Yanan Cao

(+1) 415-605-3722 | yanancao34@yahoo.com | San Francisco, CA | linkedin-yanancao21 | llan-ovoy.github.io/ycao-Profile/

EDUCATION

UNIVERSITY OF SAN FRANCISCO, San Francisco, CA

Master of Science in Data Science Jul. 2022 (expected)

Courses: Advanced Machine Learning, Deep Learning, Data Structure and Algorithm, Distributed Data System (Spark, NoSQL), Relational Databases (SQL), Time Series, A/B Testing, Product Analytics and Communication.

BEIJING NORMAL UNIVERSITY, Beijing, China

Bachelor of Art in Management, minoring in Statistics Jun. 2021

Courses: Statistical Learning with R, Linear Model, Probability Theory, Stochastic Process, Mathematical Statistics, etc.

SKILLS

Programming Python (Pandas, NumPy, Scikit-Learn, PyTorch, Seaborn, Plotly), Spark SQL, NoSQL, Bash, R (ggplot2)

Platform & Tools AWS (S3, EC2, EMR), GCP, MongoDB, Hive, Databricks, PyCharm, Tableau, Git/GitHub

WORK EXPERIENCE

Walmart Global Tech Dec. 2021 – Present

Data Science Intern, Repurchase track, Personalization team San Francisco Bay Area, CA

Repurchase Prediction

- Created new features based on interval between orders with customer transaction data on Logistic Regression model for predicting items in customer's next order, achieved an average 0.01% improvement on precision so far.
- Queried and cleaned data using Hive Beeline and Apache Spark, and developed TIFU-KNN model from newly published paper for predicting items in next order, which outperformed Walmart's initial logistic regression model by 2% in top-three-item precision so far.

Personalized Seasonal Item Recommendation

- Analyzed customers' seasonal purchase behaviors for Easter products by processing and visualizing customer transaction data including side-by-side bar plot and stack bar plot with Seaborn library.
- Developed an item re-ranking algorithm to personalize Easter items recommendation, reaching an average 159% improvement of precision; a following A/B test will be launched at the beginning of Easter season.

DeGroote School of Business, McMaster University Jul. 2020 – Nov. 2020

Research Intern Ontario, Canada

- Worked on building an economic model based on price discrimination to research mobile Apps pricing strategy.
- Developed ETL processes for different price version information of more than 25,000 apps; scrapped data from websites using BeautifulSoup; designed a mobile Apps information database and inserted periodically collected data according to website's monthly update on each app.

Lenovo Group Feb. 2020 – Jul. 2020

Data Analyst Intern, Lenovo Voice, Software Development Team Beijing, China

- Built a pipeline for testing multi-language-based task-oriented dialog systems, including metrics for slots identification result with respect to different keywords in a command.
- Embedded sentences with pre-trained Chinese BERT, and built a oneVSrest Random Forest classifier for unstructured sentences, creating more training data with predicted labels and finally reaching a recall of 70%.

PROJECTS

Predicting with implicit feedback for Booking.com Feb. 2022

- Developed travel destination cities recommendation system with implicit user feedback, using collaborative filtering and Neural Network with weighted negative sampling technique based on city popularity.
- Tackled cold start problems by adapting a combination of randomly selected cities with a mean user.
- Ranked top 10 in Kaggle competition with private binary cross-entropy to 0.41.