

Summer McGrogan

Thousand Oaks, CA • summermcgrogan@gmail.com • (408) 806-0065 • www.linkedin.com/in/summermcgrogan
https://github.com/smcgrogan

EDUCATION

University of California, Berkeley

Master of Information and Data Science
GPA 4.0

Berkeley, CA

December 2024

Loyola Marymount University

Bachelor of Science, Applied Information Management Systems
Minor, Environmental Studies
GPA 3.8

Los Angeles, CA

May 2021

HIGH LEVEL SUMMARY

I have a working background in data analytics and my previous jobs have focused on SQL and python for data cleaning and modeling. Graduate experience gave me hands on experience with LLMs and feature extraction techniques, especially for Natural Language Processing and Computer Vision.

RELEVANT COURSE PROJECTS

Abstractive Text Summarization for Domain-Specific Articles: This project focuses on creating a high-performance abstractive summarization model for environmental science articles using PEGASUS, FLAN-T5, and BART, and fine-tuning them on them on a curated subset of environmental science papers.

Artists Image Classification: This project focuses on developing an image classification model to categorize images of paintings based on a large range of image features to increase model accuracy and generalizability.

RecycleRight: This application uses a RAG (Retrieval-Augmented Generation) system to provide accurate disposal information for various waste items. By leveraging multimodal large language models (LLaVa and Llama), it ensures users receive the correct guidance on recycling and waste management.

INDUSTRY EXPERIENCE

CalAmp

Data Science Analyst

Richardson, TX

October 2021 – Sept. 2023

- Developed machine learning models to automate car accident predictions to save 5% of employee time per year
- Created SQL queries and PowerBI dashboards to reduce accident rates of large transport vehicles by 6.5%
- Automated ETL pipeline to process semi-structured data from AWS buckets to store in Azure SQL database
- Collaborated across teams to develop data analysis results critical to retaining customer base during transition

FactSet Research Systems

Client Solutions Intern

Remote (San Francisco, CA)

June 2020 – July 2020

- Created financial dashboards for clients and coworkers highlighting specific KPI's important to the client
- Performed qualitative and quantitative analysis of financial data through acquisition and database querying
- Executed queries of financial information using FQL and presented detailed analysis of query findings
- Mastered FactSet products with an in-depth review of functionality as applied to select client data

Programming Languages: Python (Pandas, NumPy, scikit-learn, Langchain), R, SQL, Java, HTML, Javascript

Visualization: PowerBI, Tableau, Matplotlib, Plotly, ggplot, ETL

Databases/ Cloud Computing: AWS, mySQL, SQLserver, PostgreSQL, NoSQL, DB2, Neo4j, MongoDB, Redis, Docker

Machine Learning: (+ deep learning, NLP) Neural networks, recurrent neural networks, convolutional neural networks, clustering, K means, K nearest neighbors, decision trees, random forest, Scikit-learn, naive Bayes, XGBoost, recommender systems, computer vision, Tensorflow, PyTorch, Keras, NLTK, natural language processing, computer vision, generative AI

Statistics: Linear Algebra, Logistic/Linear Regression, Pandas, NumPy, SciPy, RStudio

Code Management: Git/Github, Jupyter notebooks, Visual Studio

Platforms: Windows, Linux, Unix OS, Bash, MacOS