

RUTIKA AVINASH KADAM

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

EDUCATION

Stony Brook University

Master of Science in Data Science

Stony Brook, USA

August 2024 – May 2026

Coursework: Statistics, Machine Learning, Deep Learning, Natural Language Processing, Large Language Models, Big Data.

Savitribai Phule Pune University

Bachelor of Engineering in Information Technology

Pune, India

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 & ML Libraries: NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, TensorFlow, Keras, PyTorch, Hugging Face

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

Data Scientist(Graduate Researcher)

Stony Brook, USA

July 2025 – Present

- Built missing data imputation pipelines for the Study of Osteoporotic Fractures (women 65+); developed ML models to predict physical function decline & fracture risk across multiple imputed datasets, ensuring robust & reproducible results.
- Developed a Retrieval-Augmented Generation (RAG) pipeline using OpenAI GPT-4 & embedding models with LlamaIndex to convert SAS & R mixed-effects outputs into interpretable epidemiological narratives for public health research.

Tata Consultancy Services

Data Analyst

Pune, India

August 2020 – April 2024

- Collaborated with Vulnerability Management team to perform Exploratory Data Analysis on 1 Million+ system vulnerability records reporting from Qualys VMDR, uncovering trends, anomalies, & threat vectors using Python & MySQL.
- Developed machine learning models (Linear/Logistic Regression) to predict vulnerability exploitability & patch prioritization, leveraging engineered features, reducing security risks by 15% and improving efficiency by 25%.
- Applied unsupervised semantic clustering (E5 + HDBSCAN) on remediation text to identify recurring vulnerability remediation patterns at scale (1 Million+ findings).
- Trained and deployed a supervised BERT classifier on cluster-derived labels to predict remediation categories in real time, enabling bulk deployments and automated ticket assignment, cutting manual triage by 40%.
- Employed Data Analysis Expressions (DAX)-enhanced KPIs like CVSS score, threat intelligence, exploitability, risk levels, remediation timelines, deployment status, compliance rates within Power BI driven Vulnerability Analysis Dashboard.
- Designed & deployed feasible technical solutions using MS Endpoint Configuration Manager to remediate Windows & application vulnerabilities with 99% compliance; optimized configurations & automated tasks using PowerShell, boosting productivity by 25%.

Zensar Technologies

Data Intern

Pune, India

May 2019 – July 2019

- Designed and implemented ETL pipelines using Azure Data Factory to process 30GB of customer feedback and web traffic data in Azure Data Lake Gen2, enabling scalable data integration and analytics for Swiggy, India's leading online food delivery platform.
- Performed sentiment analysis on Swiggy's customer feedback data to identify loyal vs. at-risk users, enabling targeted retention campaigns and improving customer satisfaction insights.

PROJECT EXPERIENCE

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

- Designed RAG application combining FAISS vector search with Google Generative AI, enabling intelligent question answering over ingested documents & web content.
- Implemented an ingestion pipeline for PDFs, DOCX, TXT, & URLs with SBERT-based semantic chunking, integrated into a FastAPI backend & Streamlit frontend to deliver semantic search & LLM-powered contextual responses.

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

- Developed a hybrid deep learning model combining Artificial Neural Network for tabular data & Convolution Neural Network for heart-scan images for cardiovascular disease prediction.
- Achieved 81% accuracy and AUC = 0.87 with high recall (0.94) for high-risk detection.
- Deployed an interactive Streamlit app on Hugging Face Spaces for real-time, explainable predictions.

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

- Performed exploratory data analysis on Seattle Airbnb data to identify key pricing drivers (bedrooms, room type, amenities).
- Built & compared regression models (Linear, Ensemble techniques), achieving highest $R^2 = 0.68$ with Random Forest Regression for price prediction.