Snehitha Tadapaneni

Arlington, VA | snehithat001@gmail.com | 202-768-4195 | github.com | linkedin.com

EDUCATION

The George Washington University **Washington DC**
Master of Science, Data Science (GPA: 4.0/4.0) **Expected May 2026**
Relevant Coursework: Introduction to Data Science, Data Warehousing, Introduction to Data Mining, Visualization of Complex Data, Machine Learning I:Algorithm Analysis, Time Series Forecasting for Analysis

Vellore Institute of Technology **Vellore, India**
Bachelor of Technology, Computer Science and Engineering (GPA: 8.81/10.0) **May 2024**
Relevant Coursework: DSA, Statistics, DBMS, Human Computer Interaction, Machine Learning, Deep Learning, Natural Language processing, Linear Algebra

Organizational Work Experience:
Society for the Promotion of Indian Classical Music And Culture Amongst Youth VIT
ChairPerson, Management Head

- Led a team of 50 members, overseeing all organizational aspects and approving major events.
- Conducted brainstorming sessions to innovate new event ideas and implemented strategies to foster a cohesive and engaging club environment.
- Coordinated event logistics, managed audience engagement, and handled publicity to maximize event reach and impact.
- Recruited and managed volunteers to ensure smooth execution of all club activities and events.

IEEE IAS VIT
Design Team – Core Committee Member

- Developed and enhanced the frontend of our chapter website, ensuring an intuitive and user-friendly interface.

TECHNICAL SKILLS & CERTIFICATIONS

- **Programming:** Python (Sklearn, TensorFlow, PyTorch, Scikit-Learn, Flask), R, SQL, C++.
- **Technologies & Tools:** Data Version Control, GitHub, Azure – Databricks, ML Studio, AWS – EC2, Microsoft – Power BI, Power Apps, MongoDB, Excel.
- **Competencies:** Data Analysis & Visualization, Statistical Modeling, Natural Language Processing, Deep Learning.
- **Certifications:** AWS Certified Cloud Practitioner (Jul 2023), Business Analytics with Excel by SmipliLearn (Jun 2023), Foundations of Data Science by Google (Aug 2023), Python for Data Science and ML Bootcamp by Udemy (Nov 2023).

PROFESSIONAL EXPERIENCE

Innomatics Research Labs **Hyderabad, India**
Data Scientist Intern **Jan – June 2024**

- Analyzed a dataset of thousands of internship applications to understand demographics of interns.
- Conducted an examination of recovered chat data to identify instances of misconduct related to sharing of unique Zoom meeting links.
- Analyzed historical data related to order arrival and delivery times for Domino's Pizza Store.
- Conducted an open-ended Exploratory Data Analysis on the AMCAT dataset including data visualization and statistical techniques and presented the document.
- Applied statistical tests to validate hypotheses and investigate specific claims, such as relationship between education and earning potential.
- Conducted a thorough examination of the backend codebase of the Note Taking Application to identify areas of improvement and bugs.

Smart Bridge **Remote**
Artificial Intelligence Intern **May - Jul 2023**

- Received comprehensive training in diverse development tools for AI, acted on a range of use cases, and completed a healthcare project using deep learning techniques.

TECHNICAL AND RESEARCH PROJECT EXPERIENCE

Multi Agent RAG Research Assistant | Python | LLM | Agentic AI **Mar – May 2025**

- Built a multimodal research assistant using IBM Watsonx, LangChain, and Gradio.
- Designed a multi-agent RAG pipeline (Parser, Retriever, Research, Verification, Decision, TTS) for accurate and trustworthy answers.
- Enabled robust document parsing with IBM Docling and hybrid retrieval (BM25 + ChromaDB) for high relevance.
- Integrated LLMs + verification layer to ensure factual, structured responses.
- Added Text-to-Speech support for accessibility and convenience.

Environmental Insights from the NYC Street Tree Census & Air Quality Data | Tableau **Mar – May 2025**

- Analyzed NYC Street Tree Census & Community Air Survey datasets to uncover spatial patterns in tree health, biodiversity, and air quality.
- Built an interactive Tableau dashboard showcasing tree density, species diversity, pollution levels, and stewardship impact across 59 districts.
- Conducted regression analysis ($R^2 = 0.557$, $p < 0.0001$) showing a negative correlation between species diversity and PM2.5 pollution.
- Integrated geospatial data with shapefiles to align tree and pollution data, driving insights for urban sustainability and environmental equity.

Analysis of crime rates on residential property in Washington DC in 2014-18 | Python **Oct - Dec 2024**

- Conducted an in-depth analysis of the relationship between neighborhood crime rates and residential property values using a dataset of 46,252 records and 30 attributes.
- Applied regression models, including Random Forest Regression and LightGBM, to uncover patterns linking crime rates to property values.
- Developed classification models (Random Forest, XGBoost) to categorize neighborhoods into housing price tiers based on violent crime rates and income levels, achieving actionable insights for policymakers and real estate stakeholders.
- Integrated and preprocessed data from Open DC Data and Kaggle by merging attributes across census tracts and years, ensuring a robust foundation for the analysis.

The Power of Data: ML in Cyber Attack Classification | Python & Framework, HTML, CSS **Jan - May 2024**

- Implemented a state-of-the-art cyber attack classification system using supervised machine learning.
- Curated a diverse dataset including malware, phishing, and DDoS attacks to configure a robust model that accurately categorizes threats.

Web App for Text Extraction and Summarization using GPT-3 Model | Python **Jul - Dec 2023**

- Developed a web application that automates text summarization using GPT-3 model.

Deep Learning Model for Eye Diseases Prediction | Python, Flask framework, HTML, CSS, JS **May - Jul 2023**

- Developed an AI-powered application to accurately predict eye diseases using Flask serving a REST API with React as the frontend.

PUBLICATIONS

- The Power of Data: ML in Cyber Attack Classification: Published in the IJNRD International Journal (ISSN: 2456-4184).
- Deep Learning Model for Eye Diseases Prediction: Published in the Sensors Imaging System and Techniques, AAP, CRC PRESS (SIST -5010 [CHAPTER NO: 10]).

AWARDS

- Recipient of the SEAS Dean's Scholarship Award for master's studies at GWU.