Hanna Wang

ywang415@usfca.edu | (415) 602-9341 | LinkedIn | GitHub | Website | San Francisco, CA

EDUCATION

University of San Francisco   **San Francisco, CA**
Master of Science in Data Science   *July 2018 – Exp. June 2019*
- Coursework: Machine Learning, Deep Learning, Natural Language Processing, A/B Testing, Distributed Computing and Data Systems, Time Series, Visualization, Data Acquisition, Bayesian Analysis, Product Analytics, Marketing Analytics

University of Groningen   **Groningen, Netherlands**
Bachelor of Science in Business & Economics   *2016 - 2018*

Fudan University   **Shanghai, China**
Bachelor of Arts in Risk Management & Insurance   *2014 - 2018*

PROFESSIONAL EXPERIENCE

Fandom, Inc.   **San Francisco, CA**
Product Team, Data Science Intern   *Nov. 2018 - Present*
- Built a pipeline of multi-label text classification to help with content recommendation for the product team and advertisement targeting for the Ads team.
- Explored features that distinguished text by unsupervised topic modeling and visualization with PCA and t-SNE.
- Leveraged transfer learning from pre-trained word embeddings and improved F-score by 10% over the TF-IDF based model.
- Built Gradient Boosting and Neural Network models, achieved a 0.2 higher F-score than the benchmark of MTurk manual labels.

L.E.K. Consulting   **Shanghai, China**
Department of Strategy, Business Analyst Intern   *Aug. 2017 - Oct. 2017*
- Researched into Chinese hemodialysis (HD) service industry on multiple levels to support foreign entrance decision-making.
- Automated the data collection process using Python for keyword query searches and web crawlers.
- Forecasted market capacity in 5 years with time series models such as AR and ARMAX in R.

SELECTED PROJECTS

Product from Scratch - Manga Translation
- Developed a brand-new web app for manga translation and carried out market analysis on the US and global manga markets.
- Took lead of the back-end team, built a web server with Flask and organized user and image data by connection with AWS RDS and S3 databases.
- Built a model pipeline of detecting text bubbles (SSD), recognizing text from images (OCR API) and translation from Japanese to English (Google Translate API).

In-app Purchase Prediction
- Tidied 60+G raw data provided by Leanplum into 8G dataset using bash commands and pandas SQL queries.
- Generated features on user engagement, monetization, recency and trend during each weekly time window, and selected features with a forward elimination approach.
- Predicted user purchase in 7 or 14 days based on last two months' events, sessions, and user attributes data, and achieved a 0.985 AUC score using a LightGBM model.

Fraudulent Click Detection
- Detected abnormal users with an isolation forest from scratch and used their average depth among trees as a feature.
- Predicted click attributions with a Random Forest model, and achieved an AUC score of 0.96.

Social Network Analysis
- Pulled Twitter data from users' public profiles and activities using a self-coded Python web-scraper.
- Performed sentiment analysis on users' posts using NLTK and deployed a web server with Flask to display the searching and analysis results interactively.

SKILLS

Programming:	Python (Sklearn, Pandas, NumPy, SciPy, XGBoost, PyTorch, Seaborn, NLTK, Spacy, Gensim, Matplotlib, Plotly, Flask), Spark (SparkML, SparkSQL), R, C
Database:	SQL (PostgreSQL, MySQL), NoSQL (MongoDB)
Tools:	AWS (S3, EC2, EMR), SPSS, Tableau, html, Bash, Git, LaTex