

# SHAMIK BASU

Los Angeles, CA, USA

+1 (213) 255-6219 | [shamik1900@gmail.com](mailto:shamik1900@gmail.com) | [linkedin.com/in/shamikofficial/](https://linkedin.com/in/shamikofficial/) | [github.com/ShamikOfficial/](https://github.com/ShamikOfficial/)

## EDUCATION

### University of Southern California

*Master of Science, Applied Data Science (GPA: 3.7)*

- **Coursework:** Machine Learning, Data Science, Data Management, Deep Learning

**Jan 2025 - Dec 2026**

*Los Angeles, CA, USA*

### SRM Institute of Science and Technology

*Bachelors in Technology, Computer Science (GPA: 3.46)*

**May 2018 - Jun 2022**

*Chennai, TN, India*

## SKILLS

- **Programming:** Python, C/C++, R, SQL, Bash, Shell Scripting, Linux
- **Technologies:** TensorFlow, Machine Learning, Deep Learning, Keras, REST, ELK, Apache Kafka, Matlab, Spark, Salesforce
- **Data & Cloud Platforms:** Microsoft Azure, MongoDB, PostgreSQL, MySQL, Scikit-learn, TensorFlow, Predictive Modeling, Customer Cohorts, Informatica, A/B Testing, Docker, Kubernetes, NumPy, Pandas, Med-Spacy, Computer Vision, Tableau
- **Certification:** Python for Data Science and AI – IBM, Databases and SQL for Data Science by IBM, Open Source tools - IBM
- **Interests:** Keyboard/Piano, Table-Tennis, Soccer

## WORK EXPERIENCE

### Bajaj Finserv Health (Financial services: \$80B+ assets|100M+ customers)

**Nov 2023 - Dec 2024**

*Data Scientist*

- **Led cross-functional engineering teams, collaborating with Product, Architecture, and senior stakeholders delivering end-to-end AI and data-driven solutions, while also driving on-campus engagement through hackathons and technical hiring**
- Leading product development initiatives to create a Medical Health Files Processing framework using **LLM** integration through **Langchain, Langfuse, RAG** and **MCP** architectures, improving data collection and enhanced product personalization
- Designed and deployed a GTM **Agentic AI** platform service for automated Cart Recommendation, reducing TAT by 83%, achieving a 92% capture rate, and ensuring detail-oriented extraction accuracy of 96%, eliminating manual processing
- Migrated asynchronous cronjobs to **Kafka** and **MySQL** to **MongoDB** database, enhancing scalability and real-time processing.

### Bajaj Finserv Health | Associate Data Scientist

**Jul 2022 - Oct 2023**

- Optimized NER name-matching algorithm (Fraud Detection), enhancing policy holder identification accuracy by 27%
- Integrated Lead Allocation Framework through **Salesforce** for prioritized lead allocation and decision-making based on lead quality **forecasting** and geographic proximity **modeling** to field agents, reducing churn by 33%
- Created a data pipeline for **predictive modelling** of Employee Performance calculator and integrated it using **REST API** contributing to 22% Performance improvement in low performing category and 15% overall impact in operational management

### Bajaj Finserv Health | Data Engineer Intern

**Jan 2022 - Jun 2022**

- Optimized Web Engagement by **multivariate** Events analytics & retention strategy on **Google Cloud** improving conversion by 37%
- Applied Distributed Systems design to engineer Disaster Recovery, Optimization and Monitoring System for Customer CUBE (**Enterprise SQL Server**) with over 5M+ Data Points and 200+ Features wide driving product analytics
- Established Customer Cohorts, **A/B Testing** Strategies to enhance product personalization revamping conversion by 17%

### Reomnify | Data Engineer Intern

**Nov 2020 - Jan 2021**

- Engineered custom **web scraping** solutions for 500+ templates with version control, reducing manual data collection.

## PROJECTS

### EcoMateAI - Carbon Footprint Analyzer | <https://ecomateai.streamlit.app/>

**Aug 2025 - Aug 2025**

*UCLA SAIRS*

- Built and hosted a gamified Generative AI based carbon footprint analyzer using StreamLit that tracks daily activities and offers personalized suggestions to optimize sustainability goals in day-to-day tasks

### BizScoutAI – Geospatial Business Recommendation

**Mar 2025 - May 2025**

- Developed an AI-powered platform that assists in identifying optimal locations for new businesses in urban areas. The system integrates geospatial data clustering, demographic trends, and economic indicators to provide recommendations

## LEADERSHIP

### GRIDS Club (Graduates Rising in Informatics and Data Science)

**Sep 2025 - Present**

*Director of AI & Data Innovation*

- Leading executive board members responsible for projects, tech-talks and workshops while overseeing tech recruitments

### USC Viterbi School of Engineering

**Aug 2025 - Present**

*Graduate Student Mentor*

- Guided peers on academics, career, and USC life to foster a strong graduate community