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 (basic); R (academic); C

Growth & Marketing Analytics:

- A/B Testing, User Segmentation, Paid Acquisition Analysis, Marketing ROI, Funnel Analysis, SEO Analytics, Time Series Forecasting

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 the deployment of an LLM-powered research assistant (GPT-4 + RAG), auto-generating market briefs and risk reports, **cutting analyst turnaround time by 50% and accelerating decision-making**
- Experimented with Neo4j-based knowledge graphs and integrated LLMs (GPT-4) for enhanced context retrieval and entity linking in financial documents.

- **Engineered and deployed autonomous trading agents** to monitor market anomalies and execute hedging strategies in real-time, improving risk-adjusted returns by 15% (backtested) and reducing manual intervention by 80%.
- Integrated generative AI (DALL·E 2) into investor pitch decks, **automatically creating bespoke visuals and infographics on demand**.
- Spearheaded end-to-end data science initiatives in financial analytics, **optimizing trading strategies and risk management processes**.
- **Architected a novel fraud detection system** that fused graph-based community detection (on transaction networks) with NLP sentiment analysis (on market reports), improving the precision of fraud identification by 20% over legacy models.
- Developed and optimized predictive models using supervised and unsupervised learning techniques, enhancing performance through **feature engineering and hyperparameter tuning**.
- Designed and deployed interactive, real-time visualizations in Tableau and Power BI that directly delivered actionable trading and risk insights to senior leadership.
- Designed and deployed clustering pipelines (K-means, hierarchical, DBSCAN) for customer segmentation and automated risk profiling, enabling a 25% lift in actionable insights for the product team
- Integrated neural networks and deep learning models (TensorFlow, Keras) for **advanced fraud detection and investment risk assessment**.
- Managed large-scale data pipelines and optimized ETL workflows, ensuring seamless data integration from SQL databases and cloud data warehouses.
- Implemented reinforcement learning models and AI-driven automation to **improve trading algorithms and risk mitigation**.
- Promoted adoption of big data frameworks for scalable analytics, ensuring efficient model deployment and monitoring.

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

- Developed a machine learning model in Python using Scikit-learn to predict student retention, achieving 88% accuracy and identifying key contributing factors.
- Conducted extensive literature reviews and data collection for a grant proposal on AI ethics in healthcare, contributing significantly to the successful submission of a $50,000 grant.
- Preprocessed and analyzed large-scale text data using NLP techniques (NLTK, SpaCy) in Python to identify sentiment patterns in social media discussions related to university events.
- Built interactive data visualizations in Tableau to present research findings on alum engagement to university stakeholders, facilitating data-driven 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)

- Developed prototype agent-based orchestration systems to **automate ETL pipelines and model retraining**.
- **Reduced hospital readmissions by 12%** by developing a patient segmentation model using K-Means and DBSCAN on claims data, enabling the business to launch targeted, high-risk cohort care programs.
- Engineered features from structured (ICD-10, CPT codes) and unstructured (clinical notes) datasets using Spark NLP and Python, **enriching inputs for downstream models**.
- **Built and productionized a suite of predictive models** (XGBoost, Random Forest) on GCP Vertex AI to identify high-cost claimants, achieving 85% accuracy and enabling proactive resource allocation for a major healthcare client.

- Applied BERT-based NLP to extract key clinical entities from provider notes, integrating insights into patient risk scores and care-gap analyses.
- Used PySpark and MLlib to process large-scale graph-structured healthcare data for patient relationship modeling and anomaly detection.
- Engineered and automated executive-level Power BI dashboards to monitor real-time model performance and KPIs (e.g., readmission rates, cost per claim), providing critical transparency to clinical and compliance teams.
- Built HIPAA-compliant ETL pipelines in Azure Data Factory and GCP Cloud Dataflow, harmonizing claims, EHR, and pharmacy data with encryption and RBAC.
- Forecasted claims volume using ARIMA and Prophet time-series models, **supporting capacity planning and budget projections** for the health plan.
- Devised anomaly-detection workflows with Isolation Forest in Databricks, **flagging suspicious billing behaviors and reducing improper claims by 9%**.
- Created Power BI dashboards to monitor KPIs (readmission rates, cost per claim) and track model performance, providing transparency to clinical and compliance teams.
- Partnered with clinicians, data engineers, and compliance officers in Scrum ceremonies to refine analytics solutions and maintain regulatory adherence.

Client Project: Charles Schwab (Financial Services)

- **Pioneered the firm's first agentic automation systems** for ETL and anomaly detection, demonstrating the business value of emerging AI trends to client leadership.
- Contributed to developing early LLM-based chatbot workflows using transformer models and explored fine-tuning techniques with open-source models.
- Built synthetic data generators with GANs for **privacy-preserving data augmentation** in healthcare analytics.
- Strategically applied supervised and unsupervised models for high-stakes fraud detection and customer segmentation, presenting findings from rigorous EDA to guide product strategy.
- Developed a high-performance data analysis pipeline that reduced data processing time by 30%, enabling faster model development cycles.
- Explored deep learning frameworks (TensorFlow, Keras) and reinforcement learning to **enhance AI applications**.
- Effectively communicated complex analytical findings to technical and non-technical stakeholders, fostering cross-functional collaboration.

Data Analyst

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

- Utilized advanced analytical techniques to **extract actionable insights and drive data-informed decision-making**.
- Developed and implemented predictive models using supervised and unsupervised learning techniques.
- Built and validated foundational predictive models for churn and demand forecasting; conducted detailed EDA to ensure data integrity for modeling.
- Applied statistical methods and PCA for dimensionality reduction to **refine model inputs**.
- Engineered robust data pipelines through data acquisition, validation, and transformation.
- Deployed scalable data warehousing solutions and managed end-to-end ETL processes.
- Designed and maintained Tableau dashboards to track key business metrics for departmental review.
- Researched deep learning (TensorFlow, Keras) and reinforcement learning methods in internal R&D settings to explore potential applications in customer behavior modeling and process optimization.

Education

Master's in Data Analytics
Concordia University, St. Paul, MN | Sep 2023 – Dec 2024
GPA: 3.91/4.00

Bachelor’s in Science
Nizam College, Hyderabad, India | Jun 2017 – Mar 2020
GPA: 8.22/10.00

Certifications

- Machine Learning – Stanford University (Coursera)
- Data Science and Cloud Computing – Infosys (Internal Certification)
- Data Analytics – Cisco (Training Program)
- Digital Marketing, SEO, Social Media Marketing (HubSpot)

Additional Information

Open to onsite, hybrid, or remote opportunities within the U.S.; willing to relocate as needed.

Open to an on-site Interview.