GIRI SAI KUMAR PRAGADA

Binghamton, NY | +1(607)-245 9921 | gpragada@binghamton.edu | linkedin | GitHub

EDUCATION

Binghamton University, State University of New York, Thomas J. Watson College of Engineering and Applied Science
Masters of Science in Information Systems   *May 2025*
Cumulative GPA: 3.82 /4.00 | **Dean's List**: Fall 2023 – Spring 2025
Relevant Coursework: Data Structures and Algorithms,Machine Learning,Deep Learning,Artificial Intelligence,Tools for data science,Database Systems,Information Systems,Network and Security.

TECHNICAL SKILLS

Languages: C, Java, SQL,Pyspark,Python, R (basic).
Software and OS: XCode, Eclipse, Android Studio, Visual Studio Code, Jupyter Notebook, Git, Linux,Vim.
Database: MSSQl,PostgresQL,MySql,MangoDB.
Productivity & Automation: Advanced Excel (Pivot Tables, vlookup, Macros, VBA), Power Automate, PowerApps, SharePoint, Jupyter Notebooks, Git, REST APIs, Data Automation Scripting.
Data Visualization & BI Tools: Tableau, Power BI, Looker, Google Data Studio, Oracle Fusion Data Intelligence (FDI)
Cloud Platforms:AWS, Microsoft Azure (Synapse Analytics).
Emerging Technologies:AI & Machine Learning, Large Language Models, Prompt Engineering, Langchai,Langgraph.
Additional: HTML, CSS, XML Parsing, TCP/IP, SQL Workbench, Data Visualization, Predictive Analysis,Agile Methodologies (Scrum, Kanban, User Story Development, Sprint Planning), Jira, Confluence, Cross-Functional Team Collaboration, Stakeholder Management.
Certifications: AWS Generative AI, Amazon Bedrock,AWS Prompt Engineering,Langchain Langgraph.

PROFESSIONAL EXPERIENCE

Tata Consultancy Services(TCS), *Data Analyst* | Hyderabad, India   *April 2021 – July 2023*
- Utilized SQL (MSSQL) and Python (pandas, seaborn) to analyze 20M+ transaction records, identifying customer churn segments and driving a 12% increase in retention.
- Migrated 10+ legacy databases from **on-premise MSSQL** to **Azure SQL (IaaS, PaaS)**, ensuring zero downtime for 2 global clients.
- Built and maintained **ETL pipelines** using **Airflow, dbt, and SSIS**, integrating CRM, **Synapse (Azure)**, and payment data into **BigQuery** and **Power BI**.
- Developed 15+ **Power BI dashboards** with row-level security, replacing Excel models and improving access for 200+ users across 15 sales teams.
- Conducted **SQL performance tuning**, index optimization, and query refactoring, reducing report runtimes by up to **60%**.
- Created stored procedures, triggers, and T-SQL scripts for high-availability environments using **Always On, Log Shipping, and Replication**.
- Delivered predictive models using **Python (pandas, seaborn)** and **R**, improving churn prediction accuracy and boosting retention by **12%**.
- Implemented **PowerApps**, **SharePoint** tools for internal support tracking, reducing ticket resolution time by **35%**.
- Built **prompt templates** for ChatGPT-based internal ticketing support prototype, reducing manual triage time by 20% in pilot phase.
  Environment: SQL (MSSQL), Python (pandas, seaborn), Power BI, Row-Level Security, Excel, Airflow, dbt, CRM, Synapse (Azure Cloud), BigQuery, R, Regression Analysis, Customer Lifetime Value Prediction, Jupyter Notebooks.

Cybage Software, *Data Analyst Internship* | Hyderabad, India   *January 2019 – March 2021*
- Cleaned and preprocessed 15K+ rows of cloud usage data using Python (Pandas), reducing manual Excel work by 60%.
- Wrote and optimized 10+ SQL queries in **PostgreSQL** to generate weekly usage summaries and reduce reporting time by 30%.
- Contributed to building 3 **ETL** pipelines in **Snowflake** and dbt, supporting migration analytics for clients moving to Azure Synapse and AWS Redshift.
- Designed 4+ **Tableau** dashboards to visualize cloud cost patterns, helping uncover 10–15% potential savings during monthly reviews.
- Automated project status reports with Excel VBA macros, reducing manual reporting time by 40%.
- Maintained and updated **Agile** boards in **Jira** for 2 cross-functional teams, ensuring 100% sprint closure over a 10-week internship.
- Created Power Automate workflows for SLA alerts, used by global IT teams in 3 regions to monitor performance.
- Assisted in conducting **A/B tests** and statistical analysis on network data, identifying latency bottlenecks that led to a 25% uptime improvement.
  Environment: Python, Pandas, SQL, PostgreSQL, Snowflake, dbt, Tableau, Azure Synapse, AWS Redshift, Excel, VBA, Jira, Agile, Power Automate, SharePoint, Time Series Analysis, Data Cleaning, ETL, Dashboards

