

Gregory Glatzer

Email: gregoryg323@gmail.com
GitHub: www.github.com/g1776

LinkedIn: www.linkedin.com/in/gregory-glatzer
Portfolio: www.gregoryglatzer.com

Cell: (267)-229-6863

Education

The Pennsylvania State University, College of Information Sciences and Technology

August 2019 – May 2023

B.S. Applied Data Science, Cumulative GPA: 3.96 - Summa cum laude

Experience

KCF Technologies – Software Engineer I

June 2022 – December 2022

- Commended by leadership for efficiency improvements by modernizing data entry and internal savings tracking through development of a React application utilizing Amazon Web Services Amplify, S3, Cognito, DynamoDB, and Power BI. The project saved employees hundreds of hours of data entry per year and was presented to C-Suite members to outline savings and ROI.
- Supported a major client's real-time data processing needs by delivering an ETL pipeline built with Node-RED, MQTT, and AWS EC2. This involved transforming data into API payloads and analyzing MQTT broker throughput to optimize data flow by batching and caching incoming payloads from IoT devices.
- Built Dockerized Python applications for the Data Integration and Edge Enablement agile development teams to support client-specific needs to move data between Industry 4.0 systems.
- Operated in the Azure DevOps platform to manage backlog items, code reviews, and build and test CI/CD pipelines in live production environments at scale. Managed large code bases using version control systems (git).
- Designed and implemented SQL and NoSQL databases in AWS that interacted with a full-stack application built in React.js.
- Utilized MQTT and Modbus communication protocols to deliver real-time payloads between systems.

Capstone Project: Detecting Misleading Airbnb Listings with Machine Learning – Data Scientist

January 2023 – May 2023

- Enabled users to make informed decisions on whether Airbnb listings are misleading using a machine learning model that performed with an accuracy of 93% on the test set for a multiclass classification task.
- Processed a large tabular dataset (~900,000 rows) of Airbnb reviews and listing information using NLP techniques from the NLTK library, along with BERT to perform sentiment analysis on the dataset.
- Performed feature engineering, including generating N-grams, word embeddings, Principal Component Analysis (PCA) vectors, and One-Hot Encoding the existence of keywords in the Airbnb reviews.
- Communicated a data-driven story by analyzing results, comparing model performance, and ensuring transparency in our work.

Penn State Research: Grozinger Lab – Computer Vision Researcher

January 2023 – May 2023

- Accelerated etymology research by helping researchers efficiently scan videos and analyze bee behavior quickly and effectively.
- Designed a novel computer vision-driven approach to tracking movement of solitary bees using OpenCV that achieved an accuracy of 71%.
- Developed a novel approach to measure performance against ground truth labels by a human.
- Applied contour detection, Hough Transforms, image processing, and object tracking techniques.

Penn State Research: Mitra and Kinyua Lab – Data Science Researcher

May 2021 - November 2021

- Created insights for elephant experts and conservation researchers on the factors that affect elephant movement by applying DBSCAN and KMeans machine learning models on geospatial datasets of elephant movement in Africa.
- Published two research papers on the analysis of elephant movement in conservation conferences.
- Transformed data between spatial file formats and CSV. Developed pipelines to automate the processing of new datasets.
- Communicated results through weekly presentations and contributed creative problem-solving solutions.
- Created a data-driven dashboard to facilitate data exploration and extract insights through data visualizations.
- Demonstrated initiative and the ability to create self-driven work by independently learning and applying ML algorithms.

Publications

A Socio-Ecological Approach to Understanding How Land Use Challenges Human-Elephant Coexistence in Northern Tanzania

June 2022

- Peer-reviewed publication in MDPI Diversity with co-authors in Africa and the UK.

An Analysis of Elephants' Movement Data in Sub-Saharan Africa Using Clustering – First author

December 2021

- Presented at the Tanzania Wildlife Research Institute (TAWIRI) 13th Annual Scientific Conference in Tanzania, Africa.

Skills & Certifications

Skills

- Python, R, SQL, JavaScript, HTML/CSS, Node-RED, Java, React.js, D3.js, pytorch, git
- AWS: S3, EC2, Cognito, Amplify
- Docker, CI/CD Pipelines, REST APIs, Postman, Microsoft Office
- Pandas, sklearn, folium, plotly, matplotlib, Pyspark, OpenCV
- MongoDB, Neo4j, Redis, DynamoDB, MySQL

Certifications

- Agile at Work: Building Your Agile Team (LinkedIn Learning)
- Certified Data Scientist Level 1 Credential (General Assembly)
- Nittany AI Challenge 2020 and 2021 Finalists