

# 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

**Coursework:** Statistics, Data Analysis, Machine Learning, Deep Learning, Natural Language Processing, Big Data Analysis.

Stony Brook, USA

August 2024 – May 2026

### Savitribai Phule Pune University

Bachelor of Engineering in Information Technology

**Coursework:** Discrete Structures, Data Structures & Algorithms, Distribution Systems, Cloud Computing.

Pune, India

August 2016 – May 2020

## SKILLS

**Programming Languages:** Python, R, SQL (MySQL, PostgreSQL, PL/SQL), MongoDB, C, C++, HTML, CSS, JavaScript

**Data Science:** NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, TensorFlow, Keras, PyTorch, HuggingFace

**Data Engineering & Big Data:** Azure Data Factory, Data Lake, Databricks, Microsoft Fabric, Spark, Kafka

**MLOps & Deployment:** Flask, Streamlit, Gradio, Docker, GitHub

**Visualization & Analytics:** Power BI, Tableau, Microsoft Excel

**Cloud & Platforms:** Azure, AWS, Visual Studio Code, Jupyter Notebook, RStudio, SAS, Postman

**Other Tools:** ServiceNow, Jira, Microsoft Endpoint Configuration Manager, Qualys, Aternity

## PROFESSIONAL EXPERIENCE

### Stony Brook Medicine

Graduate Research Assistant

Stony Brook, USA

July 2025 – Present

- Built end-to-end data imputation and machine learning pipelines in R & SAS for Study of Osteoporotic Fractures cohort (women 65+), to predict physical function decline & fracture risk with cross-validation & hyperparameter tuning across imputed datasets.
- Engineering an AI-assisted scientific reporting pipeline leveraging Retrieval-Augmented Generation (RAG) & Large Language Models (LLMs) to automatically interpret & generate domain-specific narratives from SAS & R mixed-model outputs.

### Tata Consultancy Services

Data Analyst

Pune, India

August 2020 – April 2024

- Collaborated with Vulnerability Management team to perform Risk Analysis on 1 Million+ vulnerability records reporting from 55 Thousand+ Windows assets in Qualys VMDR, uncovering trends, anomalies, & threat vectors using Python & MySQL.
- Developed predictive models (Logistic Regression, Boosting, ANN) to predict vulnerability exploitability & patch prioritization, leveraging engineered features, reducing security risks by 15% and improving efficiency by 25%.
- Built a Databricks-based NLP pipeline that clustered 1M+ vulnerability findings by remediation using E5 embeddings + HDBSCAN, cutting manual triage by 60%.
- Fine-tuned BERT & deployed an ML classifier to predict solution buckets (macro-F1 0.87), enabling auto-routing to patch teams.
- 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.
- Built NLP-based sentiment analysis models 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 dual deep learning model combining Artificial Neural Network (ANN) for tabular data & Convolution Neural Network (CNN) 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, Ridge, Lasso, Random Forest, XGBoost, CatBoost), achieving  $R^2 = 0.68$  with Random Forest Regression for price prediction.
- Derived insights showing that entire homes, higher bedroom counts, and amenities like gym, elevator, and pool significantly increase listing prices.