Vankayal Megha Shree

St Louis, MO | (314) 910-0455 | meghavankayal@gmail.com | LinkedIn

EDUCATION

State University of New York, Binghamton

Master of Science in Computer Science *May 2025*

Relevant Coursework: Data Structures and Algorithms, Machine Learning, Operating System, Database Management Systems, Cloud Computing, Programming Languages, Data Mining, Discrete Mathematics, Object Oriented Programming, Design Patterns

TECHNICAL SKILLS

Programming Languages & Tools: Python, Java, C++, R, Shell Scripting (Bash), Git, Jenkins, MATLAB

Databases & Web Development: MySQL, SQL, Neo4j, MongoDB, SQL Alchemy, Flask, RESTful APIs, API Development

Cloud & DevOps: AWS (Sagemaker), GCP, Docker, Kubernetes, Terraform, Apache NiFi, CI/CD Pipelines, JIRA, Hadoop

Big Data & Data Engineering: Apache Spark, PySpark, ETL Pipelines, Data Ingestion, Data Aggregation, Data Modeling

Machine Learning & AI: scikit-learn, XGBoost, TensorFlow, PyTorch, Keras, OpenCV, Supervised Learning, Unsupervised Learning, Reinforcement Learning, Deep Learning, Natural Language Processing (NLP), Generative AI

Advanced AI & LLMs: Retrieval-Augmented Generation (RAG), Transformers (BERT), Hugging Face, LangChain, FAISS

Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Tableau, Power BI, Exploratory Data Analysis (EDA), Statistical Modeling, Classification, Clustering, Anomaly Detection, A/B Testing

Certification: AWS Cloud Practitioner Certification

PROFESSIONAL EXPERIENCE

AI & Data Analyst | AttainX, Inc. | |Herndon, VA May 2024 – August 2024

- Developed a **Knowledge Graph**-powered Intelligent Document Processing system using **Neo4j**, **Python**, **PySpark**, **Pandas**, **SQL Alchemy**, **Hugging Face Transformers**, and **LLMs** with **RAG**, processing text from diverse sources (PDFs, logs, surveys) in varied formats to improve query accuracy by 25%.
- Built an automated **ETL** pipeline with **PyPDF2**, **NLTK**, **spaCy**, **Regex**, **LangChain**, **FAISS**, and **SentenceTransformers** for extracting and normalizing unstructured text, enabling **similarity search** and **exploratory data analysis (EDA)** with 30% faster response time.
- Designed scalable **data pipelines** using **Apache NiFi**, **PySpark**, **Matplotlib**, and **Seaborn** for **data visualization**, analyzing telemetry and survey data in diverse formats with **Pandas**, **NumPy**, **scikit-learn**, and **Hugging Face** for **NLP** and **statistical modeling**.

Data Analyst | Safe Code Group LLC | Hyderabad, India *March 2023 – May 2024*

- Engineered **data pipelines** with **SQL**, **Python**, **Pandas**, and **NumPy** for **ETL**, streamlining text from varied sources (databases, logs, **APIs**) and formats, improving **data quality** by 10% through **data wrangling**.
- Developed automated reporting systems using **Python**, **SQL**, **Matplotlib**, and **Seaborn** for **data visualization** dashboards, enabling **data-driven decisions** with text from diverse sources, reducing report generation time by 40%.
- Implemented **machine learning** models with **TensorFlow** and **scikit-learn** for **classification** and **predictive modeling**, analyzing text in multiple formats to enhance customer behavior prediction by 20% via **exploratory data analysis (EDA)**.

RESEARCH EXPERIENCE

Analysis of LinkedIn Job Postings Data | Binghamton University, NY *Aug 2024 – Dec 2024*

- Designed a **data-driven** recommendation system using **Python**, **MongoDB**, **Pandas**, and **NumPy** to match candidate profiles with job requirements, analyzing text from LinkedIn job postings in varied formats for **exploratory data analysis (EDA)**.
- Visualized job market trends (hiring patterns, salaries, roles, regional demand) with **Matplotlib** and **Seaborn** for **data visualization**, processing diverse text sources to deliver actionable insights for job seekers and employers.
- Derived insights into salary trends and industry dynamics using **Pandas**, **NumPy**, and **statistical modeling**, optimizing job search strategies by structuring text data from multiple formats in **MongoDB**.

Training ML Agents in UNITY to Detect Explosives using AI | Vellore Institute of Technology, India *Nov 2022 – May 2023*

- Developed a **machine learning** model in **Unity** using **Python**, **TensorFlow**, and **PyTorch** to detect explosives, processing text and sensor data from simulated 3D environments in various formats with the **ML Agents toolkit**.
- Trained AI agents with **Proximal Policy Optimization (PPO)** and **Soft Actor-Critic (SAC)** algorithms via **reinforcement learning**, reducing false positives in explosive detection by analyzing diverse data sources.
- Enhanced detection accuracy by fine-tuning **machine learning** models with **TensorFlow** and **PyTorch** for **classification**, integrating **AI solutions** in **Unity** to handle varied data formats like simulation logs.

Pattern Classification Framework | Vellore Institute of Technology, Chennai, India *Jun 2022 – Dec 2022*

- Developed a novel **Probabilistic Neural Network (PNN)** framework using **Python** and **TensorFlow** for **pattern classification**, improving accuracy over **CRISP (**Cross-industry standard process**)** methods by processing text and feature data from diverse datasets.
- Reduced misclassification rates in imbalanced and noisy datasets using **Pandas**, **NumPy**, and **scikit-learn** for **exploratory data analysis (EDA)** and **statistical modeling**, handling various data formats.
- Evaluated **machine learning** models with **TensorFlow** and **scikit-learn**, defining soft decision-making criteria to optimize **classification** performance across text and numerical data sources.