# Janvita Reddy

979-709-1497 | reddyjanvita11@gmail.com | linkedin.com/in/janvita/ | github.com/JanvitaReddy11

## EDUCATION

**Texas A&M University**, College Station, TX, USA  *Aug 23 - May 25*
MS in Data Science, Data Science,*CGPA: 4.0/4.0*

**Sardar Vallabhbhai National Institute of Technology**, Surat, India  *Aug 19 - May 23*
B.Tech in Mechanical Engineering, Mechanical Engineering,*CGPA: 9.21/10.0*

## SKILLS

**Programming languages**: Python, R, SQL
**Ml libraries**: NumPy, Pandas, Pytorch, Keras, OpenCV, TensorFlow, Scikit
**Models**: Regression, SVM, KNN, Decision Trees, Neural Networks, Transformers
**Cloud products**: GCP Compute Engine, Cloud storage, Vertex AI, BigQuery, AWS EC2, AWS S3
**Data technology**: Hadoop, PySpark, Docker, Airflow, Git, MongoDB, Tableau, PowerBI
**Analytical skills**: Data preprocessing, Feature engineering, Hyperparameter tuning, Model evaluation
**Transferable skills**: Automation, Data ingestion, Pipeline tracking, Deployment, CI/CD pipelines

## WORK 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, enhancing model performance by 0.70 IoU score.
- Engineered segmentation models like Unet attention, CBAM, Pix2Pix Unet, and SwinUnet, achieving a 0.70 IoU score and improving model performance by implementing LoRA and skip connections to finetune SAM.
- Conducted correlation analysis between pixels and cotton yield, achieving a 0.91 correlation accuracy and identifying key factors influencing crop yield.

**Graduate Research Assistant — Texas A&M University**, USA  *Sep 23 - Dec 23*
- Experimented with CNN models like AlexNet, VGG-16, ResNet, EfficientNet, and 3D CNN for classification of stress-induced crops, conducting hyperparameter tuning using Keras Tuner and achieving a 6% accuracy improvement on test data with an F1 score of 0.97.
- Implemented LSTM model on time series images of cotton crops to capture spatio-temporal relations, resulting in a 6% accuracy improvement on test data with an F1 score of 0.97.

**Research Intern — IISER Bhopal**, India  *May 22 - July 22*
- Designed a domain generalization model to improve resilience across multiple domains, utilizing adversarial learning for model adaptation and integrating Grad-CAM for better interpretability.
- Applied K-means clustering to segregate domains and utilized AlexNet as a feature extractor, achieving a 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, implementing YOLOv5 to detect joints with a precision of 99.5%.
- Leveraged image processing techniques to denoise and identify edges, yielding an absolute error of Â±1mm for weld lines and Â±0.1mm 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, leveraging RAG for multi-document querying and context-aware responses.
- Integrated specialized agents into a unified master agent using ReAct framework for seamless task execution, ensuring efficient and accurate responses.

**Reproducing ChatGPT**
- Trained GPT-2 124M model from scratch on FineWeb dataset, achieving a 6.5% improvement in HellaSwag accuracy from 0.31 to 0.33 through enhancements including replacing LayerNorm with RMSNorm and incorporating Rotary Positional Encodings.

- Optimized multi-head attention by implementing Group Query Attention, resulting in enhanced model performance.

**Job Recommendation System**
- Designed a personalized job recommendation system by analyzing users' work history and applying user collaborative filtering, resulting in 89% accuracy and a hit rate of 0.78.
- Conducted text preprocessing to transform job descriptions and user profiles into word2vec embeddings and TF-IDF vectors, ensuring accurate matching.

**Wide Residual Attention Networks**
- Enhanced ResNet model by adding mask attention after residual blocks, achieving 95.4% accuracy on the CIFAR-10 dataset through applied normalization, random flipping, and Gaussian noise to training images.
- Improved model performance by incorporating attention mechanisms, enabling the model to focus on relevant features.

**Chicken Disease Classification using MLOps**
- Designed a scalable ML solution by integrating pre-trained VGG16 with custom layers, executing data ingestion, training, and evaluation using DVC for pipeline tracking.
- Deployed the model using Docker on AWS EC2, leveraging GitHub Actions for automated CI/CD pipelines, ensuring efficient model deployment and management.

**Fraud Classification using GCP**
- Collected and stored employee data in GCS bucket, transforming data and masking sensitive information using Cloud Data Fusion.
- Created dashboards using Looker and automated ETL process with Airflow, enabling real-time data analysis and visualization.