Chaitanya Dasari

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

Professional Summary

Machine Learning Engineer and **Data Scientist** with 4+ years of experience delivering high-impact AI solutions from concept to production. Expertise in predictive modeling, NLP (BERT), and Generative AI (LLMs), combined with advanced MLOps skills in AWS (SageMaker), Databricks, and data engineering (AWS Glue). Proven ability to drive measurable business value, evidenced by reducing patient readmissions by 12% and optimizing data pipelines for a 30% reduction in processing time.

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)

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,Snowflake, Redshift)

Statistical Methods
- Regression (Linear, Logistic), 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), Azure, Snowflake, Databricks

Data Visualization & BI
- Tableau, Power BI, Matplotlib, Seaborn, Interactive Dashboards

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

Professional Experience

AI Agent Evaluator (Independent Contractor)
Mercor Intelligence (Remote) | Sep 2025 - Present

- Evaluated frontier **AI coding agents** by running real-world software engineering tasks across large open-source repositories (200+ files).
- Conducted **Agent Code Comparison workflows**: iterative development, debugging, component refactoring, and backend enhancements.
- Benchmarked agents on task success, tool use, instruction following, and code best practices.
- Documented **behavioral issues** (tool undertriggering, hallucinations, agentic laziness) and delivered structured preference ratings.

Data Science Expert (Independent Contractor)
Mercor (Remote) | Aug 2025 - Sep 2025

- Designed and evaluated **rubrics** to assess AI/ML model outputs across data science tasks, including forecasting, hypothesis testing, and descriptive analysis.
- Built **golden responses and Python workflows** (pandas, visualization) to benchmark model accuracy and ensure reproducibility.
- Collaborated with reviewers to refine tasks, ensuring **objective, precise, and field-aligned criteria** for AI evaluation.

- Developed predictive models with feature engineering + hyperparameter tuning to optimize risk management.

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 **AWS SageMaker**.
- **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 (AWS Glue) 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)