Candice Stauffer

[redacted]

SUMMARY

Data Scientist with 6 years of extensive experience in data analytics, machine learning, and predictive modeling. Proven track record of deriving insights from data, developing sophisticated algorithms, and designing data-driven strategies.

EXPERIENCE

Data Scientist | **Roc360**

August 2022 - Present, New York, NY

• Produced company's first fully-automated dashboard (Sigma) to visualize real-estate market hot spots in the U.S., replacing inefficient Excel-based visualizations. Resulted in a 50% reduction in data processing time and eliminated the need for 20 hours of manual updates per week.
• Developed a custom NLP script to detect and mask personally identifiable information (PII) specific to Roc360, enhancing data privacy and compliance. Achieved a 95% accuracy rate in identifying and masking PII, reducing the risk of unauthorized access and protecting client confidentiality.
• Developed and maintained databases, including customer demographics and sales figures, using SQL and Snowflake.
• Regularly presented aforementioned work to the CEO and other company executives and received consistent positive feedback.

Astrophysics Research Scientist | **Northwestern University**

August 2018 - Present, Evanston, IL

• Published 4 articles in leading peer-reviewed journals with several multi-institutional collaborations.
• Led the development of a novel tool for an international collaboration of astronomers that utilized regression trees to streamline the classification of supernovae with a test accuracy of over 93% (https://github.com/cosmiccandice/DomesticCats).
• Utilized high-performance computing cluster to conduct extensive analysis on large-scale datasets.
• Developed a tool to predict galaxy redshift (distance) using untapped photometric data and leveraging random forest regression models. Achieved average bias of 1E-3 and an outlier rate of only 1.48%, which surpassed previous models.

Software Engineering Apprentice | **Code/Astro, California Institute of Technology**

June 2022 - July 2022, Pasadena, CA

• Developed and deployed "AstroMVP" (https://github.com/edenacadia/astroMVP), a Python package that automates the identification of the most relevant research paper associated with a specific keyword. Implemented a semantic similarity algorithm based on word embeddings, resulting in a 30% increase in accuracy.
• Demonstrated proficiency in employing state-of-the-art software development, testing, and code release strategies. Successfully reduced software defects by 25% through rigorous testing.
• Invited to teach Code/Astro workshops the following year and received overwhelmingly positive feedback from workshop leaders.

EDUCATION

Ph.D. Astrophysics | **Northwestern University**

Evanston, Illinois, 2023

B.S. Physics | **Portland State University**

Portland, Oregon, 2018

CERTIFICATIONS

Integrated Data Science | **Northwestern University**

2021

Certified in advanced data analytics skills spanning statistics, data mining, machine learning, and visualization, enabling impactful contributions to scientific research and data-intensive industries.

SKILLS

Machine Learning, Data Analysis, Python (Pandas, Numpy, Scikit-learn, PyTorch, Tensorflow, & more), SQL, JavaScript, Matlab, Sigma, Tableau, GitHub/git, HPC Clusters, Snowflake, AWS, MacOS, Unix, Linux, Bash