

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, 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 technology: Hadoop, PySpark, Docker, Airflow, Git, MongoDB, Tableau, PowerBI, Lookup, Excel

Analytical skills: Image processing, CNN, LSTM, YOLOv5, K-means clustering, Grad-CAM, adversarial learning, domain generalization

Miscellaneous: robotic welding, natural language processing (NLP), machine learning (ML), deep learning (DL)

WORK EXPERIENCE

Graduate Research Assistant — Texas A&M University, USA

Jan 24 - Present

- Annotated over 1800 images using LabelMe software integrated with SAM to identify cotton balls, ensuring data quality for machine learning model development.
- Engineered and implemented segmentation models, including Unet attention, CBAM, Pix2Pix Unet, and SwinUnet, achieving a 0.70 IoU score and enhancing model performance through LoRA and skip connections.
- Conducted correlation analysis between pixels and cotton yield, achieving a 0.91 correlation accuracy and providing insights for crop yield prediction.

Graduate Research Assistant — Texas A&M University, USA

Sep 23 - Dec 23

- Experimented with CNN models, including AlexNet, VGG-16, ResNet, EfficientNet, and 3D CNN, for classification of stress-induced crops, and conducted hyperparameter tuning using Keras Tuner.
- Implemented an 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 and developed a domain generalization model to improve resilience across multiple domains, utilizing adversarial learning for model adaptation and Grad-CAM for better interpretability.
- Applied K-means clustering to segregate domains and used AlexNet as a feature extractor, achieving a 3% increase in accuracy over the baseline model through the incorporation of 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, leveraging YOLOv5 for joint detection with 99.5% precision.
- Utilized image processing techniques to denoise and identify edges, yielding an absolute error of $\hat{A} \pm 1\text{mm}$ for weld lines and $\hat{A} \pm 0.1\text{mm}$ 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, 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 automation of tasks.

Reproducing ChatGPT

- Trained GPT-2 124M model from scratch on FineWeb dataset, achieving a significant improvement in HellaSwag accuracy from 0.31 to 0.33 through enhancements such as replacing LayerNorm with RMSNorm and incorporating Rotary Positional Encodings.

- Implemented Group Query Attention to optimize multi-head attention, resulting in improved model performance.

Job Recommendation System

- Designed and developed a personalized job recommendation system by analyzing users' work history, applying text preprocessing techniques to transform job description and user profile to word2vec embeddings and TF-IDF vector.
- Modeled a neural ranker using user collaborative filtering, achieving 89% accuracy and a hit rate of 0.78, providing users with relevant job recommendations.

Wide Residual Attention Networks

- Enhanced ResNet model by adding mask attention after residual blocks, resulting in improved model performance and achieving 95.4% accuracy on the CIFAR-10 dataset.
- Applied normalization, random flipping, and Gaussian noise to training images, ensuring robustness and generalizability of the model.

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 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 data-driven insights and efficient data management.