

Danny Vilela

danny.pizza • danny.vilela@hey.com • github.com/dataframing • linkedin.com/in/dataframing

Experience

Reddit, *Machine Learning Engineer*

April 2023 – Present

- I don't know, I just got here.

Hulu, *Senior Data Scientist*

April 2022 – April 2023

- Designed, deployed, and tuned Disney+'s inaugural off-channel trigger-based recommendation system, targeting 400M+ US-based subscribers. Multiplied system coverage to include EMEA and enable global subscriber base, notably driving 3% increase in watch hours and reducing subscriber churn.
- Authored team-wide, interactive data science onboarding documentation via *Jupyter Book*. Deeply appreciated by new and existing team members as an intuitive resource for understanding internal data workflows and ecosystem.
- Led cohort of 10+ data scientists, engineers, and analysts through weekly reviews of *Full Stack Deep Learning – 2022*. Resulted in multiple members completing the official course and growing as machine learning experts.

Hulu, *Data Scientist*

August 2020 – April 2022

- Authored large-scale content representation pipeline, yielding significant performance and convergence improvements across multiple downstream personalization tasks. Published embeddings to an internally-accessible Python package.
- Led modeling for Live content personalization, driving a 6% increase in monetizable subscriber engagement hours.
- Optimized trigger-based recommendation system, increasing throughput by 60% and subscriber engagement by 7%.

Albert, *Data Scientist, NLP*

May 2019 – June 2020

- Deployed production-ready NLP models, reducing live annotation time by 62% and time to first response by 22%.
- Contributed back-end support for A/B testing, shadow/canary deployments, and live model performance monitoring.
- Authored an internal package for standardized machine learning tasks with 95% code coverage and static typing.

Applecart, *Data Scientist*

May 2018 – May 2019

- Developed and scaled de-duplication and entity resolution pipelines, enabling vital political targeting campaigns.
- Optimized internal pipelines via automatic parameter tuning, yielding an average F1 score increase of 12%.

Skills

Languages: Python, SQL, Bash, Go, R, Rust (for fun)

Development Tools: Docker (Engine, Compose), Poetry, Git, Jupyter (Notebook, Lab), conda, make

Cluster and Remote Computing: AWS (EC2, S3), Apache Spark, Databricks, DigitalOcean

Grit: relentlessly exploring, discovering, and implementing. Give me a dataset and a problem — I will figure it out.

Education

New York University, BA in Computer Science

September 2014 – May 2018