Kristina Halbig
Data Scientist

Pasadena, CA | 201-232-8213 | kristinadhalbig@gmail.com | linkedin.com/in/khalbig | https://fragariachemist.github.io/

Data scientist specializing in quantitative analysis, data visualization, and machine learning with strong background in analytical chemistry. Completed Data Science immersive bootcamp, gaining expertise in advanced data techniques, including predictive modeling, natural language processing, and data wrangling. Skilled at extracting insights from diverse datasets and driving business impact through data-driven decision-making.

SKILLS
Programming and Tools: Python, SQL, Pandas, NumPy, Scikit-learn, TensorFlow, Keras, APIs, Jupyter, GitHub, Streamlit
Data Analysis and Machine Learning: Data Wrangling and Preprocessing, Supervised and Unsupervised Learning, Ensemble Methods, Deep Learning, NLP, Predictive Analytics, Model Evaluation and Optimization
Data Visualization: Matplotlib, Seaborn, Tableau
Software and Collaboration: MS Teams, Sharepoint, Slack, Agile Methodologies

EXPERIENCE
Encore Labs   **Pasadena, California | 07/2018 - 02/2024**
Senior Analytical Chemist
- Led data wrangling for over 10,000 data points per day, supporting 270 clients and driving $600K in monthly revenue.
- Developed Tableau dashboards to visualize 500 geographic data points, improving decision-making.
- Automated data processing using QBench LIMS, reducing data errors from 15% to less than 5%.

The J.M. Smucker Company   **Burbank, California | 06/2015 - 07/2018**
Analytical Chemist
- Managed data analysis for ingredient datasets across 100 product lines, processing 20 samples daily.
- Conducted regulatory audits, ensuring compliance with 75 SOPs and assisting with ISO 17025 certification.
- Performed cost analysis, resulting in $140,000 annual savings through optimized purchasing processes.

PROJECTS
SpaceLab Exoplanetarium Hackathon   **October 2024 – Present**
Python, PyTorch, AWS, Agile Methodologies, Data Collection and Storage
- Collaborated with a cross-functional team of 10 data engineers and UX/UI designers to define data requirements for the SpaceLab Exoplanetarium project, leveraging Agile methodologies for effective project management.
- Transferred and optimized over 1,000 lines of code for AWS SageMaker notebooks, employing S3 buckets for efficient data collection and storage.

Ravelry Knitting Pattern Recommender System   **May 2024 – Present**
NLP (spaCy), Streamlit, API Scripting, Unsupervised Learning
- Developed machine learning recommender system for 15,000 patterns, extracting 13 key features through API scripting.
- Integrated data into Streamlit app, enabling users to receive 5 personalized pattern recommendations.

Subreddit Post Classifier Using Logistic Regression and NLP   **March 2024 – April 2024**
Logistic Regression, NLP (spaCy), Naïve Bayes, Scikit-learn
- Built classifier to analyze 2,800 subreddit posts, achieving 85% accuracy in predicting categories using logistic regression and NLP.
- Improved model performance by 9% through GridSearch and optimized classification pipelines, reducing misclassifications.

EDUCATION
General Assembly   **Remote | March 2024 – June 2024**
Data Science Immersive Certificate of Completion
- Completed 12-week immersive bootcamp, utilizing data exploration and data science techniques to execute 4 machine learning projects and capstone presentation.

California State University, Northridge – *Bachelor of Science (BS) in Chemistry*   **Northridge, CA**