Giri Sai Kumar Pragada| Data Scientist

📍 Binghamton, NY | ✉ girisaikumarpragada@gmail.com | ☎ (607) 245 – 9921 | 🔗 LinkedIn

SUMMARY

- 3.5+ years of experience in developing scalable data science solutions using Python, with a strong ability to automate complex data processing workflows and streamline business operations.
- Skilled in designing robust ETL pipelines using PySpark and SQL, capable of efficiently processing large-scale datasets and improving data availability for enterprise reporting systems.
- Adept at building interactive web-based interfaces using HTML, CSS, and JavaScript, enhancing the usability and functionality of data visualization tools for internal stakeholders.
- Strong foundation in machine learning model development using TensorFlow, with experience in predictive analytics that supports business forecasting and strategic planning.
- Strong foundation in machine learning, deep learning, and predictive modeling using TensorFlow, with experience in building analytics solutions that support business forecasting and strategic planning.
- Experienced in advanced data manipulation and statistical programming using NumPy and Pandas, enabling efficient data preparation and transformation for analytical tasks.
- Knowledgeable in applying Scikit-learn and SciPy for feature engineering and model evaluation, with a focus on improving model performance and accuracy.
- Competent in creating clear and insightful data visualizations using matplotlib and Seaborn, translating technical findings into actionable insights for non-technical teams.
- Hands-on expertise in relational database management using MySQL, MS SQL Server, and PostgreSQL, with a focus on performance tuning and large-volume data processing.
- Experienced in developing BI dashboards using Power BI, integrating business KPIs to support decision-making with real-time, interactive visual insights.
- Skilled in using Tableau for designing user-centric dashboards and monitoring performance metrics, improving engagement and campaign tracking effectiveness.
- Proficient in Microsoft Excel for financial modeling and reporting, including advanced functions like Pivot Tables, VLOOKUP, and automation techniques to streamline repetitive tasks.

TECHNICAL SKILLS

Languages:	Python, PySpark SQL, HTML, CSS, JavaScript
Frameworks:	TensorFlow, NumPy, Pandas, Scikit-learn, SciPy, matplotlib, Seaborn, PyTorch,
Databases:	MySQL, MS SQL Server, PostgreSQL
Data Visualization:	Power BI, Tableau, MS Excel
AI/ML Capabilities:	Machine Learning, Deep Learning, Natural Language Processing (NLP), Predictive Modeling, AI-Powered Solutions
Tools:	Git, MS Office Suite (Word, Project, Excel, Visio, Excel Pivot table, V-looks, PowerPoint), JIRA, SAP, Airflow
Analytic Skills:	Requirements gathering, Statistical Analysis, Process modeling, Stakeholder management, Gap analysis
Cloud Platforms:	AWS
Methodologies:	Agile (Scrum, Kanban, User stories, sprint planning, Retrospective meetings)

EXPERIENCE

Data Scientist (Intern)
Northern Trust | Binghamton | Jan 2025 – Current

- Deployed ML and NLP models using AWS SageMaker and Lambda, creating scalable AI-powered solutions and reducing infrastructure costs by 22% through serverless deployment.
- Python developed a customer segmentation model using K-Means and XGBoost that improved campaign engagement by 27% and reduced churn by 18%.
- PySpark SQL processed over 2 TB of structured and unstructured data from multiple sources, improving data availability by 35% through optimized ETL pipelines.
- TensorFlow implemented deep learning models to predict customer lifetime value with 93% accuracy, driving personalized marketing strategies.
- MySQL wrote complex queries and optimized database performance to reduce report generation time by 40% for monthly business reviews.
- Tableau built executive dashboards with real-time KPIs for sales and customer acquisition, increasing stakeholder visibility and insights.
- Git managed version control across 30+ data science notebooks and production scripts, enabling collaborative development and code reliability.

Data Analyst
Tata Consultancy Services | India | Apr 2021 – Jul 2023

- Leveraged Machine Learning and Deep Learning frameworks such as TensorFlow and Scikit-learn for predictive modeling, including regression, classification, clustering, and Natural Language Processing (NLP) tasks.
- Developed and maintained data pipelines using PySpark and SQL across AWS environments, ensuring scalable and secure data transformations.
- Conducted comprehensive data analysis using Python (Pandas, NumPy, SciPy, Scikit-learn) and statistical modeling to support business decision-making and identify trends.
- Designed and deployed interactive dashboards in Power BI and Tableau, transforming complex datasets into actionable insights for stakeholders.
- Collaborated cross-functionally in Agile/Scrum teams, participating in sprint planning, backlog grooming, and retrospectives to deliver data-driven features.
- Performed ETL operations using Airflow, automating workflows and ensuring reliable data ingestion from sources like MySQL, PostgreSQL, and MS SQL Server.
- Created and optimized SQL queries for ad-hoc reporting, business intelligence, and advanced analytics in relational databases.
- Leveraged TensorFlow and Scikit-learn for predictive modeling and machine learning, including regression, classification, and clustering algorithms.
- Delivered executive-level reporting and visualizations using Excel PivotTables, VLOOKUPs, and PowerPoint, supporting strategic business reviews.

Data Analyst
Cybage Software | India | Jan 2020 – Mar 2021

- Optimized complex PostgreSQL queries, reducing database query execution time by 40%, ensuring faster data retrieval for analytics teams.
- NumPy utilized matrix operations and statistical methods to analyze pricing trends for over 500 products across multiple retail channels.
- Seaborn created correlation and distribution visualizations for marketing datasets, helping to optimize budget allocation by 12%.
- PostgreSQL extracted and joined data using advanced SQL queries including CTEs and window functions for campaign analytics.
- JavaScript contributed to building interactive dashboards using JavaScript chart libraries, improving user engagement with internal tools.
- Excel VLOOKUP automated reconciliation processes using Excel formulas and macros, saving over 30 hours per month in reporting.
- MS Office Suite prepared executive presentations and documentation using PowerPoint, Word, and Visio for project tracking and communication.
- Requirements Gathering conducted stakeholder interviews and translated business KPIs into technical specifications for dashboard development.

EDUCATION

🎓 **Master of Science in Information Systems**   **May 2025**
   Binghamton University, State University of New York Binghamton

🎓 **Bachelor of Technology in Electronics and Communication**   **June 2016 – Apr 2020**
   Andhra University, India Visakhapatnam

PROJECT

AI-Powered Business Insights Chatbot

Designed and developed an interactive GenAI-based chatbot that enables users to ask business questions in natural language and receive dynamic responses by executing real-time SQL queries on sales and operations data, with results visualized using charts and dashboards. The solution enhances stakeholder decision-making by bridging the gap between raw data and strategic insight through conversational analytics.

- Integrated OpenAI's GPT model to translate user queries into optimized SQL code for PostgreSQL and MySQL databases.
- Built a clean frontend interface (HTML/CSS/JS) for interactive user input and visual feedback.
- Used Pandas & PySpark SQL for data processing and Seaborn/matplotlib for in-app visualizations.
- Embedded Power BI/Tableau dashboards for deeper executive-level analytics and reporting.
- Automated data refresh and ETL workflows using Apache Airflow and scheduled jobs.
- Followed Agile methodology (sprint planning, retros, user stories) and tracked tasks via JIRA & Git.