

Ethan Caldwell

Senior Data Scientist - ML & GenAI

ethan.caldwell@example.com · +1 (415) 555-0142 · LinkedIn · GitHub

SUMMARY

Data scientist with 8+ years building production ML and GenAI systems across fintech and retail. Specialise in LLM agents, forecasting and large-scale data pipelines that move real business metrics - not dashboards nobody reads. Comfortable owning a problem end-to-end, from framing and experimentation to deployment and stakeholder buy-in.

EDUCATION

- Easton Institute of Technology**
B.E. Computer Science

2012 - 2016

SKILLS SUMMARY

- Languages:** Python, SQL, Scala, R
- Frameworks:** PySpark, LangGraph, LangChain, PyTorch, scikit-learn
- AI / ML:** LLMs, RAG, Forecasting, NLP, Time-series, A/B testing
- Cloud / Platforms:** Azure Databricks, AWS, GCP
- Databases:** Postgres, Snowflake, MongoDB
- Tools:** Airflow, MLflow, Docker, Git, Tableau

EXPERIENCE

- Northwind Analytics** Bengaluru
Senior Data Scientist 2021 - Present
 - Built a LangGraph QC agent that auto-validated ~500 factsheets/month, lifting discrepancy catch-rate from ~60% to 95% and saving 20 analyst-hours/week.
 - Shipped a what-if forecasting service that cut scenario modelling from 3 days to 2 hours for the planning team.
 - Led a media-attribution model that reallocated 15% of quarterly spend into higher-ROI channels, lifting digital share 4%.
 - Mentored 4 data scientists and set the team's code-review and experiment-tracking standards on MLflow.
- Brightford Consulting** Pune
Data Scientist 2018 - 2021
 - Automated a pharma ETL pipeline ingesting 10 source systems / 5M records a month, cutting data-quality failures by 40%.
 - Designed a dynamic pricing & segmentation model contributing a 3% improvement in net revenue realisation.
 - Migrated 50+ CSV sources (2TB) to a Hadoop/PySpark platform, reducing daily pipeline runtime from 6 hrs to 45 min.
- Veritas Labs** Hyderabad
Data Analyst 2016 - 2018
 - Built churn and propensity models that informed a retention programme reducing monthly churn by 1.8 pts.
 - Automated weekly executive reporting with SQL + Airflow, eliminating ~12 hrs of manual work per week.
 - Partnered with product to A/B test onboarding changes, lifting 30-day activation by 6%.

PROJECTS

- Naive RAG Toolkit (Python, LangChain):** Open-source RAG starter with a grounding-eval harness; 300+ GitHub stars across 20+ forks. Ships with reranking, citation tracing and a one-command eval against a labelled QA set.
- Forecast-Bench (Python, PyTorch):** Benchmarking suite comparing 8 time-series models on retail demand; published notebooks and a leaderboard. Reproducible Docker setup adopted by two internal teams for model selection.

HONORS AND AWARDS

- Speaker, DataWeekend 2023 - 'LLM agents for automated QC'.
- Top-5 finalist out of 200+ teams in the firm-wide GenAI hackathon.
- Two mentees promoted to Data Scientist within a year.

CERTIFICATIONS

- Certified Machine Learning Professional (DataCert Academy | 2023)
- Big Data Engineering Certificate (DataCert Academy | 2022)
- Deep Learning Specialist (NeuralPath | 2021)