Elaine Zhao

(+1)415-8232658 | jzhao54@dons.usfca.edu |Linkedin| github.io/Jinghui| San Francisco, CA 94111

EDUCATION

***MS in Data Science* - University of San Francisco** *July 2018 - June 2019*

Courses: Machine Learning, Deep Learning, Design of Experiments (A/B testing), Distributed Computing (Spark), SQL, Time Series, Data Structures and Algorithms, Linear Regression, and Data Visualization.

***BS in Engineering Management* - Beijing Jiaotong University** *Sep 2014 - June 2018*

WORK EXPERIENCE

Data Scientist Intern | Orange Silicon Valley | San Francisco, CA *Nov 2018 - June 2019*

- Applied a novel TFIDF algorithm on anonymized location data to detect similar users for ad targeting. Implemented the algorithm using Python and Spark.
- Calculated relative store location optimality by comparing user movements and travel patterns using a large dataset (4TB) of mobile user information processed on a 9-node Spark cluster.
- Created dynamic and interactive visualizations of users' locations with R and folium.
- Built statistical and neural language models for Wolof, a language used in the African market of Orange. Scraped web for Wolof corpora, cleaned the text (Python, CLD2), built n-gram models for text generation, and built bi-directional LSTM model to predict text sources. Achieved an accuracy score of 0.845.
- Discovered the abstract topics of Wolof corpora with Latent Dirichlet Allocation (Gensim).

Data Analyst Intern | China Telecom | Xi'an, China *Dec 2017 - Feb 2018*

- Categorized customers on a Recency-Frequency-Monetary basis with K-means. Designed strategies and services for each segment to increase customers' loyalty.

Data Analyst Intern | JD.com | Beijing, China *June 2017 - Aug 2017*

- Optimized Inventory by identifying economic order quantity and optimal reorder point. Conducted data wrangling and inputting in warehouse management system and order management system.

SELECTED PROJECTS

Dog Image Visual Search and Recommendation *[dogfind.us]* *Mar 2019 - May 2019*

- Designed a web application (Flask, Jinja, SqlAlchemy) which uses Convolutional Neural Nets to search dog shelters to help find missing dogs. Users upload photos of their pets, image similarity scores are calculated and matches are presented.
- Presented the product in a poster session to a panel of VCs.
- Implemented back-end development including web scraping (BeautifulSoup and Selenium), setting up an EC2 server and S3 file system, and managing users' profile data using PostgreSQL.

Air Quality Index Prediction *Jan 2019*

- Predicted nation-wide AQI for better air quality forecasting and human activity planning. Created features with Spark SQL, built ML regression models (XGBoost, Random Forest) achieving an RMSE of 15.13.
- Implemented distributed storage and computational system with S3, MongoDB and Sagemaker.

User In-App Purchase Prediction *Mar 2019*

- Estimated the probability of users making in-app purchases over 7-day and 14-day windows using over 60GB of data with XGBoost, lightGBM, and Random Forest models.
- Data was taken from over 3 million users of a mobile application provided by LeanPlum, a mobile marketing firm.
- Designed and created 64 features based on users' attributes and in-app behaviors with Pandas and Pyspark, achieved an AUC score of 0.97.

Style Transfer Application *June 2019*

- Implemented a style transfer model which takes in a photo from the user as well as an image from a particular artist and generates a new image in the style of that artist. (PyTorch, CNN).
- Improved upon existing models by applying gradient descent on the content/style loss functions.

SKILLS

- **Programming:** Python, R, Spark, PyTorch, C
- **Database:** SQL (PostgreSQL), NoSQL (MongoDB)
- **Tools:** AWS (EMR, Sagemaker, EC2, S3, RDS), Keras, Flask, Latex, Git, HTML, Jinja2, Excel, Tableau