ZIHAN YE

Data Scientist with Product Management Background

@ zihanye96@gmail.com 718-751-6187 New York, NY linkedin.com/in/zihanye github.com/zihanye96

EDUCATION

M.A in Statistics
University of California, Berkeley

August 2019 – May 2020 Berkeley, CA

- **Coursework:** Applied Machine Learning, Advanced Probability, Mathematical Statistics, Linear Models, Statistical Computing, Pedagogy of Statistics and Probability
- **Cumulative GPA:** 3.90/4.00
- **Awards:** Betty Scott Excellence Scholarship
- **Activities:** Graduate Data Science Organization, Statistics Graduate Students Association, Board Games at Berkeley

B.A in Statistics & Asian Studies
Williams College

August 2014 – June 2018 Williamstown, MA

- **Coursework:** Discrete Math, Linear Algebra, Real Analysis, Abstract Algebra, Data Mining, Bayesian Statistics, Statistical Inference, Categorical Data Analysis, Regression & Forecasting, Game Theory
- **Cumulative GPA:** 3.79/4.00
- **Awards:** Mu Sigma Rho (Statistics Honor Society), Dean's List
- **Activities:** Information Technology Committee, Student Math and Statistics Advisory Board, Asian Dance Troupe

WORK EXPERIENCE

Graduate Student Instructor
University of California, Berkeley

August 2019 – May 2020 Berkeley, CA

- Taught statistical topics such as: regression, probability, confidence intervals, and hypothesis testing to a non-technical audience.
- Received the 2019-2020 Outstanding Graduate Student Instructor Award in recognition of excellence in teaching.

Product Manager
MBI, Inc.

August 2018 – August 2019 Norwalk, CT

- Managed competing project deadlines of 70+ consumer products totaling over $1MM in annual sales with advertising expenditures of over $200M.
- Developed and implemented marketing strategy by analyzing market segments and performance data to achieve industry-leading profit margins via direct mail and digital platforms.
- Collaborated with cross-functional teams to conceptualize and execute creative initiatives, including product development and marketing materials.

SKILLS

- **Tools:** Python, R, SQL, Bash, git
- **Libraries:** Pandas, Numpy, Matplotlib, sklearn, Keras, dplyr, ggplot2, Shiny, Caret

PROJECTS

Classification of Toxic Comments

- Constructed various classification models to correctly identify toxic online comments over 93% of the time (Naive Bayes, Random Forest, Logistic Regression, Neural Network, SVM).
- Utilized feature engineering techniques and problem context to develop 11 intuitive features used for modelling.

Cluster Analysis of Digital Art Collection

- Clustered the digital art collection of the Williams College Museum of Art (WCMA) via K-means clustering.
- Developed a web application (Shiny) to give users customized recommendations from the WCMA collection based on their artistic preferences.
- Analyzed image data for over 5000 digitized artworks to create new features that capture (dis)similarity outside of traditional criterion (ex. artist, date, and genre).

Fair Value Estimator for Stock Investing

- Created a Jupyter (IPython) notebook to automate stock valuation process by using Selenium to manually scrape financial data from Morningstar.com.

Predicting Annual Income on Airbnb

- Constructed a model to accurately predict annual income for listings on Airbnb.com using web-scraped data.
- Researched existing literature and utilized data cleaning and feature engineering techniques to prepare dataset for modeling (OLS, LASSO, Random Forest, GBM) and inference.

R Package: Genetic Algorithm for Variable Selection in Regression

- Led a team of 3 to create an R package that implements a genetic algorithm for variable selection in regression using modular functions.