# Bhanu Sunku

LinkedIn | Github | +91-9059461340 | bhanukedhar09@gmail.com

## WORK EXPERIENCE

**Data Analytics Intern**

*Elevance Health, Atlanta* | May 2024 - Dec 2024

- Designed and integrated a **ML model** using **Python**, **Scikit-Learn**, which reduced manual review hours by **60%** and enhanced team efficiency, resulting in faster file actions and **reduced fines** from performance guarantee clients.
- Achieved **95%** predictive accuracy in file status classification by applying advanced statistical techniques, gradient boosting, random forests with rigorous cross-validation and hyperparameter tuning, leading to more accurate and timely decision-making.
- Introduced interactive sunburst charts using **Plotly** and **Matplotlib** to track file progress across stages and increased the enrollment success rate by improving the overall process efficiency**.**
- Analyzed over **144k** audit reports to extract insights on file update failures, leading to a **20%** boost in model performance and more accurate predictions of potential issues.

**Graduate Assistant**

*Data Analytics Theory and Applications (DATA) Lab, Northern Illinois University, Dekalb* | Aug 2023 - Present

- Maintained large-scale, multi-node **Ray-vllm** inference pipelines on NIU Metis **HPC** deploying distributed jobs across 32 A100 GPUs optimizing throughput with high GPU utilization.
- Worked as Teaching Assistant(**TA**) for CSCI 502 (Java) and CSCI 503 (Python) leading lab sessions, grading assignments, and mentoring **40+** graduate and undergraduate students.

**DevOps Engineer Intern**

*Alation, Chennai(Remote)* | Apr 2021 - Aug 2021

- Improved systems cost per instance by reducing instance costs by **30%** through migrating Mac machines from MacStadium Anka to **AWS** Mac VMs with **Jenkins** integration, while designing **auto-scaling** and **load-balancing** for VMs that optimized resource utilization during peak demand.

## SKILLS & INTERESTS

**Programming languages** : Python, Java, SQL, JavaScript.

**Technologies** : AWS, Docker, GCP, Azure cloud, Git, Kubernetes, Tableau.

**Frameworks & Libraries** : PyTorch, Scikit-Learn, OpenCV, Pandas, NumPy, D3.js.

**Systems & MLOps** : Ray, vLLM, HPC, Distributed Inference, FAISS, RAG.

**Cloud & DevOps** : AWS, GCP, Azure, Docker, Kubernetes, Git.

## PROJECTS

**LLM-Driven Scientific Figure Analysis & Flowchart Reconstruction Pipeline**

- Created a framework that processes **~110 K arXiv research papers** and extracted **~ 700k figures** along with captions and associated text via **PDFFigures2.0**, **GROBID** and **PaddleOCR** creating a large-scale dataset for visual and text analysis.
- Finetuned **Llama4-Scout** and **Qwen2.5-VL** on SPIQA and ChartLlama dataset to generate concise figure summaries and high-quality question–answer pairs, enhancing figure comprehension.
- Designed an ensemble labelling pipeline by combining outputs from **Phi-4**, **Qwen 2.5-VL** and **InternVL**, then using **Llama 4-Scout** as an **LLM-as-judge** to select the most accurate labels, enhancing fine-grained figure annotations.
- Developed a multi-stage flowchart reconstruction pipeline leveraging Landing AI **OCR** for component detection and **Claude 3.5 Sonnet** for structural layout parsing, transforming outputs into targeted retrieval queries within a **RAG** framework to generate Chain-of-Events (**CoE**) representations, and generating conceptually aligned architecture diagrams in **TikZ**.

**Obesity Prediction: Leveraging LLM and Advanced Machine Learning Techniques**

- Built a predictive system to classify obesity levels from lifestyle, dietary, and demographic data for early high-risk detection.
- Implemented Decision Trees, Random Forest, **SVC**, **KNN**, and **GPT-4** with K-fold CV and Optuna, then combined them into a majority-vote ensemble, boosting accuracy from **94%** to **98%** while mitigating overfitting and underfitting.
- Applied feature engineering, data scaling (**StandardScaler**), and class balancing (**SMOTE**) to address imbalanced datasets, improving robustness; evaluated with **accuracy, weighted F1-score, recall, and precision**.

**GrantsFlow: Interactive dashboard for grants distribution for Louisiana hurricane Disasters(Rita&Katrina)**

- Developed a **React** + **D3** analytics app with six coordinated view hexgrid map, time histogram, Sankey, radial network chart, pop-out linked windows, and a funding house glyph to analyze Road Home grant distributions.
- Implemented a **hexgrid** brush/parish selector that cross-filters all views and powers parish-level analysis of reimbursements, provider relationships, and time trends; the house **glyph** encodes repair vs. rebuild, with a fixed stacked bar that updates on brush.
- Introduced direct-manipulation interactions: **Copy-as-Highlight** (persist a selection as a token across views) and **Drag-as-Filter** (drop tokens to apply scoped filters), with an optional link overlay to reveal cross-view relationships.

## EDUCATION

**Northern Illinois University** — **Dekalb, IL**

*Master's in Computer Science* — Aug 2022 -May 2024

**Amrita Vishwa Vidyapeetham (Amrita University)** — **Amritapuri, India**

*Bachelor's in Computer Science* — Jun 2017 - Jun 2021