Anurima Saha

San Diego, California

Ph: (+1) 6173048280 Email: asaha8669@sdsu.edu LinkedIn: Link to Profile GitHub: Link to Profile

SKILLS

Statistical Computing Language – Python, R, SAS, STATA

Machine Learning Algorithms –Linear & Logistic Regression, Support Vector Machine, Clustering, Decision Trees, Random Forest, XGBoost

Deep Learning (Pytorch) – CNN, RNN, LSTM , Transformers, LLM

Statistical Skills – Regression, Classification, Clustering, PCA, Time Series, Statistical Testing, Multivariate Statistics

Visualization and Report – MS Excel, Tableau, PowerBI

Database Querying – SQL, Snowflake

EXPERIENCE

AAA — *Data Science Research Intern*

MAY 2024 – PRESENT

- Worked on **model refresh for risk segmentation** of Electric Vehicles using **statistical analysis and machine learning**
- **Collaborated** with team to extract data using Snowflake and develop understanding of insurance domain
- Performed extensive **exploratory data analysis** to identify new feature for EV loss data modeling
- Developed an **ensemble machine learning model** in **Python** using **AWS** to understand the impact of newly identified feature
- Incorporated feature into existing segmentation model to achieve a **Gini improvement of 11.6%**
- **Presented insights** to business leaders and cross-functional teams for implementation

HSBC – *Senior Analyst*

SEP 2021 – AUG 2022

- Actively **managed projects** of small/medium complexity effectively **communicating analytical solutions** to business heads
- **Defined** a business problems, **collected** required **data**, **analyzed** the results **synthesizing compelling insights**
- Developed **advanced analytics solutions** including forecasting, predictive modeling, clustering, and prescriptive analytics to solve business problems using **Machine Learning in Python**
- **Collaborated with cross-functional stakeholders** to understand their business needs and formulated end-to-end analysis

HSBC — *Analyst*

JUL 2019 – AUG 2021

- Performed exploratory data analysis on sales and revenue data followed by **forecasting using machine learning** in **R, RShiny and Python**

- Built reusable, and maintainable models that **handle large amounts of data**
- Implemented data driven solutions using **reports and visualizations** in **Excel and Tableau** to communicate data insights to stakeholders.
- Assisted with **data pulling** from data lake followed by **data engineering** and reduction.
- Developed **professional competency** in **MS Office tools** like PowerPoint, Word, Outlook and Teams

EDUCATION

San Diego State University — *MS (Big Data Analytics and Data Science)*
AUG 2023 - MAY 2025
GPA - **4.0**

University of Calcutta — *MSc (Economics)*
AUG 2017 - JUN 2019
GPA - **3.78**

AWARDS

Graduate Research Excellence Award, 2024 (SDSU) - Across all departments of the university

Star of Business Service, 2022 (HSBC) - Company Level (1 among 12000+ employees)

Team Star, 2020 (HSBC) - Organisation Level (1 among 2000+ employees)

KEY PROJECTS

Deep Learning Based Recipe Discovery and Recommendation (associated with SDSU)

Deep learning and Natural Language Processing used to develop an automated system capable of accurately recognizing a wide range of fruits and vegetables from their images using advanced Vision Transformer. The methodology involves the use of transfer learning to fine-tune pre-trained vision models on the Fruits and Vegetables dataset from Kaggle followed by NLP to offer personalized recipe recommendations based on the identified ingredients.

Identifying Conspiracy Theories from Reddit Texts (associated with SDSU)

This project involved the use of Natural Language Processing for data cleaning followed by term extraction using *Spacy*. LDA model from *tomotpy* was used to identify 20 topics. Embedding was done using *Spacy* and *SBERT* followed by using *TruncatedSVD* to transform data into a 20 dimensional vector. HDBSCAN was used to identify 20 clusters, plotting topics in interactive *bokeh* plot.

Predictive Client Prospecting Model (associated with Global Trade, HSBC UK)

Ensemble Modelling techniques used in Python to recommend ~200 clients (base >22k customers) for upsell initiative rendering benefit of ~11 million GBP with prediction accuracy of ~84%.

Model for Automatic Credit Approval Decision Engine (associated with Invoice Finance, HSBC UK)

Created an Explainable AI in Python for credit risk mitigation tailored for pandemic era recession generating a benefit of ~30 million GBP.

Anomaly Detection and Pattern Recognition Model (associated with Receivable Finance, HSBC UK)

Supervised Machine learning used to develop a framework for early detection and automated trigger report for clients with a likelihood of default in the next 6 months with an accuracy of ~81

⋆ **For more details please refer to my website -** Website