Xinmin (Stacy) Wang

stacywang1225@gmail.com | (+1) (206) 369 3076 | LinkedIn

CAREER OBJECTIVE

Data scientist with deep understanding of building the data processing pipeline and infrastructure. Proficiency in using data mining algorithms, machine learning based predictive modeling and quantitative techniques to deliver insights from the data and implement action-oriented solutions to complex business problems.

CORE COMPETENCIES

- **Machine Learning:**
  - Regression: GLM, Ridge, Lasso, Kernel Regression, KNN
  - Classification/Clustering: Logistic Regression, Decision Trees, Random Forest, XGB, Gradient/Ada Boosting, SVM, Naive-Bayes, K-Means, K-Prototypes, GMM (EM algorithm)
- **Statistics:**
  - A/B Testing, Hypothesis Testing, Bayesian Inference, Probability
- **Programming:**
  - Python, SQL, R, Matlab, Java

PROFESSIONAL EXPERIENCE

Data Scientist

Lighter Capital, Seattle, WA / Apr. 2018 - Present

- Develop ML pipeline that automates feature selection, model construction and reporting process to evaluate borrower's monthly performance, largely reduce processing time by 100+ times (from weeks to hours).
- Construct scoring system to measure the risk of main deals. Reject suspicious deals with a precision of 95% and avoids around $500K loss per deal via model ensemble techniques applied on random forest, XGB and linear regression,
- Implement clustering algorithms such as K-Prototypes and GMM, build decision trees and create visualization using Matplotlib and Seaborn to explore main drivers of potential sales, thus improve the marketing conversion rate by 20%.
- Engineer 1000+ features and build GLM, regression trees to predict borrowers' baseline revenue and annual growth rate. The model reduces MAPE by 30%, increases overall yield by 2% and brings $800K profits to the company each year.
- Collaborate with software developers and UI designers to build data analytics platform that benefits both internal key stakeholders and external clients using Python and SQL.

Research Assistant Intern

Center for Education Data & Research, Seattle, WA / Jan. 2016 - Apr. 2018

- Processed and analyzed large-scaled time series education data in SQL and python, analyzed longitudinal data of teachers in US school districts and made data visualizations to infer rates and reasons for teacher mobility within states.
- Applied statistical inference and performed hypothesis testing on student performances data between two tests systems.
- Led group discussion and presented research results to aid policymakers in decision-making to improve education outcome.

PROJECTS

Interest Rate Prediction

- Performed variable selection procedures by ANOVA, identified leverage points, outliers and influential points, tested multi-colinearity and realized data visualization for 400000 observations via R.
- Applied GLM with PCA to predict interest rate using FICO score, revolving utility and total interest to data based on observations' past 5-year credit history.

Testing Stickiness of chewing-gums

- Designed complete randomized block experiment and performed null hypothesis testing to examine that if different standing time after chewing and different levels of pressure have no effect on stickiness.
- Adopted additive model and interaction model to investigate treatment and blocking effects.

EDUCATION

University of Washington, Department of Statistics, Seattle, WA

- M.S. in Statistics / Sep. 2016 - Mar. 2018
- B.S. in Mathematics (Dean's list) / Sep. 2012 - Jun. 2016