PROJECT EXPERIENCE

AI-Powered Business Insights Chatbot, *Full Stack ML Engineer | Independent Project*   *May 2025 – june2025*
- Created a user-friendly **chatbot interface using Streamlit** that allows users to ask business-related questions in plain English and receive data-driven answers.
- Integrated **LangChain** and **OpenAI's GPT API** to convert natural language queries into **SQL statements** for **PostgreSQL** and **MySQL** databases.
- Processed and filtered company sales and operations data using **Pandas** and **PySpark SQL** to support query responses.
- Generated charts and tables using **matplotlib** and **seaborn** to visualize insights directly in the chatbot interface.
- Implemented **input validation and output parsing** using **Pydantic** to ensure structured, type-safe communication between the frontend and backend.
- Added robust **error handling** for malformed queries and invalid inputs, achieving near **100% crash-free user interactions** during testing.

Pneumonia Detection System, *Full Stack ML Engineer* | Capstone Project   *January 2025 – May2025*
- Led development of an end-to-end **deep learning**-based diagnostic tool that detects pneumonia from chest X-ray images, completed as a Capstone Project with a team of 3.
- Trained a **CNN model** using transfer learning on a dataset of 5,000+ images, achieving **96% classification accuracy** and improving F1 score from 0.78 to 0.91 through **hyperparameter tuning**.
- Performed image preprocessing steps including resizing, normalization, and augmentation to enhance model generalization.
- Created an interactive **Streamlit** frontend allowing doctors to upload X-ray images and receive instant diagnostic results with confidence scores.
- Integrated **LangChain** for potential future expansion into conversational AI interfaces (e.g., assisting doctors with follow-up questions).
- Serialized the trained model as a Pickle file, reloaded it during inference, and provided **real-time predictions** with visual and text-based output.
- Containerized the entire application using **Docker**, ensuring consistent behavior across environments.
- Deployed the model and UI as an AWS-hosted service using **SageMaker** endpoints, enabling scalable, secure, and fast access to predictions.
- Evaluated the system using precision, recall, **AUC-ROC** curves, and confusion matrices to ensure reliability in clinical use cases.

YouTube API Data Analysis, *Data Analysis | Independent* Project   *November 2024 – December 2024*
- Extracted real-time video and channel metrics using **YouTube Data API v3** for 50+ content creators.
- Performed **trend analysis** on engagement rates, watch time, and subscriber growth.
- Created visualizations to predict future performance metrics using **matplotlib** and **seaborn**.

Netflix Subscription Forecasting, *Data Science* | Independent Project   *May 2021 – August 2021*
- Built a **time series forecasting pipeline** using **ARIMA and Facebook Prophet** to predict Netflix subscriber growth trends over 12 months.
- Performed **feature engineering and EDA** on historical data to uncover insights into seasonal viewing patterns and user retention.
- Evaluated models using **RMSE, MAE, and MAPE**, achieving <10% error across validation windows.
- Designed a **Streamlit dashboard** to visualize forecasts and simulate promotional campaign impact.

LEADERSHIP EXPERIENCE

HackBU(DSA-Focused Hackathon Project), *Team Lead*   *August 2024 – December 2024*
- Led a 4-person team during **HackBU**, guiding algorithm design and implementation for a real-time coding problem analyzer.
- Designed and implemented efficient **data structures** and applied advanced **sorting and graph algorithms** to support code evaluation and performance metrics.
- Facilitated team strategy, delegated tasks, and ensured on-time delivery of a working MVP under a 24-hour deadline.
- Fostered a collaborative environment, helping teammates understand complex algorithmic trade-offs and optimization strategies.

BUAI(Artificial Intelligence Club), *co-founder & president*   *January 2021 – May 2023*
- Founded and led a student club focused on **machine learning, deep learning, and large language models**, with 40+ active members.
- Organized weekly workshops and speaker sessions covering **NLP, computer vision,** and real-world AI use cases.
- Mentored members on AI tools and projects using frameworks like **PyTorch, TensorFlow**, and **Hugging Face**.