

SUBHAJIT BAG

(404) 578-3235 suvo.98890@gmail.com linkedin.com/in/subhajit-bag github.com/shuvo-iitkgp

EDUCATION

Georgia Institute of Technology, Atlanta, Georgia

Aug 2025 – Dec 2026

Master of Science in Analytics (Computational Data Analytics Track)

Indian Institute of Technology, Kharagpur

Jun 2019 – May 2023

Bachelor of Technology in Chemical Engineering | Minor in Mathematics & Computing

CGPA - 8.83

EXPERIENCE

American Express

Gurugram, India

Analyst - AI Research (AI Labs)

Jun 2023 – Jul 2025

- Built **GovLLM RAG system** to answer internal compliance queries, decreasing model audit findings by **50%**.
- Built **CUDA-optimized Data Shapley framework** for datasets $\geq 1M$ rows, cutting training time by **50%**.
- Explored GPU-based simulations to accelerate ML workflows, achieving **40%** faster model iteration cycles.
- Developed **LLM-powered retrieval-augmented engine** (Mixtral + Pandas-Agent) that reduced report time by **30%**.
- Led design of multi-objective HP tuning system (OCC-compliant) adopted by **290+** teams, boosting lifecycle efficiency.
- Applied feature selection reducing variables by **30%** with **1–2%** lift in Gini across credit risk models.

Indian Institute of Management Ahmedabad

Ahmedabad, India

Research Intern

Apr 2023 – Jun 2023

- Boosted cybersecurity risk-scoring accuracy to **87%** ($\uparrow 2\%$) by modeling unstructured data as temporal knowledge graph
- Engineered adaptive multi-head attention to detect optimal heads, reducing parameters by **35%** with no drop in F_1 score
- Improved model interpretability using SHAP and attention heatmaps; accepted at Decision Support Systems Journal

Truminds Software Systems

Bengaluru, India

Data Science Intern

May 2022 – Jul 2022

- Engineered pipeline parsing 1 TB/day network logs to detect tunnel-down events; trained BERT model with 94% recall
- Shipped **Streamlit + MLflow** web app that surfaces top-K root-causes in 2 s, deployed on on-prem Kubernetes cluster

PUBLICATIONS

- Enhancing Cybersecurity Risk Assessment using Temporal Knowledge Graph-based Explainable Decision Support System**
Accepted by Decision Support Systems Journal (2025)
Subhajit Bag, Sobhan Sarkar, Indranil Bose
- SENE: A Novel Manifold Learning Approach for Distracted Driving Analysis with Spatio-Temporal and Driver Praxeological Features**
Accepted by Engineering Applications on Artificial Intelligence Journal (2023)
Subhajit Bag, Rahul Golder, Sobhan Sarkar, Saptashwa Maity
- Identifying Spammer Groups in Consumer Reviews using Meta-Data via Bipartite Graph Approach**
Accepted by International Conference on Data Analytics for Business and Industry (2022)
Varun Balakrishna, Subhajit Bag, Sobhan Sarkar

COMPETITIONS

Open IIT Data Analytics | Sponsored by Brillio

May 2021

Overview: Predicted popularity of **4000+** songs to help a company bid to maximize revenue

Position: 01 out of 48

- Employed Voting Classifier model, with Random Forest and SVC and Gradient Boosting to predict popularity of song
- Attained an accuracy of **91.2 %** and generated a revenue of **10,500** which was **85 %** the maximum amount possible

TECHNICAL SKILLS

Programming & Analytics: | Python | R | C++ | SQL | Pandas | Numpy | Seaborn | Matplotlib | MS Excel | Tableau | D3.js |
ML & GenAI: | Gradient Boosting | XGBoost | LlamaIndex | Optuna | PyTorch | Langchain | RAG (FAISS) | Transformers |
Cloud & Other Technology: | Spark | Hadoop | GCP | Git | CUDA | Docker | Streamlit | MLFlow | Azure ADF | Databricks |

EXTRACURRICULAR ACTIVITIES

Mentor | Students Welfare Group | IIT KGP: Responsible for mentoring **10** junior UG students in their academics

Volunteer | Child Rights & You | IIT KGP: Successfully conducted teaching sessions for **90+** underprivileged students