

Alan Feder

<https://www.alanfeder.com>

11402 Gainsborough Road, Rockville, MD 20854

alanfeder@gmail.com

(347) 625-9709

SUMMARY

I am an experienced data scientist with over ten years of expertise in harnessing artificial intelligence, machine learning, and statistical analytics.

I have developed models using tools like generative AI and natural language processing techniques, including transformer-based neural networks for classification and LDA for topic modeling . Additionally, I've utilized xgboost, generalized linear models, and time series analysis for structured data. I have experience in all stages of the data science project pipeline, from project conceptualization and exploratory data analysis to model training and deployment to production.

I am seeking a role where I can leverage AI & ML to design products, models, and solutions that benefit customers, employees, or the general public.

I live in the Washington, DC metro area and am open to working remotely or in-person locally.

EXPERIENCE

- **Invesco**
 - Senior Principal Data Scientist* *February 2023 - September 2023*
 - Principal Data Scientist* *September 2018 - February 2023*
 - **Created an LLM-powered tool** analyzing public comments on proposed government regulations
 - * Using GPT 4, classified stakeholder sentiment, pinpointing support or opposition
 - * Using GPT 4 combined with RegEx, grouped and synthesized arguments, identifying key themes and measuring consensus
 - **Enhanced internal chatbot** by enabling summarization and comparison of related PDF documents
 - * Utilized retrieval-augmented generation (RAG) through LangChain and Qdrant vector databases
 - **Devised a real estate-specific NLP-based sentiment signal** to forecast index prices across real estate sub-sectors
 - * Detected sector-wide price fluctuations 5 months ahead of market data recognition
 - **Organized initiatives** for internal Citizen Data Scientist program
 - * Oversaw capstone projects including the development of a new internal document search tool and cloud cost forecasting, leading to 20% increase in forecast accuracy
 - * Taught 12 investment professionals & 15 tech professionals lessons about machine learning
 - **Developed an ensemble predictive model** for multifamily residential real estate, using geospatial and machine learning programming tools
 - * Predicted which neighborhoods will increase in rent, outperforming the benchmark by 8%
 - * Utilized machine learning interpretability techniques such as partial dependence plots and SHAP to extract insights
 - **Fine-tuned neural network** (distilBERT) using Huggingface to classify the appropriateness of a news article for an internal ESG tool, achieving 92% accuracy and a 0.94 AUC
 - **Introduced a trending topic tracker** using LDA to highlight newly popular news & social media themes
 - * Enabled Invesco to develop new thematic ETFs at least 3 months before competitors
 - **Developed a natural language processing tool** to associate companies with given themes
 - * Constructed an internal word vector model using BlazingText
 - **Won an internal Kaggle-style competition** predicting outflows from Invesco mutual funds
 - * Leveraged categorical embeddings via neural networks for xgboost models and visualized them with t-SNE

- **AIG Science**

Data Scientist, Senior Manager

July 2014 - September 2018

- **Managed a team** of three junior data scientists and one junior business analyst.
- **Advised junior data scientists** developing a Random Forest model to predict behavior of third-party financial advisors, and improved F1 metric from 0.6 to 0.8.
- **Simulated insurance losses** that could be used by the investments division in order to match the risk liabilities
 - * Used Bayesian statistics and Markov Chain Monte Carlo simulation to build variance bands
- **Enhanced workers' compensation claims forecasting** boosting the accuracy of cash flow projections by 50%
 - * Derived actionable business insights from the model using marginal contribution plots
- **Streamlined travel insurance claims handling**
 - * Improved throughput by 30% by deploying a gradient boosting model (GBM), pre-approving low-risk claims
- **Built a set of automated analytical reports** for business leaders to assess whether a claims simplification process would benefit their line of business.

- **Swiss Re America**

Pricing Actuary

June 2010 - June 2014

- **Utilized Poisson regression** to innovate the pricing of mid-sized casualty commercial risk insurance policies
- **Built a mathematical model to calculate basis risk** in parametric insurance contracts for hurricane risks based on historical data
 - * Incorporated geographic and time series data into models, comparing the payout from a hypothetical parametric structure to actual property insurance payouts

TECHNICAL SKILLS

- **Languages & Tools** : Python, SQL, R, git, bash, AWS (S3, EC2, Athena), Airflow, markdown, quarto
- **Packages** : Huggingface transformers, langchain, openai, spacy, gensim, keras, pandas, scikit-learn, xgboost, streamlit
- **AI/ML Methods** : LLMs, BERT, prompt engineering, xgboost, LASSO, LDA, word2vec

EDUCATION

- **Graduate School of Arts & Sciences, Columbia University** New York, NY
Master of Science in Statistics *May 2010*
- **Columbia College, Columbia University** New York, NY
Bachelor of Arts - Major: Mathematics, Concentration: Economics *May 2009*

PRESENTATIONS, PUBLICATIONS, AND COMPETITIONS

- **Risk and Reward** : Elizabeth Cohen, **Alan Feder**, et. al. "[Stocks move on surprises: Using sentiment information for active portfolio management.](#)" vol. Q3 2022, Invesco, 14 Oct. 2022, pp. 21-25.
- **Data Science Salon** : [NLP in Finance: Beyond Predicting Alpha](#), March 2022
- **Data Science Salon** : [Machine Learning Interpretability: How to Understand what your ML Model is Doing](#), Feb 2021
- **rstudio::global(2021)** : [Categorical Embeddings: New Ways to Simplify Complex Data](#), Jan 2021
- **Kaggle** : 4th Place (out of 5,156) [Porto Seguro's Safe Driver Prediction](#), Nov 2017