Alan Feder

https://www.alanfeder.com alanfeder@gmail.com
11402 Gainsborough Road, Rockville, MD 20854 (347) 625-9709

Summary

I am an experienced machine learning engineer with over ten years of expertise in harnessing artificial intelligence, machine learning, and data science to drive value. I have led projects developing models using both supervised and unsupervised methods, including generative AI, natural language processing techniques, and structured predictive ML models.
I am looking to help businesses and government utilize AI/ML and data science methods to solve their needs.

Technical Skills

- **Languages & Tools** : Python, R, SQL, git, bash, AWS (S3, EC2, Athena), Airflow, markdown, quarto
- **Packages** : Hugging Face, ollama, spacy, openai, scikit-learn, xgboost, statsmodels, streamlit, pandas
- **AI/ML Methods** : LLMs, RAG, BERT, prompt engineering, xgboost, LASSO, GLM, GAM, word2vec

Experience

- **Alan Feder Consulting** *January 2023 - present*
  Founder
  - **Developed a knowledge base RAG-powered chatbot** for a cybersecurity company, built on an AWS EC2 instance, utilizing the open-source AI models *Mistral-7B* and sentence-transformers
  - **Created a predictive model** for identifying asset types in a cybersecurity context, leveraging advanced AI techniques
  - **Extracted and processed raw text, tables, and charts** from financial documents to support private equity market research, using GPT-4V, unstructured.io, and other AI methods
  - **Implemented an AI-based solution** for matching group names from disparate lists using embedding models and semantic similarity measures
- **Johnson & Johnson (contract)** *March 2024 - May 2024*
  AI Technical Consultant
  - **Designed and implemented an AI system** that automatically converts protocols from oncology clinical trials into Case Report Forms
    - Led to significant time and cost savings of over a month and $60,000 per clinical trial
- **Invesco** *September 2018 - September 2023*
  Senior Principal Data Scientist
  - **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
    - Explicitly included expert real estate knowledge by working with industry leaders to develop bespoke sentiment dictionaries and tune NLP parameters

        - Developed streamlit web app which allowed research team to explore the model and run their own backtesting
    - **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%
    - **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
- **AIG Science** *July 2014 - September 2018*
  Data Scientist, Senior Manager
    - **Managed a team** of three junior data scientists and one junior business analyst.
    - **Enhanced workers' compensation claims forecasting** boosting the accuracy of cash flow projections by 50%
    - **Streamlined travel insurance claims handling**
        - Improved throughput by 30% by deploying a gradient boosting model (GBM), pre-approving low-risk claims
- **Swiss Re America**
  Pricing Actuary *February 2012 - June 2014*
  Associate Risk Management Analyst *June 2010 - February 2012*
    - **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
    - **Analyzed basis risk** for extreme mortality bonds

Education

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

Teaching, Presentations, Publications, and Competitions

- **NYR Conference** : RAGtime in the Big Apple: Chat with a Decade of NYR Talks, May 2024
- **Bethesda Data Science Meetup** : RAGs to Richer Answers: Using ChatGPT to Query Documents & Limit Hallucinations, Nov 2023
- **University of Baltimore, Merrick School of Business** : Adjunct Professor, Fall 2023 - OPRE 605, Business Analytics
- **Risk and Reward** : Elizabeth Cohen, **Alan Feder**, et. al. "Stocks move on surprises: Using sentiment information for active portfolio management." vol. Q 3 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
- **Kaggle** : 4th Place (out of 5,156) Porto Seguro's Safe Driver Prediction, Nov 2017