RUTIKA AVINASH KADAM

Stony Brook, NY 11790 · +1 934-949-8653 · rutikakadam2727@gmail.com · LinkedIn · Portfolio · GitHub

EDUCATION

Stony Brook University — **Stony Brook, USA**
Master of Science in Data Science — August 2024 – May 2026
Coursework: Statistics, Machine Learning, Deep Learning, Natural Language Processing, Large Language Models, Big Data.

Savitribai Phule Pune University — **Pune, India**
Bachelor of Engineering in Information Technology — August 2016 – May 2020
Coursework: Discrete Structures, Data Structures & Algorithms, Distribution Systems, Cloud Computing.

SKILLS

Programming Languages: Python, R, SAS, SQL (MySQL, PostgreSQL, PL/SQL), C++, HTML, CSS
Data Science: NumPy, Pandas, Seaborn, scikit-learn, TensorFlow, Keras, PyTorch, Hugging Face, Langchain, Langraph
Data Engineering & Big Data: Azure Data Factory, Data Lake, Databricks, MS Fabric, SSMS, SSIS
MLOps & Deployment Tools: Flask, Streamlit, Gradio, MLflow, FastAPI, Docker, Kubernetes
Visualization & Analytics: Power BI, Tableau, Microsoft Excel
Tools & Platforms: Visual Studio Code, Jupyter Notebook, SAS, RStudio, Postman, Azure, AWS, ServiceNow, Jira, Microsoft Endpoint Configuration Manager, Qualys, Aternity, Git, GitHub

PROFESSIONAL EXPERIENCE

Stony Brook Medicine — **Stony Brook, USA**
Data Scientist (Graduate Researcher) — July 2025 – Present

- Built **missing-data imputation pipelines using multiple imputation** & trained **supervised machine learning models** on imputed datasets to predict **physical function decline & fracture risk** in the **Study of Osteoporotic Fractures cohort**.
- Accelerated interpretation of statistical results by developing a **Retrieval-Augmented Generation (RAG) pipeline** using **OpenAI GPT-4** & embedding models with **LlamaIndex** to translate **SAS/R mixed-effects outputs** into research narratives.

Tata Consultancy Services — **Pune, India**
Data Analyst — August 2020 – April 2024

- Collaborated with Vulnerability Management, performed **exploratory data analysis** on **1 Million+ system vulnerability records** reporting from **Qualys VMDR** using **Python & MySQL**, uncovering trends, anomalies, & threat vectors.
- Reduced **security risk exposure by 15%** by developing **Linear & Logistic Regression models** predicting exploitability & patch prioritization using engineered vulnerability features.
- Discovered recurring remediation patterns across **1M+ records** by applying **semantic clustering using E5 embeddings with HDBSCAN** on remediation text data.
- Fine-tuned **BERT classifier** on cluster-derived labels to predict remediation categories in real time, enabling automated ticket routing, cutting **manual triage by 40%**.
- Improved vulnerability monitoring visibility by designing **Power BI dashboards with DAX-driven KPIs** including **CVSS score, exploitability risk, remediation timelines, & compliance metrics**.
- Achieved 99% patch compliance across enterprise systems by deploying automated remediation workflows through Microsoft Endpoint Configuration Manager & automated tasks using PowerShell, boosting productivity by 25%.

Zensar Technologies — **Pune, India**
Data Intern — May 2019 – July 2019

- Implemented **ETL pipelines in Azure Data Factory with incremental batch loading** to ingest **multi-channel website traffic & operational data** into **Azure Data Lake Gen2** for scalable data integration for **Swiggy-India**.
- Developed **Power BI dashboards** to visualize **conversion funnel metrics** & **marketing channel attribution**, identifying drivers of traffic variability & campaign performance, increasing **high-intent traffic by 15%**.

PROJECT EXPERIENCE

AskYourDocument | Retrieval-Augmented Generation (RAG), Natural Language Processing

- Enabled **semantic document question answering** on ingested documents & web content by building a **RAG system combining FAISS vector search with Google's Generative AI models**.
- Implemented an ingestion pipeline for **PDFs, DOCX, TXT, & URLs using SBERT-based chunking**, integrated with **FastAPI backend & Streamlit interface** to deliver **semantic search & LLM-powered contextual responses**.

Cardiovascular Disease Prediction | Deep Learning, TensorFlow, Keras, Scikit-learn, Streamlit, Docker

- Achieved **81% model accuracy (AUC 0.87)** by developing a **hybrid deep learning architecture combining Artificial Neural Network & Convolution Neural Network** for cardiovascular disease prediction.
- Improved **early cardiac-risk detection with recall of 0.94** & deployed an **interactive Streamlit application on Hugging Face Spaces** for real-time predictions.

Airbnb Price Prediction | Pandas, Numpy, Seaborn, Matplotlib, Scikit-learn, Machine Learning, Bagging & Boosting

- Identified **key pricing drivers across Seattle Airbnb listings** through **exploratory data analysis** of location, room type, amenities, & occupancy factors.
- Improved price prediction performance by benchmarking multiple regression models and achieving best performance with **Random Forest ($R^2 = 0.68$)**.