

Aishwarya Bhanage

✉ aishwaryamahad.bhanage@stonybrook.edu 📞 +19342460445 📍 Stony Brook, New York 🔗 LinkedIn 📁 Portfolio 🏠 GitHub

Seeking full time opportunities starting May 2026

Education

Stony Brook University

Masters of Science in Data Science

Stony Brook, NY

Aug 2024- May 2026

Courses: Statistics, NLP, Vision Language Models, Data Analysis, Probability, Big Data Systems, Cloud Computing, HPC

Sinhgad Institute of Technology and Science

Pune, MH

Bachelor of Engineering in Computer Engineering

May 2016- May 2020

Courses: Data Structures & Algorithms, RDBMS, OOPS, Machine Learning, Data Mining and Warehousing, Statistics

Experience

Veatch & Company

Founding Machine Learning Engineer

New York City, NY

Jul 2025- Present

- Built semantic retrieval pipelines leveraging **vector search (FAISS, OpenAI embeddings) and GPT-4 & Hugging Face Transformers** reranking, improving relevance by **15%**.
- Built and deployed a **Retrieval-Augmented Generation (RAG)** chatbot using **FAISS vector search and OpenAI's text-embedding-3-small**, automating compliance document queries and reducing manual review time by **10%**.
- Fine-tuned **Falcon-7B, Mistral-7B with LoRA adapters** for domain-specific question answering, reducing hallucinations by **22%**.

Accenture Solutions Pvt Ltd

Data Scientist (Banking Analytics)

Pune, MH

Jun 2022 - Oct 2023

- Orchestrated **ETL** and model retraining workflows using **Databricks Workflows and Delta Live Tables (DLT)** to automate ingestion, transformation, and validation of user-session and transaction data from the Banking application.
- Cleaned, transformed, and integrated **1M+ daily** web app sessions using **PySpark and Databricks SQL**, improving data quality and enabling near real-time compliance analytics.
- Implemented predictive **anomaly-detection models (XGBoost, Logistic Regression) with 90% AUC**, and optimized **feature extraction** from behavioral data streams deployed in Kubernetes environments for scalable and reliable execution.
- Designed and evaluated **A/B testing frameworks using t-tests and regression analysis** to compare performance of app feature rollouts, ensuring statistically valid experiment outcomes that accelerated deployment by **20%**.
- Visualized KPIs** and user journey metrics in **Tableau(Certified Desktop Specialist) & Power BI**, providing actionable insights to risk and product teams that improved uptime by **15%** and enhanced user retention analytics.

Accenture Solutions Pvt Ltd

Associate Data Engineer

Pune, MH

Dec 2020 - May 2022

- Optimized **SQL queries using joins, CTEs, & Windows functions** reducing query latency by **35%** for key metrics extraction for dashboards and reports.
- Created real-time **anomaly-monitoring dashboards in Tableau and AppDynamics**, shortening incident-response time by **40%** and enabling proactive system-health tracking.
- Partnered with **product and engineering teams** via **Jira and ServiceNow** to translate business requirements into technical data models and drive production-ready BI workflows.

Leadership

Stony Brook University

Graduate Student Organization Treasurer

Stony Brook, NY

May 2025 - Present

- Managed transparent financial operations for **8,000+ graduate students**, compliantly managing five reimbursement lines and led effective budget meetings resulting in a 98% budget approval to provide financial security.

Skills

Functional Competencies: Machine Learning & NLP | Generative AI (LLMs, Transformers, RAG) | Semantic Search & Ranking Models | Embeddings & Vector Databases (FAISS, Pinecone) | A/B Testing & Statistical Analysis | Predictive Modeling | ETL & Data Orchestration (Databricks Workflows, Delta Live Tables) | Clip, Blip, SigLIP, Qwen, Global & Local Prompting | Data Visualization (Tableau, Power BI) | Cross-Functional Collaboration.

Technical Tools: Python (Pandas, Scikit-learn, PyTorch, Hugging Face) | SQL | PySpark | Databricks | TensorFlow | AWS (S3, SageMaker) | Azure | Docker | Git | Snowflake | Kubernetes | Docker | Streamlit.

Projects

Mythology-Based Story Generator & Narrator

- Implemented a retrieval-augmented generation system which used Falcon-7B and E5-base embeddings (**PyTorch**) to ingest over 5M tokens while achieving 93% accuracy and less than 3% hallucination across 20+ test conditions.

Online retail product recommendation Streamlit application

- Engineered a user-friendly Streamlit web app employing the Apriori algorithm to analyze retail data, identifying key product associations and recommending top 5 products that streamline the shopping experience for users.