Venkata Harshitha Bathala

+1 (682)-406-1948 venkatasfharshitha555@gmail.com LinkedIn GitHub Portfolio

Education

UNIVERSITY OF TEXAS AT ARLINGTON **Jan 2024 – May 2025**
Master of Sciences in Computer Science
GPA: 3.9/4.0

Coursework: Data Mining, Web Development, Artificial Intelligence, Machine Learning, Computer Vision, Advance Software Engineering, Distributed Systems

Experience

University of Texas at Arlington | Arlington, TX *May 2025 – Present*
Volunteer Research Assistant

Project 1: *Capture the Flag (CTF) Infrastructure Deployment*

- Migrated and redeployed the Capture the Flag (CTF) environment by setting up the CTFd platform, configuring Nginx and MariaDB, and hosting web-based challenges securely on a new system.
- Installed and configured essential components including Docker, PHP, Socat, and MariaDB to support a wide range of CTF challenges, and documented the complete deployment process for future scalability and maintenance.

University of Texas at Arlington | Arlington, TX *Jul 2024 – May 2025*
Graduate Research Assistant

- I am responsible for both **development** and **Administration Tasks**, prioritizing and executing them based on the specific requirements and needs of the department.
- Designed and developed a comprehensive **web portal** for UTA's CSE Faculty utilizing **HTML5**, **CSS3**, **JavaScript**, and **PHP**, significantly enhancing resource accessibility and departmental efficiency.
- Implemented complete **LAMP stack architecture** with **Apache** web server configuration, **MySQL database** design, and optimized **PHP back-end** to support secure authentication and dynamic content delivery.
- Executed **front-end development** focusing on responsive design principles and intuitive **user interface** elements that simplified navigation and improved faculty adoption rates.
- Consolidated essential information from the UTA Main Portal into a centralized **web application**, including inventory details, networking and security policies, supported operating systems and tools, completed projects, important forms, and faculty request submissions.
- As a system administrator, contributed to the setup and configuration of lab systems for the CSE department
- Developed secure **user authentication** mechanisms and role-based access control systems to maintain strict departmental information security while enabling streamlined workflow processes.
- Implemented **LUKS encryption** and advanced **Linux security** protocols across server environments to ensure comprehensive protection of sensitive web application data and user information.
- Implemented automated **Bash scripts** for system maintenance and security, including user management frameworks that streamlined account provisioning and enforcing granular **access control**within lab environments.
- Optimized **virtualization resources**, such as VirtualBox, to ensure students receive enhanced computing capabilities for coursework.
- Monitored and maintained **UTA's two servers**, specifically tracking the percentage increase in resource usage to ensure system health and maintain uptime.
- Operating systems worked on include **Debian-based** (Ubuntu) and **RPM-based** (CentOS, Red Hat, Fedora) systems.

DXC Technology | Chennai, India *Jan 2023 - Dec 2023*
Associate Professional Software Engineer

- Served as a key technical contributor across **Lean Process Engineering** and **Data Science** teams, implementing integrated solutions combining **process optimization** methodologies with **advanced analytics** frameworks.
- Engineered **Python-based automation solutions** utilizing **Pandas**, **NumPy**, and custom libraries, achieving measured 5% **FTE savings** and quantifiable improvements in **operational efficiency** metrics across business units.
- Developed sophisticated **Value Stream Mapping (VSM)** analytics using **Python data visualization** libraries to identify process bottlenecks, resulting in streamlined workflows and reduced cycle times.
- Led cross-functional **automation initiatives** leveraging **JIRA**, **UCML**, and **STA** tools, while implementing custom **Python scripts** for workflow integration, significantly enhancing process reliability and operational metrics.
- Designed and implemented **machine learning pipelines** using **scikit-learn** and **TensorFlow**, creating predictive models that improved forecast accuracy by analyzing historical operational data patterns.
- Collaborated directly with enterprise clients in automotive and luxury sectors, developing interactive **Power BI** and **Tableau dashboards** integrated with **Python-generated datasets** to drive strategic decision-making processes.

- Architected end-to-end **ETL processes** using **Python** to transform raw operational data into structured formats suitable for **machine learning model** training and business intelligence applications.
- Implemented **statistical analysis** and **time-series forecasting models** using **Python's statsmodels** and custom algorithms to optimize resource allocation and improve operational planning accuracy.
- Leveraged **NLP techniques** with **NLTK** and **spaCy** libraries to analyze customer feedback data, extracting actionable insights that informed product development and service improvement initiatives.

