

Adrian Cole

Data Scientist Specializing in
Machine Learning and Predictive
Analytics

456 Mission Street, San Francisco, California, United States, 94110

📞 (415) 555-0142

✉️ adrian.cole@example.com

🌐 <https://adriancole.dev>

🐙 @adriancole • 🌐 @adriancole-data

Basics

Data scientist with 6+ years of experience building machine learning systems that turn messy data into measurable business impact. Strong background in statistics, causal inference, and natural language processing, with deep expertise across the full model lifecycle—from problem framing and data pipelines to deployment and monitoring. Passionate about creating interpretable, reliable models and translating data insights into clear recommendations for product and leadership teams.

Education

Aug 2016–May 2018 **Master, Information and Data Science, Score: 3.9, University of California, Berkeley,** <https://www.ischool.berkeley.edu/>

- Concentrated on machine learning, causal inference, and large-scale data systems
- Capstone project: built a real-time churn prediction pipeline for a telecommunications partner
- Graduate research assistant supporting a study on algorithmic fairness in lending decisions

Courses: Applied Machine Learning, Statistics for Data Science, Natural Language Processing, Big Data Systems, Experimentation and Causal Inference, Data Visualization

Sep 2012–Jun 2016 **Bachelor, Statistics, Score: 3.8, University of Washington,** <https://www.stat.washington.edu/>

- Built strong foundations in statistical theory and computational methods
- Undergraduate research assistant in the Statistical Learning Lab

Courses: Statistical Inference, Regression Models, Probability Theory, Computational Statistics, Data Mining

Work

Mar 2021–Present **Senior Data Scientist, Lumina Analytics,** <https://www.lumina-analytics.com/>

- Lead development of predictive models for customer lifetime value across a portfolio of SaaS clients, improving average revenue forecast accuracy by 18%
- Architect feature store and model monitoring pipelines using Python and MLflow, reducing model retraining time from days to hours
- Partner with product and engineering teams to design A/B tests and causal inference studies for new feature adoption initiatives
- Mentor a team of four data scientists and establish best practices for code review, reproducibility, and documentation

Keywords: Python, MLflow, Feature Store, Causal Inference, Mentorship

Jul 2018–Feb 2021 **Data Scientist, HealthInsight,** <https://www.healthinsight.com/>

- Built risk stratification models using electronic health records and claims data, enabling early intervention for high-risk patients
- Designed and implemented an NLP pipeline to extract clinical concepts from unstructured notes, achieving 87% F1 score
- Created interactive Tableau dashboards used by clinical leadership to track quality metrics and reduce hospital readmission rates by 12%

Keywords: Healthcare, NLP, Tableau, Electronic Health Records

Languages

English Native or Bilingual Proficiency

Spanish Full Professional Proficiency

Keywords: IELTS 8.5

Keywords: DELE C1

Skills

Machine Learning	Expert	Keywords: Python, scikit-learn, XGBoost, TensorFlow, PyTorch
Statistical Modeling	Advanced	Keywords: R, Bayesian Methods, Time Series, Survival Analysis, Experiment Design
Data Engineering	Advanced	Keywords: SQL, Spark, dbt, Airflow, AWS, Snowflake
Natural Language Processing	Advanced	Keywords: Hugging Face, spaCy, Transformers, NER, Topic Modeling
Data Visualization	Intermediate	Keywords: Tableau, matplotlib, Plotly, ggplot2

Awards

- Dec 2023 **Lumina Analytics**, *Outstanding Contribution in Data Science*
Recognized for leading the development of a company-wide model monitoring platform that reduced unplanned model retraining incidents by 40%.

Certificates

- Nov 2021 **Amazon Web Services**, *AWS Certified Machine Learning - Specialty*, <https://aws.amazon.com/certification/certified-machine-learning-specialty/>
- Apr 2020 **Google**, *TensorFlow Developer Certificate*, <https://www.tensorflow.org/certificate>

Publications

- Aug 2022 **A Scalable Framework for Real-Time Customer Churn Prediction**, *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, <https://doi.org/10.1145/3523227.3547400>
- Presented a production-ready churn prediction framework combining gradient-boosted trees with online feature stores
 - Demonstrated a 23% relative improvement in AUC versus batch baselines across three industrial datasets

Projects

- Jan 2023–Jun 2023 **An NLP-powered search and question-answering tool for unstructured clinical documents.**, *DocQuery*, <https://github.com/adriancole/docquery>
- Developed a retrieval-augmented system using sentence embeddings and a fine-tuned transformer model to answer clinician questions from EHR notes
 - Deployed as a REST API with Docker, reducing average answer lookup time from 20 minutes to under 5 seconds
- Keywords:** NLP, Transformers, Q&A, Docker, FastAPI
- Aug 2020–Dec 2020 **Click-through rate prediction model for online advertising campaigns.**, *AdScore*, <https://github.com/adriancole/adscore>
- Engineered features from impression logs and built a distributed training pipeline in PySpark
 - Achieved a 15% lift in offline AUC and enabled real-time bidding through model export to ONNX
- Keywords:** CTR, PySpark, Feature Engineering, ONNX

Interests

- Open Source Contributing to scikit-learn, Jupyter, Streamlit
- Hiking Yosemite, Pacific Crest Trail
- Chess Tournament Play, Puzzle Solving

Volunteer

Mar 2022– **Data Science Volunteer**, *Bay Area Data Corps*, <https://www.bayareadatacorps.org/>

- Dec 2023 ○ Partnered with local nonprofits to analyze service utilization data and guide resource allocation decisions
- Built reusable notebook templates to standardize weekly reporting and make findings accessible to nontechnical staff