SHAMIK BASU

Los Angeles, CA | +1 (213) 255-6219 | shamik1900@gmail.com | Linkedin | GitHub | Portfolio

Professional Summary

Graduate Data Scientist pursuing M.Sc in Data Science at USC with experience implementing and deploying large-scale ML Models and GenAI systems. Built and maintained 200+ CI/CD pipelines, delivered 4 production grade analytics frameworks, and developed 5+ FastAPI services powering lead-scoring and churn models that increased conversion by 40%, reduced churn by 22%, and drove 70%+ operational efficiency gains.

Education

University of Southern California **January 2025 - December 2026**

Master of Science, Data Science | GPA: 3.78 *Los Angeles, CA, USA*

- **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 *Chennai, TN, India*

- **Coursework:** Machine Learning, Artificial Intelligence, Data Structures & Algorithms, Probability & Queueing Theory

Work Experience

Data Science Associate Intern | *KCC Capital Partners* | Los Angeles, CA **January 2026 - Present**

- Worked with the Automation Team to develop a seamless chatbot service using JavaScript and Docker, to streamline service request handling and seamless delivery speed for client requests

Data Scientist | *Bajaj Finserv Health* | Pune, India **November 2023 - December 2024**

- Architected and deployed a real-time medical document analytics system using RAG architecture delivered through REST APIs, processing 5M+ records/month with 92% accuracy to bridge the gap between experimental ML models and production services.
- Engineered NLP and LLM-based inference workflows (using GPT-3.5) to automate complex decision processes, directly reducing operational costs with 72% yield
- Built modular analytics and monitoring pipelines using LangChain and Langfuse to enhance model observability, reducing computational re source utilization by 15% during experimentation phases
- Optimized and aggregated ML pipelines using CI/CD best practices and ELK based logs monitoring, reducing pipeline errors by 20-45% accelerating model deployment
- Collaborated with product and engineering teams to integrate ML outputs into Power BI dashboards, reducing turnaround time of ad-hoc reporting tasks by 42% and enabling self-serve analytics

Associate Data Scientist | *Bajaj Finserv Health* | Pune, India **July 2022 - October 2023**

- Developed supervised Machine Learning model (Logistic Regression), improving workforce efficiency outcomes by 22% through actionabl e and statistical insights and reward based model
- Implemented significant changes to NER based name-matching algorithm for fraud detection system, increasing user identification accuracy by 27% reducing False Positives and significantly mitigating fraudulent claims against non-insured members
- Constructed forecasting and scenario-analysis models trained on 10M+ records using cross-validation metrics, delivering reliable insights that improved strategic planning accuracy for senior leadership
- Processed big data (10M+) in Azure Synapse delivering reliable insights with SQL to support business decisions for senior stakeholders

Data Engineer Intern | *Bajaj Finserv Health* | Pune, India **January 2022 - June 2022**

- Performed A/B testing and cohort analysis using statistical methods, identifying key user behavior that improved conversion metrics by 37%
- Designed a distributed analytics system using C++, Trino, and Docker to support over 10 million records and 200+ features, accelerating data-driven decision-making for product teams
- Built and monitored robust data pipelines using SQL Server and Cloud services, reducing reporting latency by 20% to enable faster business intelligence reporting

Projects

EcoMateAI | ***UCLA SAIRS 2025 Hackathon*** **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 recomendations.
- Implemented LLM reasoning + embeddings to estimate carbon impact and recommend behavior optimization.

Leadership

GRIDS Club, USC | *Director of AI & Data Innovation* **September 2025 - Present**

- Led analytics workshops, ideathons, and data-driven projects for 250+ members.

USC Viterbi School of Engineering | *Graduate Student Mentor* **January 2025 - Present**

- Mentored 6 graduate students on analytics careers, communication, and professional development.

Technical Skills

Programming: Python, SQL, Bash, C, C++, JavaScript
Analytics: Ad-Hoc Analysis, Predictive Analytics, Demand Planning, KPI Design, Scenario Analysis, Executive Reporting
Machine Learning: Logistic Regression, Classification, Clustering, NLP (NER, BERT, SpaCy), Deep Learning (OCR, TensorFlow, PyTorch)
Data Platforms: MySQL, SQL Server, MongoDB, Azure Synapse, Trino, Enterprise Data Warehouses, Snowflake, PostgreSQL
Visualization: Power BI, Tableau, Excel, PowerPoint, Dashboard Design
Engineering: FastAPI, Docker, CI/CD, ELK Stack, Kubernetes