Projects

Smart HealthCare Hub

Git Repository

- Architected and developed a comprehensive **full-stack healthcare platform** with multi-user functionality serving diverse stakeholders including patients, medical providers, administrators, and pharmacists, featuring **secure authentication**, **appointment scheduling**, **electronic health records**, **symptom assessment**, and **real-time messaging**.
- Engineered responsive **front-end interface** utilizing **React**, **HTML5**, **CSS3**, and **JavaScript** with mobile-first design principles, while implementing robust **back-end architecture** with **PHP** and **MySQL**, integrating **Axios** for optimized **RESTful API** communication.
- Implemented comprehensive **data security protocols** including **encryption**, **secure authentication mechanisms**, and **role-based access control** systems to ensure strict compliance with healthcare **data protection standards** and regulatory requirements.
- Designed scalable **database schema** optimized for healthcare workflows, incorporating **normalized data structures** and efficient **query optimization** techniques to maintain performance under high transaction volumes.
- Crafted elegant, intuitive **user interfaces** with fluid **DOM manipulation** and sophisticated **asynchronous JavaScript** techniques, delivering exceptional platform usability while achieving sub-second page load times and seamless real-time data updates across all healthcare workflows.

Sentiment Analysis on Song Lyrics

Git Repository

- Engineered a comprehensive **natural language processing** solution for **sentiment analysis** on imbalanced Spotify song lyrics dataset, implementing advanced **NLP preprocessing** pipeline including **tokenization**, **stop-word removal**, and **lemmatization** with **NLTK**.
- Implemented sophisticated **feature extraction** techniques utilizing **Word2Vec** embeddings with **Gensim**, while addressing dataset challenges through **stratified sampling** methodologies and **data augmentation** strategies using the **nlpaug** library.
- Architected and optimized a **bidirectional LSTM** neural network with systematic **hyperparameter tuning**, achieving 75% classification accuracy, then enhanced model performance by implementing **BERT** transformer architecture via **Hugging Face Transformers** library, reaching 77% accuracy.
- Conducted rigorous **model evaluation** using **confusion matrices** and performance metrics with **Scikit-learn**, delivering a production-ready **real-time prediction system** capable of effectively classifying emotional content in lyrical text data.
- Demonstrated expertise in **deep learning** frameworks, **text classification**, **transfer learning**, and **machine learning evaluation** methodologies throughout the full project lifecycle from data preprocessing to model deployment.

Assistive Visual Question Answering (VQA) System

Git Repository

- **Developed an AI-driven Visual Question Answering system** to assist visually impaired users by generating natural language answers to real-world image-based questions.
- Built two complementary pipelines:
  - Classification-based model using **PaliGemma-2** embeddings with a custom **MLP classifier**, achieved a high **Top-1 test accuracy of 71.7% and Top-5 accuracy of 98.4%**, demonstrating robustness against noisy inputs and unanswerable queries.
  - Generative model using **Vision Transformers (CLIP, SigLIP, ViT)** for image encoding and **GPT-2** for text decoding, enabling open-ended natural language generation with **56% semantic accuracy**.
- **Applied transfer learning, multimodal embedding alignment**, and **autoregressive decoding** to enhance contextual understanding and open-ended answer generation capabilities.
- **Preprocessed the VizWiz dataset** (31,000 image-question pairs), performed data cleaning, managed annotation noise, and implemented an **80-10-10 train-validation-test split**.
- **Integrated LLMs and NLP pipelines** for semantic alignment between image embeddings and question prompts, boosting image-question comprehension and accuracy.

Protocols for Fault Tolerance in Distributed Systems

Git Repository

- Implemented the **SWIM protocol** for scalable **failure detection** using **gRPC** and **Python**, enabling efficient node monitoring and **membership management** across **Dockerized containers** within a microservices architecture..
- Engineered a **Two-Phase Commit (2PC)** system with distinct vote and decision phases, leveraging **Protocol Buffers** for cross-language **serialization** and robust **transaction coordination**.
- Developed a streamlined **Raft consensus** implementation featuring **leader election** and **log replication**, utilizing **Docker** multi-container orchestration to ensure consistent state management.
- Created comprehensive **fault injection testing** framework in **Python, NodeJS** to simulate network partitions and node failures, validating protocol resilience under adverse conditions.
- Implemented **asynchronous communication** patterns and **connection pooling** techniques across all distributed protocols, significantly improving system throughput and reducing latency.

Technical Skills

Operating Systems: Linux, Unix, Windows

Web Development: HTML, CSS, JavaScript, React, Nodejs, Angular, PHP

Programming Languages: C, Java, Python

Data Analysis & Visualization: Power BI, Excel, Python, Tableau

Cloud Platforms: Microsoft Azure Administrator, AWS

DevOps: Docker, Kubernetes, CI/CD (Git, Ansible)

Databases: MYSQL, MongoDB, MariaDB

AI/ML Models: BERT, Bidirectional LSTM (BiLSTM), PaliGemma-2, SigLIP, Vision Transformer (ViT), CLIP, GPT-2

Libraries & Frameworks: TensorFlow, Keras, PyTorch, Scikit-learn, NLTK, Pandas, NumPy, Matplotlib, Seaborn

Shell Scripting: Bash

Tools: JIRA, UCML, STA, Apache, Tomcat

Professional Certifications

AWS Certified Cloud Practitioner

Microsoft Certified: Azure Fundamentals (AZ-900)

AIGPE Lean Six Sigma Yellow Belt Certification