Andrew Chang – Data Scientist

andrewjych@gmail.com – 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

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-facing teams.
  - Architected and engineered a multi-stage ML pipeline that consumes over 100000 data entries, across 5 sources of structured data, to inference, score, and filter relevant inventory.
  - Implemented this pipeline into a user-friendly web app, increasing the efficiency of the product team's initial workflow by over 3000% while achieving a 43% purchase rate in the first month of deployment.
- Collaborated within the data science team to research, train, and inference a wide range of machine learning models/techniques for implementation into the recommendation pipeline.
- Utilized extensive data-wrangling and machine learning techniques (image classification) in order to preprocess and clean company data, increasing the efficiency of this process by over 500%.
- 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

Recommender System (RecoBot) (September 2021 - Present)
- A scalable, multi-stage, machine learning pipeline that is trained and inferenced on 3+ years of transactions data and 8000+ unique items in order to generate personalized clothing recommendations.
- The machine learning pipeline built through Python (TensorFlow, mlxtend, Pandas, NumPy). Data pulled through Shopify API, MySQL (through SQLAlchemy). Frontend is implemented through Streamlit. Deployed as a locally-hosted web app on an AWS server.

Fashom Analytics (October 2021 - Present)
- An automated analytics pipeline that processes 3 sources of data, to generate insights on sales performance across different client demographics, increasing the efficiency of this process by 3600%.
- The analytics is performed through Python (Pandas, NumPy), while distribution to teams occurs on Google Sheets through various libraries and APIs (pygsheets, Drive API).

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 rank-learning 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. It is the only open-source TensorFlow 2.x library to implement a matrix factorization model using the Weighted Margin Rank Batch Loss (WMRB).
- 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.

Education

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