Hanieh Haeri

925-768-0299 | hhaeri0911@gmail.com | Alamo, CA94507
https://hhaeri.github.io/portfolio/ | https://www.linkedin.com/in/hanieh-haeri-9319b024/

PROFESSIONAL SUMMARY

Seasoned Data Scientist with a proven track record in leading and executing impactful projects. Bringing a decade of expertise in modeling, statistical analysis and machine learning augmented by my great enthusiasm for delivering innovative data and technology driven solutions to complex business problems.

EXPERIENCE

Machine Learning Researcher **2023 - Present**
University of California – San Francisco, CA

- Improving performance of a **Generative model** that uses a a novel **Variational Autoencoder** (VAE) architecture to provide interpretable deep learning models for medical imaging data (link)
- Building the trust between medical practitioners and network's decision-making processes by providing interpretable AI

Data Scientist **2023 - Present**
Freelance, Alamo, CA

- Trained a **Semantic Segmentation** model to extract key geologic features from historic USGS maps (Dice Coefficient = 0.89, Precision = 0.85, Recall=0.93). The model provides a robust automated solution for large-scale geologic feature extraction from diverse maps, effectively reducing manual feature extraction efforts by 54% (link).
- Deployed model on Databricks and AWS using PySpark to accommodate large volume of data

Engineer **2014 - 2022**
Montgomery & Associates , Oakland, CA

- Pioneered Python and SQL integration, replacing Excel for streamlined data wrangling, data analysis and data visualization tasks. Achieved a 60% boost in data processing efficiency through automation.
- Leveraged physics-based and data-driven surrogate modeling, predictive modeling, experimental design, statistical analysis and machine learning to identify, investigate, integrate and deliver clear, actionable, and timely data-driven decisions to stakeholders of varying technical experience.
- Produced informative and visually compelling geospatial analysis products, maps, and web-based GIS applications in support of project development, planning, execution, and operations

SELECTED PROJECTS

- Built an image classifier interpretation tool using a pre-trained RESNET model to create predictions for selected images and OmniXAI to delve into the underlying rationales guiding the model's decisions (link).
- Performed Natural Language Processing and sentiment analysis using the Yelp review dataset on the Hugging Face Hub. Built bag-of-word and bigram models. Built a Naive Bayes model to find the most polar words. Analyzed reviews of restaurants to find food bigrams (link).
- Designed a machine learning-based project, forecasting daily COVID-19 case counts at the county level in California. Created a user-friendly Streamlit app and deployed it on Heroku for accessibility (link)
- Constructed a social network for New York's elite by web scraping the New York Social Diary. Analyzed the assembled social graph to identify popular socialites, influential individuals, and tightly linked pairs (link)

SKILLS

Programming Languages(Python, MATLAB, Spark, SQL), Machine Learning (Scikit-Learn), Neural Networks and Deep Learning (PyTorch, TensorFlow, Keras), Computer Vision (OpenCV), Statistical Data Analysis and Hypothesis testing (SciPy, Statsmodels), Data Wrangling & Visualization (SQL, Numpy, Pandas, BeautifulSoup, Matplotlib, Seaborn, Tableau), Natural Language Processing (spacy, nltk, regex), Distributed Version Control (Git), Distributed Computing (Spark), Geospatial Data Analysis and Mapping (GIS, GeoPandas)

EDUCATION

Ph.D., Civil & Environmental Engineering, University of California, Davis **2011**
B.S., Chemical & Petroleum Engineering, Sharif University of Technology **2004**

CERTIFICATIONS

- **Large Language Models Professional Certificate**, Databricks **2024**
- **Creating Dashboards and Storytelling with Tableau**, UCDavis **2024**
- **Interpreting Machine Learning Models,** Uplimit **2023**
- **Deep Learning Specialization**, DeepLearning.AI **2023**
- **Data Scientist Certification,** The Data Incubator **2022**