Sneha Sasanapuri

+1 720-238-1979 | snehasasanapuri@gmail.com | snsa4676@colordo.edu | https://www.linkedin.com/in/sneha-sasanapuri/ |
View my portfolio

Education

Master of Science in Data Science, University of Colorado Boulder (GPA:4/4) **Expected: May 2025**

Skills

- Programming Languages: Python, C, R
- Machine Learning & AI: TensorFlow, PyTorch, Scikit-learn, OpenCV, KNIME
- Data Visualization: Libraries (seaborn, ggplot, matplotlib, plotly), Power BI, Tableau, Excel
- Data Manipulation: Pandas, NumPy, SQL
- Cloud Platforms: AWS Lambda, S3, SageMaker, ELB, Docker & Kubernetes
- Natural Language Processing (NLP): NLTK, spaCy
- Gen AI: Langchain, RAG
- Version Control: Git
- Scripting/Automation: VBA, Robotic Process Automation
- Statistical Analysis: Regression, distributions, hypothesis testing
- Familiar with MapReduce framewrok in Hadoop, PySpark, HTML

Work Experience

Research Data Analytics Intern, Research & Innovation Office, CU Boulder **Dec 2024-Present**

- Support data-driven decision-making by collecting, cleaning, and organizing datasets related to research activities and institutional performance.
- Conduct **advanced analysis**, **predictive modeling**, and trend studies to inform strategy and resource allocation.
- Develop insightful **dashboards** and visualizations using Tableau, Power BI, and Excel.
- Implementing **Capacity, Program, Sponsor & Funding landscape analysis**, calculating IMPACT scores.

Machine Learning & Automation Engineer, Larsen & Toubro, India **May 2020-Jul 2023**

- **Implemented ensemble learning** techniques (XGBoost) that analyzed historical demand data to identify key factors such as seasonality and consumption rates, and hyper-parameter tuning, which improved prediction accuracy and model robustness by 30%.
- Applied **deep learning techniques** (**CNN**) to process radiographic test films for weld seam defect detection, achieving 52% accuracy, automating a traditionally manual process and setting the stage for broader defect classification.
- **Transformed complex data structures** using **Python, Pandas**, and **SQL** to generate actionable business insights, leading to a **23% reduction** in operational delays by optimizing workflows and improving decision-making processes.
- Implemented **NLP** techniques to automate the classification of third-party inspection feedback, utilizing **dependency parsing** to analyze relationships between key terms, reducing nullification response time by **20%** and expediting corrective actions.
- Led **A/B testing** and **hyper parameter tunin**g to refine demand prediction models, resulting in a 15% increase in feature optimization and better alignment with business needs.

- Developed an intelligent system integrating AUTOCAD and SQL using **AutoLISP** and **pyodbc** to automate real-time tracking of component installations, resulting in a 30% improvement in workflow efficiency.
- Integrated **Synchro 4D BIM scripting** with **MySQL** & **PLM** software to streamline data flow and ensure accuracy and reliability, resulting in a 31% improvement in resource planning efficiency, enhancing project execution and material tracking.
- Optimized **SQL queries** to handle large datasets for integrating with POWERBI, reducing query run-time by ~40% and enabling real-time data insights to support model accuracy and business decision-making.

Data Analytics Engineer, Larsen & Toubro, India **Aug 2017-Apr 2020**

- Led client-facing projects, delivering business insights and optimizing production workflows with advanced predictive models in Power BI and Python.
- Created interactive dashboards and **visualizations** using **Power BI**, providing key insights to stakeholders and supporting decision-making in resource allocation and strategy planning.
- Created dynamic, interactive **Excel dashboards** with PivotTables and charts, providing real-time insights into project performance metrics by **automating with MACROS**.
- Created user guides and video tutorials to facilitate adoption of the dashboard among employees.
- Presented sophisticated data insights to non-technical stakeholders and senior management.

Projects:

- Refabricating Analysis: Conducted root cause analysis using Python and SQL, reducing rework by 37% and optimizing workflows.
- Category Classification: Automated category assignment using text mining in **KNIME,** improving resolution rates by 26%.
- Workmen Monitoring System: Collected and analyzed RFID reader's data to track workmen movement patterns on the shop floor, identifying inefficiencies and unnecessary travel by integrating with **Power BI dashboards**, resulting in 41% reduction in waiting time.
- FTR Dashboard: Created an interactive Power BI dashboard using DAX, reducing production downtime by 23% through data-driven insights.

Graduate Engineer Trainee- Project Planning & Control, Larsen & Toubro, India **Aug 2015-Jul 2017**

- Pioneered a solution for ship block movement, achieving timely milestones and saving ₹1.6 Crores in indirect costs.
- Advocated for material change from steel to bronze, creating a visual inspection matrix that saved ₹29 lakhs in maintenance costs.
- Managed project timelines, tracked deliverables, and collaborated with cross-functional teams using Microsoft Project & Concerto to streamline workflows and ensure on-time project delivery.
- Maintained L1 & L2 schedules in Concerto and presented milestones progress to PMO, escalating budgetary constraints and manpower allocation.

Academic Projects

WhatsApp Chat Visualizer

Data mining and **NLP techniques** applied to extract key communication patterns from chat logs using **spaCy** and **NLTK**, and visualized findings using **matplotlib** for detailed analysis.

Decoding Cricket's evolution

Using **two-sample mean** and **proportion tests** in **R** programming, explored batting averages, toss outcomes, and home-ground advantages. Visualizations such as box plots and choropleth maps highlighted performance trends across nations, contributing to insights on cricket's evolution during this period.

RAG based Chatbot

Developed an RAG-based chatbot using **LangChain** to simplify research paper analysis by retrieving document segments from **Pinecone** and generating context-aware responses via OpenAI's GPT API. Built a user-friendly Streamlit interface for seamless interaction.

Music Separation & Captioning

- Developed an audio track separation service using DEMUCS as **REST API** in a **Kubernetes** environment, enabling real-time, efficient processing of large audio files with **Redis** for task queuing and **Min.io** for object storage.
- Implemented an end-to-end audio captioning system by integrating AWS **S3, Lambda**, and **SageMaker** to process music clips, generate chunk-wise captions using a fine-tuned **BART model**, and summarize them with the **OpenAI API**. Hosted the application on **Elastic Beanstalk** with automated artifact packaging and dependency management via custom Docker images through AWS **CodeBuild**.

Accomplishments

- **Hackathon Winner**: Automated extraction of key information from housing rental agreements using **NER**, boosting processing efficiency.
- Customer Kudos: Received "**I Appreciate Note**" for analyzing customer service metrics; identified three key areas for improvement, leading to increased customer retention.
- **Digital Ambassador of the Year**: Awarded for leading the integration of cutting-edge automation tools, which significantly improved system performance and streamlined our workflows, resulting in increased efficiency.