Hongdou Li

✉ hli111@usfca.edu   ⌂ github.com/helenali323   ☐ 408-338-7908

EDUCATION

University of San Francisco   **San Francisco, CA**
Master of Science - Data Science   *Jun, 2018 - Jun, 2019*

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

University of Bath   **Bath, UK**
Master of Science - Economics   *Sep, 2014 - Nov, 2015*

University of Liverpool   **Liverpool, UK**
Bachelor of Science - Financial Mathematics   *Sep, 2010 - Jun, 2014*

PROFESSIONAL EXPERIENCE

kWh Analytics   **San Francisco, CA**
Data Science Intern   *Oct, 2018 - Jun, 2019*

- Estimated the lower quantile of solar panel performance to more accurately predict future energy production for insurance purposes (Python).
- Tested multiple models (Quantile Linear Regression, Gradient Boosting and Random Forest) and eventually improved accuracy by over 30% over previous baseline model.
- Built features, cleaned and joined multiple data sets including weather related data and internal, proprietary, solar panel production data (AWS, ec-2, S3, SQL).
- Productioniazed final model code.

Jiangsu Guotai   **Suzhou, China**
Export Merchandiser   *Feb, 2016 - Aug, 2017*

- Identified fashion trends and sourced local fabrics and materials for a variety of clothing products.
- Inspect goods ensuring uniform quality and negotiate with suppliers regarding their performance.

PROJECTS

MuiMui

- Built MuiMui, is a fashion recommendation app that suggests trendy and affordable clothes with a similar style to users' favorite Instagram posters.
- Scraped more than 30,000 images and related item information from various websites using Beautiful Soup and Selenium (Python).
- Application was hosted on AWS (ec-2, S3) and built using SQLAlchemy, Flask and multiple front-end tools (Jinja, HTML5/CSS, JavaScript).
- Recommendation engine was based upon similarity scores calculated between Instagram images and retail store product images.

In-app Purchase Prediction

- Built a gradient boosting classifier with XGBoost and a Convolutional Neural Network (CNN) model with PyTorch to predict the probability of whether a customer will purchase in the next 7 days and 14 days. Achieved an AUC score of 0.98.
- Feature engineering and model estimation were done on AWS (ec-2, S3) with 20 GB of data generated from users of a mobile application.

Auction Detection

- Built Machine Learning tools (Python) to predict whether an online bid was made by a human or a robot.
- Increased model performance by 50% by applying several machine learning algorithms with Scikit-Learn and and adopting an error metric more robust to class imbalances (AUC-ROC).

TECHNICAL STRENGTHS

Programing	Python (scikit-learn, pandas, numpy, SciPy, XGBoost, PyTorch), R, MATLAB
Database	SQL, NoSQL (MongoDB)
Tools	AWS (S3, EC2, EMR), Hadoop, Scala, Spark, Flask, NLP, Deep Learning
	Jinja2, HTML5, CSS3, Matlab, Git, Latex, Eviews
Data Visualization	ggplot, matplotlib, seaborn, Plot.ly, Tableau
Certification	ACCA (Association of Charted Certified Accountants)