

BRENDAN C. SMITH

Lead Data Scientist

@ hire@brendansmith.ai 816-522-2998 linkedin.com/in/b-c-s github.com/brendancsmith

EXPERIENCE

Lead Data Scientist 09/2024 – 10/2025
Best Egg Remote

- Created a next-gen customer expansion XGBoost model for the Flexible Rent Platform including alternative data, increasing Gini coefficient by 23%.
- Designed and implemented Metaflow flows (DAGs) to automatically train challenger models for our primary credit risk and underwriting XGBoost model.
- Trained a team of 3 other Data Scientists to contribute flows to the project.
- Worked on a RAG-based chat system to generate answers based on our internal documentation. Also integrated with Snowflake to run SQL queries based on questions.

Senior Data Engineer – Contract 04/2024 – 07/2024
Burns & McDonnell Kansas City, MO

- Built Airflow DAGs to ingest full-load and incremental data from various SQL databases.
- Transformed records between medallion stages using dbt within the Databricks Medallion Architecture.
- Designed a logging strategy integrated with Azure Monitor / Log Analytics.

Graduate Learning Facilitator – Machine Learning 01/2024 – 04/2024
University of Texas at Austin Remote

- Served as a teaching assistant for a graduate-level Machine Learning course, supporting instruction, grading, and administration for a cohort of approximately 465 students.

Data Scientist 05/2023 – 08/2023
Propense.ai Remote

- Augmented enterprise knowledge graphs with internal data to cold-start a B2B recommendation system, solving the new-platform data sparsity problem for market insights.
- Identified patterns in sales gaps by analyzing client and sales history data. Presented actionable insights to clients, securing 5 initial contracts at launch.

Quantitative Investments & Data Science Intern 02/2023 – 05/2023
Nexus Equities Remote

- Developed a PyTorch computer vision model to estimate the useable land area on outdoor storage facilities using satellite imagery, accelerating approximately 50 investment decisions.

Software Development Engineer – EC2 Core Platform 10/2018 – 01/2021
Amazon Web Services (AWS) Seattle, WA

- Owned and operated two critical services: one to drain customer instances from unhealthy EC2 hosts, and another to proactively recycle older hosts for re-provisioning. Identified and resolved deadlock conditions, resulting in a \$300k/month reduction in 'unsellable' rate.
- Built a scalable data lake to centralize previously siloed EC2 internal data sources.
- Used PySpark to ingest real-time data from various AWS data stores such as AWS DynamoDB, RDS (MySQL), S3, and Amazon Athena.
- Deployed ETL pipelines using AWS Glue Jobs, Crawlers, and VPC Elastic Interfaces to extract cross-regional data from 300+ internal AWS production service accounts.
- Created stored procedures and QuickSight dashboards, increasing the velocity of insights and business decisions for TPM stakeholders.
- Implemented a capacity forecasting model and integrated it into the proactive re-provisioning workflow, increasing host turnover rate by up to 18% per region.

Software Development Engineer – HDInsight 08/2016 – 10/2018
Microsoft Azure Redmond, WA

- Served as cross-org liaison for root cause analysis of regressions in Apache Hadoop ecosystem products (Spark, Kafka, etc.), coordinating fixes with upstream Apache engineers.
- Proposed, designed, and deployed time series anomaly detection ML models on Azure, improving alarm triggers and identifying cluster configurations with high customer impact. Reduced average TTD by ~55% and TTR by ~20%.
- Joined a small v-team to refactor the HDInsight control plane, enabling flexible cluster shapes. Closed the feature gap with competitors while increasing service reliability KPIs and reducing COGS.

Design Studio Software Developer – Decision Science 09/2015 – 05/2016
Hudl Lincoln, NE

- Collaborated with a team of data scientists to develop a temporal convolutional neural network which automatically detected basketball highlights from user-uploaded video.
- Conducted continuous A/B testing to collect user metrics (engagement, bounce rate, stickiness, and view count) to iterate on model effectiveness.
- Assessed model launch success through user engagement metrics and presented regular reports to stakeholders. Achieved significant growth in active users (DAU / WAU).

EDUCATION

M.S. in Data Science
University of Texas at Austin
01/2022 – 12/2024

- Created DNNs using PyTorch for the vision system of a racing simulator. Implemented networks for object detection, keypoint estimation, semantic segmentation, multi-action networks, and reinforcement learning for autonomous driving.
- Created neural networks for Natural Language Processing (NLP) applications such as Semantic Parsing & Labelling, Sentiment Analysis, and Language Generation.

GPA | 3.89 / 4.00

B.S. in Computer Science
University of Nebraska – Lincoln
08/2012 – 05/2016

- Student in the honors program, Jeffrey S. Raikes School of Computer Science and Business Management.
- Minor in Business Management and Mathematics.
- Capstone project involved working with local businesses to implement software solutions.

GPA | 3.90 / 4.00

SKILLS

Languages & Frameworks

Python · SQL · R · PySpark · Java · JavaScript / TypeScript · C# · PyTorch · NumPy · Pandas · Polars · Scikit-learn · Matplotlib · Plotly · Seaborn

Machine Learning & MLOps

Deep Learning · Computer Vision · Forecasting · Recommendation Systems · Hyperparameter Tuning · ML Pipelines · Model Monitoring · Metaflow

Generative AI

LLMs · Agentic AI · Tool Use & Function Calling · Fine-tuning · RAG · Prompt Engineering · Embeddings · LangChain

Data Engineering

ETL Pipelines · Data Warehousing · Spark · Kafka · Hive · Hadoop Stack · Databricks

Cloud & Infrastructure

AWS · Azure · Databricks · Docker · REST APIs · FastAPI · SQL & NoSQL Databases · Microservices

Data Analysis

Statistical Modeling · Causal Inference · Time Series Analysis · Data Visualization