Chaitanya Dasari

763-286-9003 | chaitanyadasari09@outlook.com | LinkedIn | GitHub

Professional Summary

Data Scientist with 5 years of experience specializing in the design and production deployment of advanced AI systems. Expert in **Generative AI**, with hands-on experience building and scaling **LLM-powered RAG systems, autonomous agents (Agentic AI), and knowledge graphs (Neo4j)**. Proven track record of delivering measurable business value, including a **15% reduction in churn** and a **50% reduction in analyst research time**. Highly skilled in Python, SQL, and MLOps on Azure, GCP, and AWS.

Skills

Programming Languages
- Python; SQL; Java; R

Machine Learning & AI
- Supervised/Unsupervised Learning, NLP, Recommendation Systems, Information Retrieval, Reinforcement Learning, LLMs(GPT-4, Claude, LLaMA), Generative AI (Stable Diffusion, DALL·E 2)

Frameworks
- PyTorch, TensorFlow, Keras, XGBoost, Scikit-learn, Spark MLlib, LangChain, LangGraph, Hugging Face

Data Tools & Technologies
- Pandas; NumPy; REST APIs; ETL Pipelines; Automation; Feature Engineering; Dimensionality Reduction (PCA); SQL (MySQL, BigQuery, Snowflake, Redshift)

Statistical Methods
- Regression (Linear, Logistic, Ridge), Hypothesis Testing, ANOVA, Bayesian Inference, Time Series (ARIMA, Prophet), Clustering (K-means, DBSCAN, Hierarchical)

DevOps & MLOps
- CI/CD Pipelines, Git, Unit Testing, Automation, Model Deployment & Monitoring.

Cloud & Big Data
- AWS (SageMaker, Redshift), GCP (Vertex AI, BigQuery), Azure, Snowflake, Databricks

Data Visualization & BI
- Tableau, Power BI, AWS QuickSight (familiar), Matplotlib, Seaborn, Interactive Dashboards

Professional Strengths
- Cross-functional Collaboration, Leadership, Technical Communication, Innovation, Agile/Scrum, Mentorship, Data Storytelling

Professional Experience

Data Scientist,
AT&T - Dallas, TX | July 2025 – Present

- **Reduced** voluntary churn **by 15%** by designing, building, and deploying an end-to-end predictive model using **XGBoost** and **Python**.
- **Generated** natural-language summaries explaining churn drivers by building a Generative AI assistant **using RAG** to analyze unstructured support tickets and call transcripts.
- **Automated** proactive retention is offered by deploying an agentic AI system that uses the **Model Context Protocol (MCP)** to query CRM/Billing APIs and trigger interventions securely.
- **Ensured** model reliability and scalability by engineering the complete MLOps pipeline on **Azure Databricks** and **MLflow** for real-time feature engineering, inference, and monitoring.

Data Scientist,
Piper Sandler, USA | April 2024 – July 2025

- **Reduced** financial analyst research time **by 50%** by deploying an **LLM-powered assistant (GPT-4, RAG)** to query and summarize financial documents.
- **Enhanced** context retrieval and RAG output accuracy by building a **Neo4j Knowledge Graph** integrated with LLM embedding models.

- **Improved** risk-adjusted returns **by 15%** (backtested) by designing autonomous **trading agents** in **Python** to detect market anomalies and execute hedging strategies.
- **Boosted** fraud identification precision **by 20%** by architecting a detection pipeline combining **Graph-Based (Neo4j)** community detection with **NLP** sentiment analysis.

Graduate Research Assistant,
Concordia University St. Paul | Nov 2023 – Mar 2024

- **Achieved** 88% accuracy by building a student retention prediction model using **Python (Scikit-learn)**.
- **Co-authored** a successful $50K grant proposal by researching AI ethics in healthcare
- Processed and analyzed large-scale **social media text data (NLP, SpaCy, NLTK)** to uncover sentiment patterns.
- Designed **Tableau dashboards** to present findings on alum engagement, supporting strategic planning.

Data Scientist (Functioned as ML Engineer),
Infosys, India | Jun 2022 – Aug 2023

- Served as an ML Engineer and Data Scientist on engagements with primary healthcare and financial services clients, delivering advanced AI/ML solutions.

Client Project: CVS Aetna (Health Care)

- **Reduced** patient readmissions **by 12%** by building and deploying a patient segmentation model **(K-Means, DBSCAN)** on claims data to enable targeted care programs.
- **Identified** high-cost claimants with **85% accuracy** by developing predictive models **(XGBoost, Random Forest)** on **GCP Vertex AI**.
  - **Improved** patient risk scoring by applying **BERT-based NLP** to extract clinical entities from provider notes.
- Applied **BERT-based NLP** to extract clinical entities from provider notes, improving patient risk scoring.
- Engineered HIPAA-compliant ETL pipelines (Azure Data Factory) integrating claims, EHR, and pharmacy data.
- Used **GANs for synthetic data generation**, supporting privacy-preserving healthcare analytics.

Client Project: Charles Schwab (Financial Services)

- **Explored** Generative AI capabilities by building prototype **LLM-based chatbot workflows** and fine-tuning open-source transformer models.
- **Improved** data pipeline efficiency by building **agentic automation** for ETL and anomaly detection using **Databricks** and **Python**.
  - **Reduced** data processing time **by 30%** by optimizing analysis pipelines in **Databricks**.

Data Analyst
Twilight Software Solutions, Hyderabad, India | Oct 2019 – Jun 2022

- **Strengthened** model accuracy for churn and demand forecasting models by applying **statistical methods** and **PCA** for dimensionality reduction.
- **Supported** all business intelligence reporting by building and managing scalable **ETL pipelines** and data warehouses.
  - **Tracked** critical business KPIs for leadership by developing and maintaining interactive dashboards in **Tableau**.

Education

- Master's in Data Analytics | Concordia University, St. Paul, MN | **GPA: 3.91/4.00**
- Bachelor's in Science | Nizam College, Hyderabad, India **| GPA: 8.22/10.00**

Certifications

- Machine Learning (Stanford)
- Data Science & Cloud Computing (Infosys)
- Data Analytics (Cisco)