Andrew Chang - Data Scientist

andrewjych@gmail.com - About Me - LinkedIn - GitHub - Emeryville, CA 94608

Languages, Tools, Skills

Languages: Python, C++, SQL
ML Frameworks: TensorFlow, Keras, Sci-kit Learn, PyTorch, xgboost, SpaCy, Spark NLP
Libraries: Pandas, NumPy, SciPy, Matplotlib, OpenCV, PySpark, HuggingFace (Transformers)
Databases: MySQL, PostgreSQL, SQLite
Tools: Git, Excel, phpMyAdmin, LaTeX

Work Experience

Fashom.com ("**Affordable Styled Clothing Delivered to Your Doorstep**") Data Scientist/ML Engineer (September 2021 - Present)

- Led the end-to-end development of a recommender system for internal use by product-focused teams.
  - Architected and engineered a ML pipeline, utilizing multiple ML models that consolidate over 5 sources of data, spanning 100k+ entries, into client-specific product recommendations.
  - Collaborated within the data science team to research, train, and implement a wide range of ML models.
  - Further deployed this workflow into a web app to serve to product teams. Data-driven product recommendations resulted in a 43% purchase rate and 3000% increase in workflow efficiency.
- Coordinated business intelligence efforts within the data science team, utilizing both ML methods and fundamental data structures/algorithms in order to generate insightful analytics on clientele.
- Built automated data pipelines for KPI extraction and low-level operations. This increased the sales rate by 10%, from 2021 to 2022, and increased client retention by 13% from Q3 to Q4 of 2021.

Projects

TeAMOFlow (GitHub, PyPI, Colab)

- A TensorFlow-based, Python library, serving as a modular, highly customizable, and scalable framework for deep learning and matrix factorization-based recommender systems.
- A generic ranking-based model attains up to 7.5% recall@10, 18% recall@30, 24.2% recall@50, on a 20% test split of the MovieLens 100k dataset.
- TeAMOFlow is built on Python, through TensorFlow, Keras, NumPy, SciPy. It is in active development.
- Hosted on GitHub and available on PyPI for installation into Python 3.7+ virtual environments.

Recommender System (RecoBot) (September 2021 - Present)

- A Python-based ML pipeline that generates recommendations. Consists of sequentially-placed data pipelines, and a variety of independently-trained NLP, collaborative/content-based filtering models.
- The data is sourced and queried from a MySQL database. The workflow is encapsulated in Streamlit and is served as a web application hosted on an AWS server, for use by client-facing product teams.
- The ML-specific implementations are done in TensorFlow, PyTorch, Sci-kit while the data wrangling and matrix/tensor manipulations are done with NumPy, Pandas.

Fashom BI (October 2021 - Present)

- A set of Python tools built to deliver actionable insights through business intelligence.
- Automated analytics pipeline that extracts sales-focused KPI's and posts the insights on a Google Sheets dashboard. Resulted in a 3600% increase in efficiency of KPI extraction while guiding business strategy.
- Consolidated influencer outreach efforts in a tree data structure, with quick node retrieval methods that generate insights on influencer-specific outreach and guide marketing strategy.

Education

University of California, Berkeley - B.A. in Physics, B.A. in Applied Mathematics (2016-2020)