Jian Yang Lum

https://jianyanglum.github.io

408 Cole Street | San Francisco
+1 (310) 469-3694 | jian.yang.lum@gmail.com

Personal Summary

Machine-Learning-focused Data Scientist with strong end-to-end and production experience. Capable of finding insights in data and turning that into models with strong business impact.

Employment

- 07/2017-Present: Data Scientist, NerdWallet, San Francisco, CA, US
  - Online deployment of Likelihood of Approval (user-facing feature):
    - **Discovered features** and **wrote production ETL** for feature engineering
    - **Trained models** using ensemble trees
    - **Wrote production pipeline** (model deployment, retraining, metrics)
    - Result: **20% increase** in product conversions
  - Offline deployment of NLP-centric merchant-name standardization project
    - **Trained models**: Conditional Random Fields, modified preprocessing
    - Result: **85% standardized-merchant accuracy**
  - Offline deployment of Targeted Email Campaign Models:
    - **Trained models** and performed **feature discovery**
    - Result: 5x increase in email CTR, **3x increase in conversion rate**
  - User segmentation via event-based embedding approach
    - Trained `gensim`/`word2vec` along with dimensionality reduction
- 06/2016-09/2016: Data Science Intern, NerdWallet, San Francisco, CA, US
  - **Created auto-generating sitemap**, **designed and performed experiments on sitemap crawl rate**
  - **Created LDA-based comment filter** to alert editors for important comments left behind on articles
- 06/2015-08/2015: Data Analyst Intern, Beijing WuShuang Technologies (AnG)
  - Performed conversion path analysis (channel/time lag/keyword comparison) and uncovered relationships and patterns in the Chinese travel industry

Education

- M.S. Statistics (GPA 3.91/4), Stanford University
- B.A. International Relations with Honors (GPA 3.91/4), Stanford University

Technical Skills / Frameworks

- Languages: `python`, `R`, `SQL`, `bash`
- Data Munging: `pandas`, `tidyverse`, `SQL`, `pyspark`, `requests + beautifulsoup` (scraping)
- ML: `sklearn`, `keras`, `tensorflow`, `xgboost`, `nltk`, `gensim` (and more)
- Viz: `matplotlib`, `seaborn`, `ggplot`