Joseph Denby

jgdenby.github.io

joseph.denby0@gmail.com
+1 814 404 0817

SKILLS

- **Languages**: Python (`pandas, scikit-learn`, etc.), R (`tidyverse`), SQL, Stan/PyMC3, LaTeX
- **Tools**: LLMs, Airflow, Spark, PyTorch, AWS/GCP/Azure, Databricks, Github, Langchain, Looker/Tableau
- **Domain Knowledge**: Generative AI / NLP, Machine Learning, Deep Learning, Data Visualization

EMPLOYMENT

Navy Federal Credit Union — New York, NY
Lead Data Scientist |Python, SQL, LLMs, Langchain, Azure — *August 2023 - Current*

- **LLM Expert System:** Designed, built, and implemented a domain-expert GenAI Langchain chatbot using with state-of-the-art techniques in fine-tuning and RAG, deployed via Azure Containers. Allows employees to interactively query an internal knowledge graph of regulatory process documents.
- **LLM Cybersecurity Classification:** Built and deployed a multilabel document classification pipeline with a GPU-enabled fine-tuning of Llama3. Categorized 250k unstructured cybersecurity exploit descriptions with >99.5% success.
- **AML Watchlist Pipeline:** Using Python and SQL, constructed a data processing pipeline for screening of 350M+ bank customers. Designed address disambiguation and name parsing algorithms to reduce false negatives by over 60%.

Socure — New York, NY
Senior Data Scientist |Python, R, SQL, Airflow, Spark, Looker — *April 2021 - March 2023*

- **Selected Clients:** Wells Fargo, Capital One, Chime, DraftKings, Public
- **Pre-Sales Analytics:** Conducted hands-on statistical analyses in Python and R using massive financial datasets to uncover fraud. Developed, evaluated, and deployed machine learning classification models to guide $10M+ in sales deals through proof-of-concept analysis using Python, R, and SQL.
- **Pipeline Automation:** Automated machine learning model evaluation using Python and R. Queried 500GB+ datasets using SQL on hosted AWS instances to conduct 100+ model iterations without manual intervention.
- **Post-Sales Management:** Led a team of junior data scientists to generate $5M+ in value for clients using cutting-edge anomaly detection techniques. Built out real-time monitoring dashboards in Shiny and Looker to prevent fraud attacks on major financial institutions.

Merrill Lynch — Jersey City, NJ
Data Scientist |Python, SQL, NLP, Scikit-Learn, PyTorch — *September 2020 - April 2021*

- **E-Communication Supervision:** Built and deployed end-to-end machine learning pipeline using neural networks, including state-of-the-art transformer architectures (BERT, XLNet). Upgraded firm-wide communications compliance monitoring, leading to $5M+ in yearly savings.
- **Algorithmic Research:** Evaluated synthetic oversampling and anomaly detection techniques on algorithm development. Delivered concrete recommendations for future research to upper management based on expected value.

Decode_M — New York, NY
Data Scientist |Python, NLP, Webscraping, Bayesian Inference — *July 2019 - July 2020*

- **Selected Clients:** Airbnb, Lyft, Estée Lauder, Brooklyn Brewery, R/GA
- **MFactor:** Leveraged webscraping and sentiment analysis to generate novel statistical metrics to quantify brands' cultural relevance from online content. Built client-facing data dashboards using visualization libraries.
- **Conversation Analytics:** Leveraged unsupervised topic modeling & neural network document embedding (`Doc2Vec`) to offer brands strategic guidance at scale, uncovering nuanced online conversations from social media, news, and blogs.
- **Bias Detection:** Employed machine learning to extract low-level linguistic features from online documents and measure impactful drivers of conversation. Crafted a parallelized machine learning pipeline using AWS (EC2 instances), achieving a 24x reduction in runtime.

EDUCATION

University of Chicago — Chicago, IL
Master of Arts in Computational Social Science — *September 2017 - June 2019*

Cornell University — Ithaca, NY
Bachelor of Arts in Philosophy & Psychology — *August 2013 - May 2017*