Faisal Baig

US Citizen

☐ 678-770-5117 | ✉ me@faisalbaig.com | ⌂ faisab.github.io | in linkedin.com/in/baig-faisal

Education

Georgia Institute of Technology Atlanta, Georgia

M.S. in Computer Science *Grad December 2024*

- Graduate Student Assistant for CSE 6104 (Graduate CS Algorithms) - Wrote lecture notes & facilitated exams

Georgia Institute of Technology Atlanta, Georgia

B.S. in Computer Science – Zell Miller Scholar + Deans List *Grad May 2023*

Experience

Boston Consulting Group Los Angeles, California

AI Engineer Intern *Jun. 2023 – Aug. 2023*

- Incorporated multiple APIs within ingestion pipeline & built unified functionality for processing multiple formats: earnings calls, 10Ks, 10Qs
- Developed custom parsing functions for multiple formats, creating contextually richer content within token constraints (Python, LangChain, NLTK)
- Wrote ingestion functions for raw & semi-processed input data to ensure compatibility with tokenizer & vector DB
- Implemented data quality checks, filtering criteria, & post-processing for LLM output enhancing quality and reliability in user experience
- Created unit & integration tests to validate functionality and performance of data processing components (Python, Pytest, Docker, Terraform)

CertiK New York, New York

Platform Engineer Intern *Sep. 2022 – Dec. 2022*

- Automated 10+ anomaly detection models using Databricks/PySpark/Snowflake, implementing access controls, database integration, and model performance metrics logging with Datadog and MLflow
- Developed automated pipeline for creating & updating monitors of current data platform using Github Actions/Terraform. Automated migration of 320+ existing agents/monitors to infra-as-code (MonteCarlo/Datadog/Github Actions/Terraform)
- Deployed Terraform-based infrastructure to dev, staging, & live environments using Github Actions CI/CD pipelines for public and internal APIs and set up logging and monitoring (ECS, S3, Docker, Airflow, Terraform)
- Policy creation, environment setup, & IAM for ML Infrastructure and data pipelines (Databricks, Github, Snowflake)
- Investigated economic impact and proposed optimization for high consumption or high runtime Snowpipe loads & Snowflake queries

Equifax Atlanta, Georgia

Software Engineer Intern *May 2022 – Aug. 2022*

- Endpoint development and UI design of B2B portal using Springboot and Canva for the MVP and beyond of a new project (Java/BigQuery)
- Wrote functions to integrate novel data format in an active Beam ETL pipeline processing millions of records(Dataflow/Apache Beam/PostgreSQL)
- Conducted topic modeling with LDA on over 4 million support cases, top results populating unknown causes field (Python/Gensim/NLTK)
- Performed data exploration on support case dataset, reducing the feature set from 100+ columns to 13 key attributes (Python/Pandas)

Research

Georgia Tech AI Research and Discovery Lab Atlanta, Georgia

Undergrad Researcher *Jan. 2021 - May 2022*

- Developed full-stack app using Azure Blob Storage & Node.js serving REST API to allow categorization of CT Scans by radiologists at Emory University
- Test & log metrics for ML models trained on novel data to allow for categorization, localization, and gradation of different conditions lung conditions

Projects

Image Processing & Quality Assessment Pipeline | Python, OpenCV, skimage

- Multi-step image processing for consistent quality metrics: sharpness calculation, noise estimation, distortion measurement , and color accuracy.

Ensemble Learning Trading Algorithm | PyTorch, scikit-learn, REST

- Ensembles multiple supervised methods to maximize profit on historical trading data. Relevant features & stock groupings using K-Means w/DTW

Parallelized Machine Learning Library | C++, OpenMP, Eigen

- Parallelized algorithms (Linear Regression & K-means Clustering) and preprocessing in C++, leveraging Eigen for optimized matrix operations

Skills

Languages & Frameworks: Python, C++, JavaScript, SQL, Golang, PyTorch, PySpark, scikit-learn, OpenMP, d3.js

Tools: Git, AWS, Kubernetes, Docker, CI/CD, DataDog, Databricks, LangChain, NLTK, Microservices, Apache Airflow, Apache Beam, Jupyter, MLflow, Terraform, Make, Bash, Linux

Coursework: DS& A, Algorithm Analysis, Computer Architecture, Operating Systems, Software Design, Databases, Statistics, (MS) Machine Learning, (MS) Deep Learning, (MS) Network Security, (MS) High Performance Computing, (MS) Natural Language, (MS) Visual Analytics