

THEJESH MALLIDI

+1 517-730-4554 | mallidit@msu.edu
github.com/thejesh-m | [linkedin.com/in/thejeshm](https://www.linkedin.com/in/thejeshm) | www.thejesh.in | East Lansing, MI

EDUCATION

Michigan State University, USA	08/2023 - 04/2025
Master of Science in Data Science	GPA: 3.95/4.0
Madanapalle Institute of Technology & Science, India	06/2017- 07/2021
Bachelor of Technology in Computer Science	GPA: 9.23/10.0

EXPERIENCE

Michigan State University	East Lansing, MI, USA
Graduate Research Assistant	01/2024 – Present
<ul style="list-style-type: none">Built a domain-specific knowledge graph containing over 2 million entities and approximately 1.4 million relationships as part of collaborative research under Professor Shinhan Shiu, leveraging advanced NLP techniques such as Named Entity Recognition (NER) and Relation Extraction to streamline real-world knowledge management and data processing solutions.Leveraged state-of-the-art Large Language Models (LLMs) and engineered a custom BERT-based multi-relation extraction model, achieving 90% accuracy in entity extraction and 75% accuracy in relation identification, to accurately identify and extract complex relationships from textual data.Integrated Retrieval-Augmented Generation (RAG) techniques with LLMs for natural language querying of the knowledge graph, enabling efficient data retrieval from a dataset of 400,000+ records.	
Quantiphi Analytics	Bangalore, India
Machine Learning Engineer	05/2021 - 07/2023
<ul style="list-style-type: none">Designed and implemented Automation Intelligent Document Processing by combining Computer Vision and NLP-based LLMs for the BFSI domain, enabling streamlined data processing by extracting relevant information from unstructured and semi-structured documents, thereby reducing manual effort and improving data accuracy and processing speed.Delivered solutions for Document digitization and extraction, Document Classification, Object Detection, Image Classification, and Semantic Segmentation tasks across 10+ Proof of Concepts (PoCs).Developed predictive models using Linear and Logistic Regression, Decision Trees, Bagging, Random Forest, and Gradient Boosting Machines (GBM) to address complex business challenges.Built and optimized inference pipelines for Nvidia Edge devices using Nvidia DeepStream, enabling real-time processing with inference speeds of 27 frames per second (FPS), significantly enhancing performance for time-sensitive applications.Deployed and scaled 20+ machine learning models in production using AWS SageMaker, GCP Kubernetes, Flask, REST APIs, and other deployment tools.	

SKILLS

- Core Expertise:** Data Science, Machine Learning, Deep Learning, Natural Language Processing (NLP), Computer Vision, Predictive Modeling, Decision Analytics, Graph ML, Knowledge Graph & Large Language Models (LLMs), Generative AI
- Programming Languages:** Python, R, C, C++, Bash
- AI/ML Libraries & Frameworks:** Keras, TensorFlow, PyTorch, TensorRT, Hugging Face Transformers, LangChain, LangGraph, LlamaIndex, VectorStores and Retrievers, LLMops (MLFlow), NLTK, SpaCy, NumPy, Pandas, OpenCV, Matplotlib, Seaborn
- Big Data & Streaming:** Hadoop Spark, Apache Kafka
- Databases:** MySQL, PostgreSQL, MongoDB
- Development & Deployment Tools:** Flask, REST API, Git, Docker, AWS, GCP, Heroku, Nvidia Deepstream & TLT

PROJECTS

Analysis and Quantification of Deflected Regions in Road Surfaces	09/2023 - 12/2023
<ul style="list-style-type: none">Developed a comprehensive methodology to identify and quantify deflected or damaged regions on road surfaces using Point Cloud data, achieving a 95% accuracy in defect detection and reducing inference time by 60% through parallel processing and Nvidia Rapids.Utilized technologies such as Python, VTK, Clustering techniques and Deep Learning models (PointNet++) for 3D data analysis and defect detection.	
Image Captioning Using Attention Models	01/2021 - 05/2021
<ul style="list-style-type: none">Implemented an image captioning system using the Flickr 8k dataset, leveraging pre-trained models like VGG16 and ResNet for feature extraction and deep learning architectures (CNNs and LSTMs) to generate captions for over 8,000 images.Enhanced model performance and achieved an 80% BLEU score by integrating advanced NLP-based attention mechanisms, enabling precise contextual representation and improved captioning quality.	

CERTIFICATIONS

- AWS Machine Learning Specialty Certification
- Data Science Specialization