CHAKRAPANI GAJJI

Phone: +1(785)317-5938 | E-Mail: cgajji@ksu.edu | LinkedIn: chakrapanigajji
GitHub: chakrapani2122 | Portfolio : Chakrapani

SUMMARY

Data Scientist specializing in predictive modeling and machine learning applications. Proven ability to translate complex environmental and sales data into significant business impact, demonstrated by projects that enhanced data accuracy by 98% and improved forecast precision by 15%. Expert in the Python data stack (Scikit-Learn, TensorFlow) and committed to architecting data solutions that drive revenue and efficiency.

TECHNICAL SKILLS

- **Data Ingestion & Pipelines**: SQL, Python (Pandas, NumPy), MySQL, R
- **Machine Learning & Modeling:** Scikit-Learn, TensorFlow, PyTorch, Statistical Analysis, NLP, Computer Vision (OpenCV)
- **Deployment & MLOps:** Docker, Streamlit, Git, CI/CD (GitHub), Hadoop
- **Visualization & BI Tools:** Tableau, Power BI, Matplotlib, Seaborn
- **Cloud Platforms:** Databricks, AWS

EDUCATION

Master of Science in Data Analytics (Expected May 2026),
Kansas State University, Manhattan, KS
GPA: 3.83/4.0

Bachelor of Technology in Computer Science Engineering (AI & ML) (2024)
Sri Indu College of Engineering and Technology, Hyderabad, India
GPA: 3.4/4.0

EXPERIENCE

Graduate Research Assistant, Regenerative Agriculture Project (June 2025 – Present)

- Designed and automated **data pipelines with Python, Pandas, and Excel**, ensuring accurate cleaning, validation, and organization of large-scale agricultural datasets for ongoing research.
- Developed and deployed a **Streamlit web application** that enabled researchers to upload, update, visualize, and analyze soil quality data directly on the project website.
- Acted as the **data steward for the project**, maintaining structured records and providing curated datasets and analysis to support researchers' publications and theses.

Graduate Research Assistant, RAIN Project (August 2024 – June 2025)

- Collected, consolidated, and managed **8+ years of soil, crop, and hydrology data** across 10,000+ acres by collaborating with researchers and GRAs, ensuring consistency with historical datasets.
- Built scalable **data management workflows** in Python and Excel that streamlined data cleaning, storage, and retrieval, significantly improving reliability for modeling and analysis.
- Created an **interactive Streamlit dashboard** for real-time visualization and exploration, boosting collaboration and accelerating research insights across interdisciplinary teams.

Data Science Intern, Oasis Infobyte, Hyderabad, India (February 2023 – March 2023)

- Improved sales forecast precision by 15% over previous benchmarks by developing a Random Forest model, optimizing data pipelines, and re-engineering feature selection strategies.

PROJECTS

An Inflation Story: A Deep Dive into Food Prices in the USA

- Designed and developed an **interactive Tableau dashboard** analyzing U.S. food price inflation trends, volatility, and drivers, highlighting differences between food at home vs. away and uncovering seasonal/long-term patterns.
- Implemented **advanced visualizations** (dual-axis line charts, heatmaps, stacked area charts, forecast funnels) with calculated fields, interactivity, and time-series analysis to deliver actionable insights for decision-making.

Anatomy of a Tremor: A Global Look at Recent Earthquakes

- Transformed raw seismic datasets into a **comprehensive Tableau dashboard** featuring dynamic maps, scatter plots, histograms, and ranking visuals to reveal global earthquake patterns and data reliability.
- Engineered an **interactive, user-centric BI solution** with cross-filtering actions, storytelling design, and professional UI/UX, enabling non-expert audiences to explore complex geospatial and time-series data.

Institutional Research KSU: Response & Non-Response Analysis

- **Analyzed 5,000+ alumni and senior survey responses** using machine learning to uncover participation patterns, informing a new communication strategy.
- **Built predictive models in Scikit-Learn and proposed Tableau dashboards** to visualize response trends, providing a data-driven framework to optimize outreach, increase engagement, and boost future survey participation.

Vivid Tones: Image Colorization with CNNs

- Engineered a TensorFlow-based image colorization pipeline with a CNN, trained on 100,000+ images, and reduced model training time by 15% through optimized data preprocessing techniques.
- This technology is directly applicable to enhancing archival media, improving medical imaging diagnostics, and creating compelling marketing content from monochrome assets.

Precision Object Counting System

- Increased object count accuracy by 15% by developing a real-time object detection tool in Python and OpenCV that integrates edge filtering and contour tracking.

Email Spam Detection

- Built a Logistic Regression model using NLP that achieved 98% accuracy in classifying spam emails.

Advertising Sales Prediction

- Pioneered a Random Forest regressor to forecast ad sales with 98% accuracy: enhanced performance through features and cross-validation.

PUBLICATION & CERTIFICATIONS

- **Publication:** A Survey on Large Language Models: Overview and Applications, International Research Journal of Engineering and Technology (IRJET), June 2024
- **Certifications:**
  - Lean Six Sigma Yellow Belt Certification (Kansas State University, Aug 2025)
  - Data Analysis with Python (IBM)
  - Introduction to Data Science (Infosys)
  - SQL Essential Training (LinkedIn Learning)
  - Python Object-oriented Programming (LinkedIn Learning)
  - Python Basics (IBM)
  - Learning C (LinkedIn Learning)