

SHAMIK BASU

Los Angeles, CA | +1 (213) 255-6219 | shamik1900@gmail.com | LinkedIn: <https://www.linkedin.com/in/shamikofficial/> |
Github: <https://github.com/ShamikOfficial> | Portfolio: <https://shamikofficial.github.io/>

Professional Summary

Data Scientist and MS Data Science candidate at USC (GPA 3.78) with 2+ years of production experience building RAG pipelines, LLM inference workflows, and ML-driven analytics systems at scale. Designed systems processing 5M+ records/month, cut operational costs by 72%, and won 1st Place at UCLA SAIRS Hackathon 2025, judged by panelists from Microsoft, IBM, NVIDIA, Google, and LinkedIn.

Education

University of Southern California

January 2025 - December 2026

Master of Science, Data Science | GPA: 3.78 / 4.0 |

Los Angeles, CA

Coursework: Machine learning, Data Science, Deep Learning, Data Management

SRM Institute of Science and Technology

May 2018 - June 2022

Bachelor of Technology, Computer Science | GPA: 3.46 / 4.0 |

Chennai, India

Coursework: Machine Learning, Artificial Intelligence, Data Structures & Algorithms, Probability & Queuing Theory

Work Experience

Data Science Associate Intern | *KCC Capital Partners* | Los Angeles, CA

January 2026 - Present

- Fine-tuning and integrating an open-source SLM into a production JavaScript/Docker chatbot service to automate client service request classification and routing, reducing handling overhead for the automation team.

Data Scientist | *Bajaj Finserv Health* | Pune, India

November 2023 - December 2024

- Architected a real-time medical document analytics system using RAG, LangChain, and REST APIs processing 5M+ records/month at 92% extraction accuracy, replacing a fully manual workflow.
- Engineered LLM-based inference pipelines (GPT-3.5 Turbo) to automate high-complexity decision workflows, cutting operational costs by 72% and freeing downstream teams to expand into new service lines.
- Built modular monitoring pipelines with LangChain and Langfuse for model observability, reducing GPU compute utilization by 15% during experimentation and yielding sustained infrastructure cost savings.
- Integrated ML model outputs into Power BI dashboards in collaboration with product and engineering, cutting ad-hoc reporting turnaround time by 42% and enabling self-serve analytics for stakeholders.

Associate Data Scientist | *Bajaj Finserv Health* | Pune, India

July 2022 - October 2023

- Developed a supervised ML model (Logistic Regression) for workforce performance prediction delivering actionable, statistically-validated insights that improved workforce efficiency outcomes by 22%.
- Redesigned the NER-based name-matching algorithm in the fraud detection pipeline, increasing policyholder identification accuracy by 27%, reducing false positives, and mitigating fraudulent claims against non-insured members.
- Processed 10M+ records in Azure Synapse using SQL to deliver reliable business intelligence reports directly supporting senior stakeholder decisions.

Data Engineer Intern | *Bajaj Finserv Health* | Pune, India

January 2022 - June 2022

- Ran A/B tests and cohort analyses to identify key user behavior patterns, surfacing insights that improved web conversion by 37%.
- Designed a distributed analytics system in C++, Trino, and Docker to support 10M+ records across 200+ features, accelerating data-driven decision-making for product teams.

Projects

EcoMateAI | *UCLA SAIRS Hackathon - 1st Place, Sustainability Track* |

April 2025 - April 2025

- 1st Place (Sustainability Category) - Presented to industry panellists from Microsoft, IBM, NVIDIA, Google, and LinkedIn.
- Built analytics-driven Streamlit dashboards using Generative AI (Gemini 2.5 Flash) to produce insights and recommendations.

Custom CUDA Library for CNN Pre-Processing

January 2025 - February 2025

- Engineered custom CUDA kernels for matrix multiplication and image convolution, benchmarking CPU, naive CUDA, shared memory tiled CUDA, and cuBLAS across matrix sizes up to N=2048 on NVIDIA Tesla T4 GPUs.
- Compiled optimized CUDA kernels into .so shared libraries with Python bindings, achieving exponential throughput gains and enabling seamless GPU acceleration inside standard data science workflows.

Leadership

GRIDS Club, USC | *Vice President*

September 2025 - Present

- Lead analytics workshops, ideathons, and data-driven projects for 250+ members across USC's largest data science organization.

USC Viterbi School of Engineering | *Graduate Student Mentor*

January 2025 - Present

- Mentoring 6 graduate students on data science career paths, technical communication, and professional development at USC.

Technical Skills

Programming: Python, SQL, C/C++, CUDA, Bash, JavaScript

Analytics: Ad-Hoc Analysis, Predictive Analytics, Demand Planning, KPI Design, Scenario Analysis, Executive Reporting

ML / AI: Logistic Regression, Classification, Clustering, NLP (NER, BERT, spaCy), Deep Learning, LLMs, RAG, scikit-learn, LangChain, Langfuse, TensorFlow, PyTorch

Data Platforms: Pandas, NumPy, Azure Synapse, MySQL, MongoDB, PostgreSQL, Snowflake, Trino, SQL Server

Visualization: Power BI, Tableau, Plotly, Matplotlib, Streamlit

Engineering: FastAPI, Docker, Kubernetes, CI/CD, REST APIs, ELK Stack