Kyna (Feiran) Ji

San Francisco, CA 94105 | (510) 676-9602 | fji3@usfca.edu | kyna-ji.com

EDUCATION

University of San Francisco, M.S. Data Science **Jul 2017 – Expected Jun 2018**

- Coursework: Advanced Machine Learning, Data Visualization, Distributed Computing, Data Acquisition, Time Series Analysis, Big Data Business Strategies, Computational Statistics, A/B Testing, Web Analytics, Experiment Design

University of Southern California, B.S. Business & B.A. East Asian Studies **Aug 2012 – May 2017**

- Coursework: Statistics, Programming in Python, MATLAB, Relational Database, Data Mining, Operations Management

EMPLOYMENT

UBISOFT | Data Science Intern | San Francisco, CA **Nov 2017 – Present**

- Predicted users' purchase behavior for future games using machine learning techniques (Random Forest, Boosting and Logistic Regression), improving both the precision and recall rates by 20%
- Implemented data extraction and feature engineering using SQL and Python
- Deployed end-to-end pipeline to put the random forest model into production on Hadoop clusters using Spark ML
- Visualized insights about user segmentation using Tableau and presented to marketing team including VP and SVP
- Developed an interactive dashboard on a web application using Plotly and Dash in Python

DELOITTE CONSULTING | Analytics Intern | Shanghai, China **Jun – Aug 2016**

- Designed and built a business intelligence system using Oracle OBIEE for a leading media group in China
- Conducted market research, drafted project proposals and presentations to pitch Deloitte solutions

WIDSLEY | Intern | Tokyo, Japan **Mar – Aug 2015**

- Identified topics and words that increased click-through rates on our blog

PROJECTS

"Lego it" Web Application Development: Design a web platform that converts picture into LEGO bricks by matching appropriate color and shape of bricks, generating optimized assemble instruction and purchase plan by web scraping real time data from third-party sellers. (Link)
Tools: Python, Flask, MySQL Database, HTML, CSS, AWS, Github, Web Scraping, Dynamic Programming

Distributed Energy Usage Forecasting: Forecasted single day energy usage for each household with scalable machine learning models in Python. Deployed automated data pipeline to bridge Amazon S3, MongoDB on AWS EC2 and Amazon EMR. Optimized YARN cluster types and specifications based on computational time, cost, and predictive accuracy. (Link)
Tools: Python, Feature Engineering, Random Forest, Bash, AWS, Spark ML, NoSQL

Canadian Bankruptcy Rate Prediction: Predicted Canada monthly bankruptcy using multivariate time series models.
Tools: R, Time Series, Box-Jenkins, SARIMAX, Holt-Winters, VARX

Sales Data Integration: Cleaned, integrated and standardized messy data from different sales platforms into a PostgreSQL database using Python; Automated reporting of sales performance of products through different channels.
Tools: Python, SQL, Vertabelo, Data Warehousing, ETL

Artsy Targeted Promotions Analysis: Identified customers more attracted by coupon strategy and separated customers into different groups; estimated the performance of coupon promotions by conducting A/B testing
Tools: R, JMP, Tableau, Decision Tree Classification, K-means Clustering, Hypothesis Testing

PUBLICATION

Co-First Author. "Forecasting Smart Meter Energy Usage using Distributed Systems and Machine Learning".
(Under Review) ***The 16th IEEE International Conference on Smart City.***

SKILLS

Programming: Python(Sklearn, Pandas, NumPy, SciPy, XGBoost, PyTorch, Matplotlib), R, Spark, MATLAB
Database: SQL(PostgreSQL, Oracle, MS SQL Server, Teradata), NoSQL(MongoDB)
Tools: AWS(S3, EC2, EMR), Hadoop, Git, Latex, Bash, Tableau