

# Janvita Reddy

979-709-1497 | [reddyjanvita11@gmail.com](mailto:reddyjanvita11@gmail.com) | [linkedin.com/in/janvita/](https://www.linkedin.com/in/janvita/) | [github.com/JanvitaReddy11](https://github.com/JanvitaReddy11)

## EDUCATION

**Texas A&M University**, College Station, TX, USA

*Aug'23 – May'25*

MS in Data Science, GPA: 4.0/4.0

**Sardar Vallabhbhai National Institute of Technology**, Surat, India

*Aug'19 – May'23*

B.Tech in Mechanical Engineering, GPA: 9.21/10.0

## SKILLS

**Programming Languages:** C++, Python, R, SQL, MATLAB

**ML Libraries:** NumPy, Matplotlib, Pandas, Pytorch, Keras, OpenCV, TensorFlow, Scikit, Seaborn, NLTK

**Models:** Regression, SVM, KNN, Decision Trees, Neural Networks, Transformers, Autoencoders, GANs, LLMs

**Cloud Products:** GCP Compute Engine, Cloud storage, Vertex AI, BigQuery, AWS EC2, AWS S3, SageMaker

**Data Technoogy:** Hadoop, PySpark, Docker, Airflow, Git, MongoDB, Tableau, PowerBI, Lookup, Excel

## RELEVANT EXPERIENCE

**Graduate Research Assistant — Texas A&M University, USA**

*Jan'24 - Present*

- Annotated 1800+ images using **LabelMe** software integrated with **SAM** to identify cotton balls. Engineered segmentation models like **Unet attention**, **CBAM**, **Pix2Pix Unet**, **SwinUnet** achieving 0.70 IoU score.
- Enhanced model performance by implementing **LoRA** and skip connections to **finetune** SAM. Conducted **correlation** analysis between pixels and cotton yield, achieving 0.91 correlation accuracy.

**Graduate Research Assistant — Texas A&M University, USA**

*Sep'23 - Dec'23*

- Experimented with CNN models like **AlexNet**, **VGG-16**, **ResNet**, **EfficientNet**, **3D CNN** for classification of stress induced crops and conducted hyperparameter tuning using **keras tuner**.
- Implemented **LSTM** model on **time series** images of cotton crops to capture **spatio-temporal** relations, resulting in 6% accuracy improvement on test data, with F1 score of 0.97.

**Research Intern — IISER Bhopal , India**

*May'22 - July' 22*

- Designed a **domain generalisation** model to improve resilience across multiple domains. Utilized **adversarial learning** for model adaptation, integrated with **Grad-CAM** for better interpretability.
- Applied **K-means clustering** to segregate domains, and AlexNet as feature extractor. Achieved 3% increase in accuracy over the baseline model by incorporating **multi-domain discriminators**.

**Research Intern — IIT Kharagpur, India**

*May'21 - Feb'22*

- Automated the detection of weld path by developing a machine learning algorithm for robotic welding.
- Implemented **YOLOv5** to detect joints, attained precision of 99.5%. Leveraged **image processing** techniques to denoise and identify edges, yielding an absolute error of  $\pm 1\text{mm}$  for weld lines and  $\pm 0.1\text{mm}$  for gaps.

## PROJECTS

**AI Assistant**

- Developed a personal AI assistant that automates tasks like sending emails, scheduling meetings, performing internet searches, and answering queries from PDFs. Leveraged **RAG** for multi-document querying and context aware responses.
- Integrated specialized agents into a unified master agent using **ReAct** framework for seamless task execution.

**Reproducing ChatGPT**

- Trained GPT-2 124M model from scratch on FineWeb dataset, revamped HellaSwag accuracy from 0.31 to 0.33.
- Enhancements included replacing LayerNorm with **RMSNorm**, incorporating **Rotary Positional Encodings**, and implementing **Group Query Attention** to optimize multi-head attention.

**Job Recommendation System**

- Offered personalized **job recommendations** by analyzing users work history. Conducted **text preprocessing** to transform job description and user profile to **word2vec embeddings** and **TF-IDF** vector.
- Applied user collaborative filtering, modeled a neural ranker securing 89% accuracy and hit rate@20 of 0.78.

**Wide Residual Attention Networks**

- Enhanced ResNet model by adding **mask attention** after residual blocks. Applied **normalization**, **random flipping**, and **Gaussian noise** to training images, achieving 95.4% accuracy on the CIFAR-10 dataset.

**Chicken Disease Classification using MLOps**

- Designed scalable ML solution by integrating pre-trained **VGG16** with custom layers. Executed data ingestion, training, and evaluation using **DVC** for pipeline tracking.
- Deployed model using **Docker** on **AWS EC2**, leveraging GitHub Actions for automated **CI/CD** pipelines

**Fraud Classification using GCP**

- Collected and stored employee data in **GCS bucket**. Transformed data, masked sensitive information, using **Cloud Data Fusion**. Created dashboards using **Looker** and automated **ETL** process with **Airflow**.