Vineeth Sai Varikuntla

vineethsai4444@gmail.com —(415)-919 8646
LinkedIn GitHub GeeksForGeeks HackerRank Tableau Kaggle

Education

University of the Pacific Expected May 2025
Master of Science in Data Science *San Francisco, CA*

- **Relevant Coursework:** Advanced Machine Learning, Natural Language Processing, Relational Databases & NoSQL, Frequentist & Bayesian Statistics, Data Wrangling, Analytics Computing & Storytelling.

Jawaharlal Nehru Technological University July 2022
Bachelor of Technology in Computer Science and Engineering *Hyderabad, India*

- **Relevant Coursework:** Machine Learning, Data Structures and Algorithms, Data Mining, Artificial Intelligence, Cloud Computing.

Experience

Foundation AI Aug 2022 – May 2023
Associate Data Scientist *Hyderabad, India*

- Extraction: Extracted data from documents using a rule based extraction. Used Hough Lines Transform to deal with skewed data and YOLO V5 as an object detection and image segmentation model in a relevant use-case. Worked with Lilt for information extraction.
- Unitization: Worked on document unitization by using deep learning models such as BERT Classifier, RoBERTa and LayoutLM v3.
- Classification: Used some pre-trianed models such as XGBoost for classification.
- Auto-Validation: Achieved a stage for extractions where no manual validation is required.
- Optical Character Recognition (OCR): Used Tesseract, Textract for data extraction in various Legal and Healthcare use-cases.

Zummit Info. Labs May 2022 – September 2022
Junior Data Scientist Intern *Bangalore, India*

- Initial image processing - Using a variety of reconstruction algorithms such as contrast, edge enhancement, etc.
- Image segmentation - Identify/create areas (2D images) or volumes of interest (3D images).
- Features extraction and qualification-Features include volume, shape, surface, density, intensity, texture, location, and relations with the surrounding tissues.
- Used Semantic as well as Agnostic features as they are commonly used in the radiology lexicon to describe regions of interest and they attempt to capture lesion heterogeneity through quantitative mathematical descriptors.

Projects

Brain Tumour Detection | *Deep Learning, NLP, Threshold Segmentation, de-noising, Convolutional Neural Network, Wiener filters.*

- Utilized deep learning models such as Convolutional Neural Network to diagnose brain tumors through MRI scans.
- Designed and deployed a scalable RESTful API using Django and Django REST Framework, achieving a 30% improvement in data retrieval speed.
- Implemented a real-time chat feature using WebSocket and Socket.io, enhancing user engagement and reducing response time by 20%.

Research Interpreter | *RAG, Llama-Index, AstraDB, FastAPI, NextJS, TypeScript, React*

- Leveraged RAG, llama-index, AstraDB, and OpenAI's API to revolutionize research model comparison.
- Research Interpreter reduces researchers' literature review time, providing a user-friendly solution for complex datasets.
- Plans include expanding the database, adding AI features, and enhancing user engagement for ongoing development.

Technical Skills

Languages: Python, R, SQL, HTML/CSS, JavaScript, C++, Java
Technologies: TensorFlow, Tableau, spaCY, NLTK, cv2, Scikit-Learn, SciPy, Flask, PyTorch, AWS, GCP, Heroku, Llama Index
Models: YOLO v5, RoBERTa, LayoutLM v3, Lilt, BERT, XGBoost, Hough Lines Transform, RAG
Concepts: Deep Learning, Natural Language Processing, Generative Adversarial Networks, Machine Learning, Data Mining, Database, SQL, Statistics