Zihao (David) Xu

xuzihao2015@gmail.com
Project Repo: https://github.com/zihaoxu/My_Portfolio (310) 962-6992

EDUCATION

Harvard University **Cambridge, MA**
Master of Science in Data Science, GPA: 4.0 *December 2020*

- Related coursework: Data Science, Advanced Analytics of Finance, Investment Strategies, Natural Language Processing, Visualization, Stochastic Methods for Data Analysis, Machine Learning

Pomona College **Claremont, CA**
Bachelor of Arts, Computer Science & Mathematics, GPA: 3.96/4.0 *May 2019*

- Magna Cum Laude, Phi Beta Kappa Member, Sigma Xi Member, Pomona Scholar
- Related coursework: Artificial Intelligence, Algorithms, Computational Statistics, Big Data, Stochastic Operations Research, Probability, Statistical Theory, Bayesian Statistics, Econometrics

WORK EXPERIENCE

McKinsey & Company **Waltham, MA**
Data Scientist – Model Development *Feb 2021 - Current*

- Independently implemented a branch optimization algorithm in Python as a firm asset; adapted model to client scenario and built production-ready data pipelines w/ 10+ data sources to project $15M of net profit
- Performed comprehensive validation of 2 statistical wholesale credit models, including assessments of model input, conceptual soundness, output performance, and ongoing monitoring
- Advised a top US banking institution on data transformation strategies, spearheaded the revamping of 2 data capabilities including metrics, and controls framework including data risk taxonomy
- Built interactive visualization dashboards using Tableau for top executives to analyze competitive landscape

Cornerstone Research **Los Angeles, CA**
Summer Analyst *Summer 2018*

- Cleaned, managed, and visualized multi-year vehicle transactional datasets to advise two major auto manufacturers on vehicle recall strategies
- Performed large-scale OCR and sentiment analysis over 1000+ regulatory correspondence to inform regulatory response

Pomona College Mathematics Department **Claremont, CA**
Summer Researcher *Summer - Fall 2017*

- Devised and implemented a new machine learning algorithm - Bag of Little Random Forests (BLRF) as an R package, utilizing a paralleled structure for faster computation
- Evaluated and visualized the statistical and computational performance of BLRF using simulated data sets

TECHNICAL SKILLS & EXPERIENCE

Tech Projects: CornBERT: Developed a transformer-based language model for multi-label gene expression level prediction using gene sequences, pioneering application of NLP techniques in genetics

CNN-Eye-Tracker: Built a robust appearance-based webcam eye-tracking tool using CNN, improving over previous methods by ~25% in prediction performance

Tech Skills: Python, highly proficient in all major DS/ML packages; version control with git; working knowledge of Machine Learning & Deep Learning (including NLP and CV); data visualization with Tableau and Python

Publications: B. Levy, **Z. Xu** et al. (2021), *FLORABERT: cross-species transfer learning with attention-based neural networks for gene expression prediction*, work in progress
Z. Xu, M. Salloum (2018), *Deep Neural Networks for Object Enumeration*, poster paper, 2018 IEEE International Conference on Big Data, available at https://bit.ly/34DXTLb
Z. Xu, J. Hardin (2017), *Bag of Little Random Forests (BLRF)*, First Prize, 2017 Fall USRESP Competition, available at https://bit.ly/2kecdFE