

Avdhoot Patil

awwdudee.space | avdhootdattatr.patil@stonybrook.edu | linkedin.com/in/avdhoot-patil

EDUCATION

Stony Brook University

Master of Science in Data Science

Relevant Coursework: Data analysis, Probability, Data Science fundamentals, Data management, Statistical Computing, Big data systems, Big data analysis, Deep Learning, Machine Learning, Smart energy in the Information Age.

Stony Brook, NY

Aug 2023 – May 2025

Xavier Institute of Engineering, University of Mumbai

Bachelor of Engineering in Information Technology

Mumbai, India

Jul 2018 – Jun 2021

SKILLS

Languages and Frameworks: Python, Javascript, SQL, GoLang, Terraform, React, Django

Tools and Platforms: GCP, AWS, Azure, Sagemaker, Kubernetes, Docker, Tensorflow, sci-kit learn, Deep learning, NLP, Langchain, PySpark, Elastic Search, Kibana, Power BI, Prometheus, Grafana, REST, Redis, Pub-Sub, Nginx, MongoDB

Certification: Google Cloud Platform Certified Associate Cloud Engineer [[Credential link \(GCP ACE\)](#)]

EXPERIENCE

Research Assistant | *Research Foundation at SUNY, Stony Brook, NY*

Jun 2024 - present

Project: Stony Brook Medicine - Long COVID

- Developed **scalable data pipelines** in **PySpark** to process over **6 million** COVID patient records, reducing processing time by 40% for the Long COVID project.
- Collaborated with **cross-functional teams** to define requirements for data analysis and statistical computations, driving insights into **healthcare** trends. Contributed to two high-impact research queries by leveraging distributed computing environments and advanced statistical techniques.

Software Engineer (Data Science) | *NeoSoft Technologies, Mumbai, India*

Jun 2021 - Jun 2023

Project 1: LabCorp Annotation Product

- Engineered a **model training and deployment framework in Python**, enabling seamless model training, deployment, and live servicing on **AWS Lambda and API Gateway**, contributing to **30% faster model deployment**
- Deployed **30+ production-level ML models**, automating orchestration for preprocessing, hyperparameter tuning, and real-time/batch inference through reusable templates.

Project 2: UPS's Internal Development platform

- Architected and implemented a modular, **event-driven ML platform** on AWS, enabling seamless integration and onboarding for over 15 global data science teams. This platform reduced the average ML workflow setup time by **40%**.
- Developed CI/CD pipelines using **Kubernetes** and **Docker** to package and **deploy models** to an **artifactory**, ensuring robust and efficient model lifecycle management.
- Implemented **model monitoring** and **explainability pipelines**, enhancing transparency and trust in AI models while ensuring compliance with industry standards, resulting in a 25% increase in adoption by key stakeholders

ACADEMIC PROJECTS

REIA - Renewable Energy Integration Assistant

- Developed an intelligent system integrating **Large Language Models (LLMs)** with **Langchain**, enabling advanced real-time data retrieval and processing.
- Built a distributed backend using Langchain and **FastAPI** REST APIs, reducing **response times** to under **200ms** while maintaining scalability and robustness.
- Enhanced retrieval accuracy by integrating **FAISS** with **RAG models**, cutting search times by 50% and sourcing insights from over **1,000 datasets**, including Tesla and SunPower installation guides.

ShieldBot: Real-Time Adversarial Detection for Customer Support Chatbots

- Developed and deployed an AI safety plugin for customer support chatbots, leveraging fine-tuned NLP models like BERT and GPT to detect adversarial inputs.
- Automated the generation of over **1,000 harmful prompts** using GPT-3.5, improving **risk detection** accuracy by 35% and reducing false negatives by 20%.
- Designed a **scalable pipeline** for dynamic **adversarial prompt generation** with 5+ persuasion techniques, boosting dataset creation efficiency by 50%.
- Integrated the plugin as a **REST API**, successfully testing over **500 prompts** in real-time, ensuring robust performance in live environments.