatment between sexes, and proposed actionable solutions for medical researchers. [Analysis]

Understanding Hate Groups on Reddit with Natural Language Processing

- Leveraged Large Language Models (LLMs) to classify and track hate groups on Reddit. Employed Python (TensorFlow, scikit-learn, sparknlp, NLTK) with Spark for advanced NLP tasks, providing unique insight into online hate speech. [LLM Analysis]