SHAMIK BASU

Los Angeles, CA | shamik1900@gmail.com | (213) 255-6219 | https://www.linkedin.com/in/shamikofficial

PROFILE

Data Scientist and Machine Learning Engineer with hands on experience developing and deploying production ready models, leading data driven projects, and building scalable pipelines using Python, SQL, TensorFlow, PyTorch, and cloud platforms including AWS and Azure. Proven success delivering measurable impact such as an 83 percent reduction in processing time, a 27 percent improvement in model accuracy, and a 33 percent reduction in customer churn through advanced predictive modeling. Experienced in leading teams, architecting end to end ML solutions, and applying deep learning and statistical methods to real world problems. Currently pursuing Data Science and AI or ML Engineering internships for Summer 2025.

EDUCATION

Master of Science, Applied Data Science | GPA: 3.7
University of Southern California (USC), Los Angeles, CA | Jan 2025 – Dec 2026
- **Coursework:** Machine Learning, Deep Learning, Data Science, Data Management, AI Systems

Bachelor of Technology, Computer Science | GPA: 3.46
SRM Institute of Science and Technology, Chennai, India | May 2018 – Jun 2022

PROFESSIONAL EXPERIENCE

Data Scientist   Nov 2023 - Dec 2024
Bajaj Finserv Health (Financial services: $80B+ assets|100M+ customers)
- Led a team of 3-5 engineers as a Team Lead, achieving an increase in project efficiency and successful delivery of multiple end-to-end GTM software solutions from inception to deployment.
- Orchestrated cross-functional engineering teams utilizing Agile methodologies to deliver end-to-end AI and data-driven solutions, resulting in increased project efficiency.
- Drove on-campus engagement by organizing and hosting National level Hackathon, resulting in an increase in student participation and technical skill development.
- Spearheaded product development initiatives to create a Medical Health Files Processing framework using innovative architectures, improving data collection efficiency, product personalization, and user adoption rates.
- Successfully launched a GTM Agentic AI platform service for automated Cart Recommendation, leading to an 83% TAT reduction, a 92% capture rate, and a 96% extraction accuracy rate, enhancing efficiency and precision in processing.
- Oversaw the transition of asynchronous cronjobs to Kafka, replacing MySQL systems with MongoDB database, to enhance scalability and enable real-time processing capabilities.

Associate Data Scientist   Jul 2022 - Oct 2023
Bajaj Finserv Health
- Successfully optimized the NER name-matching algorithm for fraud detection, resulting in a 27% increase in policy holder identification accuracy.
- Implemented Lead Allocation Framework via Salesforce to prioritize lead allocation and enhance Account management decision-making based on lead quality forecasting and geographic proximity modeling, resulting in a 23% reduction in churn.
- Spearheaded the creation of a data pipeline for predictive modeling of Employee Performance calculator and its integration via REST API, resulting in a substantial 22% performance enhancement in the low-performing category and a 15% overall improvement in operational management.

Data Engineer Intern   Jan 2022 - Jun 2022
Bajaj Finserv Health
- Improved conversion rates by 37% through the optimization of web engagement using multivariate events analytics and retention strategy on Google Cloud.
- Engineered a Disaster Recovery, Optimization, and Monitoring System for Customer CUBE (Enterprise SQL Server) with over 5M+ Data Points and 200+ features, utilizing Distributed Systems design principles to enhance product analytics.
- Revamped product personalization and increased conversion rates by 17% by establishing customer cohorts and implementing A/B testing strategies.

Data Engineer Intern   **Nov 2020 - Jan 2021**
Reomnify
- ▪ Developed and optimized custom web scraping solutions for over 500 templates with version control, significantly improving data collection efficiency through automation.
- ▪ Collaborated with cross-functional teams to build and deploy data pipelines for real-time data processing and analysis, leveraging technologies like Apache Kafka or Apache Spark.
- ▪ Contributed to the design and optimization of database structures to support efficient data storage and retrieval processes, utilizing SQL and database optimization techniques.

PROJECTS

SAIRS Hackathon Winner - EcoMateAI | https://ecomateai.streamlit.app | UCLA SAIRS |Aug 2025 - Aug 2025
- ▪ Developed and launched EcoMateAI, a gamified Generative AI sustainability assistant using Streamlit, recognized as the winning project at the UCLA SAIRS Hackathon in August 2025.
- ▪ Implemented daily activity tracking features and personalized optimization strategies within the EcoMateAI platform to engage users in sustainable practices, resulting in an increase in user engagement.
- ▪ Integrated LLM reasoning capabilities with structured environmental datasets to provide real-time carbon impact insights to users, enhancing environmental awareness and promoting sustainable decision-making.

BizScoutAI - Geospatial Business Recommender | Mar 2025 – May 2025
- ▪ Conceptualized and developed BizScoutAI, a geospatial business recommender system, leveraging clustering, demographic analysis, and economic indicators to pinpoint optimal business locations within the market.
- ▪ Utilized machine learning-driven ranking metrics and interactive visualization layers to empower users with data-driven insights for informed decision-making, resulting in an increase in strategic business placement accuracy.

LEADERSHIP

Director of AI & Data Innovation | GRIDS Club (Graduates Rising in Informatics and Data Science) | Sep 2025 - Present
- ▪ Spearhead the executive board, driving innovative projects, facilitating tech-talks, and organizing workshops to foster a collaborative environment for emerging talent in informatics and data science.
- ▪ Orchestrate tech recruitment initiatives and overseeing the recruitment process to onboard top-tier tech talent, ensuring alignment with the club's strategic goals and vision for advancement in AI and data innovation.

Graduate Student Mentor | USC Viterbi School of Engineering Aug 2025 - Present
- ▪ Mentor graduate students on academic strategies, career development, and navigating university life to cultivate a supportive and connected graduate community.

TECHNICAL SKILLS

- ▪ **Programming**: Python, C/C++, R, SQL, Bash, Linux
- ▪ **AI/ML**: TensorFlow, Keras, Scikit-learn, Deep Learning, NLP, RAG, LLMs, LangChain, Agentic AI
- ▪ **Data Engineering**: Kafka, Spark, ETL, Informatica, Data Pipelines, REST APIs, Docker, Kubernetes
- ▪ **Cloud & Databases**: AWS, Azure, MongoDB, MySQL, PostgreSQL, Elasticsearch, Redis
- ▪ **Analytics & Tools**: A/B Testing, Predictive Modeling, Computer Vision, Tableau, NumPy, Pandas
- ▪ **Certifications**: Python for Data Science & AI (IBM), Databases & SQL for Data Science (IBM)