Kristina Halbig
Data Scientist

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

Data scientist specializing in analytical chemistry, leveraging expertise in quantitative analysis and experimental design to address complex data challenges. Holding a Bachelor of Science in chemistry offers a unique perspective in data science, enabling the extraction of insights from diverse datasets and the development of innovative solutions. A decade of experience in project management, data verification, and quality assurance, further enhances the ability to design and manage strategic data initiatives that drive business success. Additionally, technical leadership and data-driven analysis skills, developed through hands-on experience in various data science projects, including machine learning, natural language processing, and predictive modeling, ensure the delivery of high-impact results.

SKILLS
Coding: Application Programming Interfaces (API), Data Cleaning, Data Wrangling, Keras, NumPy, Pandas, Python, Scikit-Learn, spaCy, SQL, TensorFlow
Visualization: Matplotlib, Seaborn, Tableau
Software: Github, Jupyter Lab, Slack, Streamlit
Machine Learning: Classification, Clustering, Decision Trees, Linear/Logistic Regression, Neural Networks, NLP, Predictive Modeling, Supervised/Unsupervised Learning

EXPERIENCE
Encore Labs   **Pasadena, California | 07/2018 - 02/2024**
Senior Analytical Chemist
- Directed data wrangling of over 10,000 data points per day, supporting 270 clients and generating $600k in monthly revenue.
- Developed and deployed Tableau dashboards to visualize 500 geographic data points, resulting in improved understanding of business performance and strategic decision-making.
- Integrated a QBench LIMS system to automate the upload and approval of data, ensuring proper handling of more than 5k data points per day and reducing errors from 15% to less than 5%.

The J.M. Smucker Company   **Burbank, California | 06/2015 - 07/2018**
Analytical Chemist
- Managed analysis and data wrangling of diverse ingredient datasets across 100 product lines, processing data packets for 20 samples daily. Conducted training, supervised analytical instrument maintenance, and documented lab procedures.
- Ensured compliance with industry standard regulations, overseeing the audit of 75 standard operating procedures to ensure compliance and assisting with the acquisition of an ISO 17025 certification.
- Conducted cost analysis, resulting in $400 daily savings through efficient purchasing.

PROJECTS

Hack For LA Volunteer   **June 2024 – Present**
Data Exploration, Dataset Evaluation
- Reviewed and assessed 85 datasets from the LA Controller's website to determine their potential for data exploration and modeling projects.
- Compiled selected datasets into a structured Excel sheet, including details such as dataset names, update frequency, categories, and links to both the datasets and their corresponding APIs.
- Identified high-value datasets to support the organization's future data-driven initiatives.

Ravelry Knitting Pattern Recommender System   **May 2024 – Present**
NLP (spaCy), Streamlit, API Scripting, Unsupervised Learning
- Built a recommender system using machine learning and spaCy NLP, leveraging a dataset of 15,000 knitting patterns.
- Developed an API script to extract and transform knitting pattern data into a structured pandas dataframe with 13 key features.
- Integrated the processed data into a Streamlit app, enabling users to input a favorite pattern and receive 5 tailored recommendations.

Subreddit Post Classifier Using Logistic Regression and NLP   **March 2024 – April 2024**
Logistic Regression, NLP (spaCy), Naïve Bayes, Scikit-learn
- Developed a classification tool using Web APIs and NLP to analyze 2,800 subreddit posts related to fountain pens.
- Applied logistic regression and Naïve Bayes models to identify top 20 phrases and brand names, providing insights into community preferences and opinions.
- Improved model performance by 9% through GridSearch and pipelines, enhancing specificity in classification.

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

California State University, Northridge   **Northridge, CA | Graduated: 2014**
Bachelor of Science (BS) in Chemistry