

Swaha Bhattacharya

[✉ itsswahabhattacharya@gmail.com](mailto:itsswahabhattacharya@gmail.com) — [in linkedin.com/swaha294](https://www.linkedin.com/swaha294) — [Personal Website](#) — github.com/Swaha294

Skills

- **Technical Skills** R, R Studio, Python, SQL, Power BI, Prompt Engineering, Retrieval Augmented Generation, Fine-Tuning LLMs, Statistical Modeling, Machine Learning, Canny Edge Detection
- **Technology** Databricks, Git, GitHub, Azure DevOps, MLFlow
- **Languages** English, Hindi, Bengali, French
- **Interests** Indian Classical Singing, Yoga, Karate

Education

Smith College

Bachelor of Arts: Mathematics, Statistical and Data Sciences

Northampton, MA |May 2023

- **GPA:** Major GPA: 4.0, CGPA: 3.99
- **Honors:** Summa Cum Laude, Dean's List 2019-2023
- **Academic Societies:** Phi Beta Kappa, Mu Sigma Rho
- **Coursework:** Adv Prog in R, Probability, Mathematical Statistics, Multiple Regression, Discrete Math, Linear Algebra, Machine Learning, Real Analysis, Research Seminar in Political Networks, Multivariate Calculus

Professional Experience

Hartford Steam Boiler Inspection and Insurance Company (HSB)

Data Scientist

Hartford, CT |Aug 2023 - present

- Developed an **XGBoost classification model** optimized through grid-search **hyperparameter tuning** on 50k+ claims to identify complex high-severity claims to enhance claim handling efficiency.
- Leveraged **few-shot prompting and prompt chaining** on unstructured notes using LLaMA3 for feature extraction in collaboration with the claims team to enrich predictors with claim-specific information for the model developed for identifying complex claims.
- Leveraged **Python mlflow and Azure DevOps** in collaboration with machine learning engineers for version control, streamlining code, and deploying the classification model developed for identifying complex claims.
- Developed a severity model for insurance claims, broken down by 3 individual coverage types, by leveraging the **Gamma and Dirichlet distributions** with a **Generalized Additive Model (GAM)**. Guided the summer 2025 intern on our team to further refine the model using **residual analysis, XGBoostLSS, and Bayesian modeling**.
- **Fine-tuned** a pre-trained open-source LLM, ALBERT using **contrastive learning loss** with 99% accuracy to augment training data for fine-tuning another LLM.
- Leveraged **binary and ranking learning losses** to fine-tune a pre-trained LLM, ALBERT to automate the identification and labeling of damaged equipment of insurance claims with 70% accuracy.
- Implemented a **Retrieval Augmented Generation (RAG)** pipeline with LLaMA3 using **parent-child chunking** to gain meaningful insights into the scope for future insurance products from 20k+ claims.
- Leveraged **Canny Edge Detection and Hough Transform** along with **multi-modal LLMs** for automated image pre-processing and **visual question answering** of claim forms submitted as handwritten or digitally completed PDFs.

Smith College

Statistical and Data Sciences Teaching Assistant

Northampton, MA |Aug 2021 – May 2023

- Worked as a TA for **introductory statistics** courses as well as **advanced programming and data science classes** like "*Introduction to Probability and Statistics*", "*Introduction to Data Science*", "*Advanced Programming for Data Science*", and "*Seminar: Capstone in Statistical & Data Sciences*"
- Helped instructor **develop course structure and student evaluation metrics** for "*Advanced Programming for Data Science*"
- Helped students navigate **Git and GitHub** and reviewed their code on GitHub pull requests for "*Advanced Programming for Data Science*"
- **Led discussions on Data Science ethics** based on assigned readings for a class of 30 students in "*Seminar: Capstone in Statistical & Data Sciences*"
- **Tutored 100+ students** in statistical theory and programming methods through weekly office hours
- Graded assignments and projects for statistical theory classes like "*Introduction to Discrete Mathematics*", "*Probability*" and "*Seminar: Mathematical Statistics*"

Munich Re America Services

Data Assets Intern

Princeton, NJ |May 2022 – Aug 2022

- Collaborated with underwriters and actuaries to create reports using **PowerBI and Excel** on existing reinsurance submission data and made recommendations for further structuring and use of client submission data

- Optimized **R**, **RStudio**, and **SQL** to clean data on commercial trucks to create predictive regression models for truck accidents

Reliance Industries

Data Analysis Intern

Mumbai, India | May 2021 – Aug 2021

- Created compelling interactive data visualizations on **Tableau** to portray and analyze existing data on trends in Reliance Petrochemical sales
- Analyzed the dashboards to find new visualizations or modifications to the existing ones to improve their efficiency

Indian Embassy in the U.S.

Qualitative Research Volunteer

Remote | Dec 2020 – Mar 2021

- Worked with volunteers from the embassy to research the best micro-innovations established in local schools across Indian states for the development of teaching methods for under-privileged students
- **Collected and cleaned data through interviews** for the qualitative analysis of the best micro-innovations established in local schools in India
- **Created a detailed report on the findings** and a plan of action for future implementation

Leadership Experience

Hartford Steam Boiler Inspection and Insurance Company (HSB)

Data Scientist

Hartford, CT | Aug 2023 - present

- **Led a three-person team** to coordinate monthly Lunch and Learn presentations for 500+ members of the Actuarial and Analytics Community across all of MunichRE North America.
- **Designed and led the 2024 Summer Intern Capstone** project across HSB to help 25 interns create a handbook on **prompt engineering best-practices** for diverse business purposes broken down by data governance and GenAI applications like text classification, text summarization, question-answering, and automated research. Further guided them on creating a streamlined presentation for HSB senior leadership.
- **Coordinated an effort across 10+ members** from the Data Engineering, IT, and Claims teams at HSB to establish efficient data pipelines in the Azure Datalake to make it available for Data Science modeling purposes.

EKTA South Asian Student Organization

Chair

Northampton, MA | Aug 2021 - May 2023

- Led the student group to **cultivate South-Asian culture** at Smith College through educational and cultural events
- Liaised with other student groups, administrative offices, and academic departments at the college to enable events inclusive of the diverse student body

Projects

R IPL Package

Contributor

- The goal of ipl is to provide data and functions related to the Indian Premier League (IPL) matches and standard cricket statistics. This package allows users to avail of IPL data and conduct analysis of cricketers and IPL teams using functions of this package.

SDS Capstone Project

Contributor

- Collaborated with teammates and experts at the Blue Cross Blue Shield of Minnesota (BCBSMN) to train algorithms to predict the probability of patients having hierarchical conditions (HCC) based on their medical history using tree-based models, logistic regression, and feature selection through L1 regularization.
- Used statistical metrics like balanced accuracy and F1 score to select the most effective model for each HCC.