

# Chaitanya Dasari

763-286-9003 | chaitanyadasari09@outlook.com | [linkedin](#) | [github](#)

## Professional Summary

---

**Senior Data Scientist** with 5 years of experience building and deploying end-to-end AI solutions from concept to production. Expert in **Large Language Models** (LLM), including RAG with GPT-4, and **Generative AI** pipelines. Proven track record of driving business impact, including **reducing manual analysis by 40%** and **optimizing automated trading strategies**. Highly skilled in Python, SQL, and cloud-native data science on GCP, Azure, and AWS.

## Skills

---

### Programming Languages

- Python; SQL; Java; Scala; R

### Machine Learning & AI

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

### Frameworks

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

### Data Tools & Technologies

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

### Statistical & Analytical Methods

- Hypothesis Testing; Regression Analysis (Linear, Logistic, Ridge, Lasso); Probability Distributions; 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, GCP, Azure, Vertex AI, SageMaker, Azure ML, Databricks, Apache Spark, Redshift, BigQuery, Snowflake

### 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

---

### Senior Data Scientist,

### Piper Sandler, USA | April 2024 – Present

- Led deployment of an **LLM-powered research assistant (GPT-4 + RAG)**, reducing analyst turnaround time by 50% and improving decision-making speed.
- Built **Neo4j-based knowledge graph + LLM integration** to enhance context retrieval in financial documents.
- Designed **autonomous trading agents** to detect anomalies and execute hedging strategies, improving risk-adjusted returns by 15% (backtested).
- Architected a **fraud detection pipeline** combining graph-based community detection with NLP sentiment analysis, boosting fraud identification precision by 20%.
- Developed predictive models with feature engineering + hyperparameter tuning to optimize risk management.

- Created **real-time dashboards (Tableau, Power BI)** to deliver actionable trading and risk insights to leadership.
- Streamlined **ETL workflows and large-scale data pipelines**, ensuring robust data availability across SQL and cloud warehouses.

### Graduate Research Assistant,

#### Concordia University St. Paul | Nov 2023 – Mar 2024

- Built a **student retention prediction model** in Python (Scikit-learn), achieving 88% accuracy.
- Conducted research and co-authored a grant proposal on **AI ethics in healthcare**, contributing to a successful \$50K grant.
- 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)

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

#### Client Project: Charles Schwab (Financial Services)

- Built **agentic automation for ETL & anomaly detection**, improving efficiency of data pipelines.
- Explored **LLM-based chatbot workflows** and fine-tuning techniques with open-source transformer models.
- Applied ML models for **fraud detection and customer segmentation**, reducing risk exposure and improving client strategy.
- Reduced data processing time by 30% via optimized analysis pipelines in Databricks.

### Data Analyst

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

- Designed predictive models for **churn and demand forecasting**, ensuring data quality through rigorous EDA.
- Applied **statistical methods and PCA** for dimensionality reduction, strengthening model accuracy.
- Built and managed **ETL pipelines + scalable data warehouses** for business intelligence.
- Developed **interactive dashboards in Tableau** to track KPIs and support business decisions.

### 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, Coursera); Data Science & Cloud Computing (Infosys); Data Analytics (Cisco)