

PRITESH SHAH

priteshshahwork@gmail.com | 317-457-9324 | Indianapolis, IN | [linkedin.com/in/pritesh96](https://www.linkedin.com/in/pritesh96) | github.com/priteshshah96

EDUCATION

Indiana University, Luddy School of Informatics, Computing, and Engineering, Indianapolis, IN
Master of Science in Applied Data Science

Expected Graduation: May 2025

University of Mumbai, Mumbai, India
Bachelor of Engineering in Computer Science Engineering

May 2017

TECHNICAL SKILLS

Programming Languages: Python (PyTorch), R, SQL, JavaScript

Web Tech & Frameworks: Django, flask, HTML/CSS, React

Databases: MySQL, MongoDB

Tools & Cloud: Power BI, Git, AWS

DATA PROJECTS

Annotation Tool for Scientific Data Labeling

Fall 2024

- Developed a web application using Node.js and React for annotating scientific data in a lab setting.
- Deployed the tool on Vercel with a MongoDB backend, enabling efficient storage and retrieval of annotations.
- Created a user-friendly interface tailored to streamline the annotation workflow, facilitating collaboration among lab researchers and improving data quality.

Financial 10K RAG

Spring 2024

- Analyzed five-year 10-K financial filings of top 10 U.S. technology companies using META's Llama-3 model, achieving 80% accuracy with enhanced inference speeds; employed LlamaIndex and Langchain framework for the RAG application.
- Developed a user-friendly frontend with Streamlit for Question-Answering from regulatory filings.

Med-Llama

Spring 2024

- Utilized the Qlora technique to accelerate fine-tuning of the META's Llama-3 model by 50% on the MEDalpaca dataset and developed a RAG application focused on medical QA generation, achieving 60% accuracy and enhanced accuracy and inference speeds using Query decomposition.
- Built the application using BGE large embeddings and stored the embeddings vector database, leveraging the LlamaIndex framework to ensure effective data management and quicker processing capabilities.

Haiku Generator and Analyzer

Fall 2023

- Engineered an AI-powered haiku generator using LLaMa 2 and GPT-J 6B, trained on a Kaggle dataset of traditional haiku. Implemented Google Colab with FastPunct for text normalization and GRUEN for quality assessment. Developed a Flask-based web interface with ngrok, providing features like sentiment analysis and word cloud visualization for each AI-generated haiku.
- Optimized the model for efficient haiku generation while maintaining the classic 5-7-5 syllable structure, integrating advanced syllable counting techniques to address complex linguistic elements like initialisms and acronyms.

RELEVANT WORK EXPERIENCE

Natural Language Processing Graduate Research Assistant

August 2023 – Present

Indiana University, Luddy School of Informatics, Computing, and Engineering

Indianapolis, IN

- Enhanced cultural relevance in language models by contributing to annotation processes and integrating cultural factors in NLP tasks.
- Developed automated pipelines using Pytorch, GROBID, and CERMINE for scientific text parsing, significantly improving data extraction and processing efficiency.
- Constructed a knowledge graph of conference papers to analyze organization and author affiliations, providing valuable insights into research networks and collaboration trends.
- Collaborated on refining entity extraction and document parsing accuracy, utilizing advanced NLP models and techniques tailored for academic research applications.

Data Analyst

October 2022 – July 2023

Malom Clothing

Mumbai, India

- Designed and developed a comprehensive HR management dashboard using Power BI to track employee shift assignments, attendance, and performance metrics.
- Employed advanced forecasting machine learning algorithms such as Random Forest and XGBoost using historical data to predict future inventory requirements.
- Executed A/B tests for marketing campaigns and website enhancements, utilizing tools like Google Analytics.
- Analyzed user interactions to optimize designs, offers, and user experiences, resulting in improved customer engagement.

Trainee Data Engineer

June 2021 – October 2022

Niku Associates

Mumbai, India

- Utilized Microsoft Power BI to analyze and manage data, contributing to the development of an employee management software and dashboard for optimizing shift duties and raw material procurement.
- Created Python scripts to automate database tasks, improving database management efficiency by 20%.
- Utilized advanced data analysis techniques, including time series analysis and simulation modelling, to optimize supply chain management, resulting in a remarkable 10% enhancement in overall supply chain efficiency.