Tony Ding

xiayiding@hsph.harvard.edu · 213-245-5570 · LinkedIn· Personal Website

EDUCATION

Harvard University *08/2022 – 05/2024 (Expected)*

Master of Science in Health Data Science

- GPA: 4.0/4.0; Recipient: Lyman & Grew Memorial Scholarship AND Harvard Central Grant Scholarship
- Cross-registration at **MIT**: *Artificial Intelligence and Decision Making*; GPA: 5.0/5.0

University of Southern California *08/2018 – 05/2022*

Bachelor's Degrees in Data Science and in Neuroscience

- Data Science major GPA: 4.0/4.0; Cumulative GPA: 3.9/4.0
- Presidential Scholar (half-tuition awarded throughout 4 years); Renaissance Scholar Distinction

RELEVANT EXPERIENCE AND PROJECTS

Data Science Team Member, **MIT Hacking Medicine**, Cambridge, MA *10/2022 – Present*

- Initiated and currently working on a project to predict re-admittance rate for diabetes patients while also examining patients' demographic integrity. I mainly implemented multinomial logistic regression, Random Forest, and XGBoost and am currently using various metrics and methods such as AIC and SHAP to evaluate our models and their covariates.

Data Scientist Intern, **AstraZeneca**, Shanghai, CN *05/2021 – 08/2021*

- Built a supervised machine learning program to predict users' behavior on AZ_MedInfo (a medical information exchange platform with over 1M users), like users' follow or unfollow activity. Increased the AUC score by 26.3% and decreased the error rate by 28.5% by applying and tuning a Random Forest classifier. Successfully identified 12 out of 161 most significant and meaningful variables that impact users' decisions.
- Designed a weighted association rule mining program, for Prof. Binghe Xu, MD, by extracting keywords "ctDNA" and "Breast Cancer" from publications' titles and abstract sections to identify and rank his associations with fellow colleagues and determine his individual academic rankings among all researchers in that field.

Data Science & Visualization Intern, **Takeda Pharmaceutical Company**, Cambridge, MA *05/2020 – 08/2020*

- Used Python to pre-process the clinical data and implement unsupervised machine learning models on raw baseline data (variables regularized using LASSO in R) for PANDA, a Takeda's oncology program for Ponatinib. Successfully identified 4 significant patient subpopulations for Ponatinib and the key risk drivers ($p<0.001$) of MACE (Major Adverse Cardiac Event) occurrences among Chronic Myelogenous Leukemia patients in the clinical trials.
- Utilized MS SQL Server Management Studio and Excel to quantitatively analyze Takeda R&D partnerships around the globe. Used the analytical results to create a new R&D partnerships visualization paradigm for Takeda by designing 8 elaborate time-based interactive network visualizations.

Data Analytics Intern, **MeetSocial Group Digital Technology**, Shanghai, CN *05/2019 – 08/2019*

- Implemented supervised machine learning, including kNN and SVM, to classify potential business sizes for new incoming clients and predict optimal advertising patterns for them. Increased the click-through rate of advertisements by 68%.

RESEARCH EXPERIENCE

Statistics Research Assistant, **USC Health, Emotion, & Addiction Lab** *12/2020 – 07/2022*

- Wrote thousands of rows of SPSS syntax and conducted various statistical tests and analysis in R for ADVANCE (*Assessing Developmental Patterns of Vaping, Alcohol, Nicotine, and Cannabis Use and Emotional Well-being*) School Reports project to determine the significance of associations among variables. Calculated RCADS (Revised Children's Anxiety and Depression Scales) scores from hundreds of variables for each participant. Applied syntax to and created analytical reports for all 6 high schools in the Greater Los Angeles area.

LEADERSHIP EXPERIENCE

Data Team Lead, **LinkedIn Campus Editing Team, USC** *09/2020 – 05/2021*

- Cleaned data sets using Python and built decision trees in RapidMiner to analyze post-graduation career plans and job offers received by graduate students at USC across all academic fields. Built visualizations and dashboards on Tableau and demonstrated findings at the annual Post-graduation Outcome event hosted by USC Career Center.

ADDITIONAL SKILLS & AWARDS

- Expert in Python, R, and SQL; Proficient in Java, machine learning and data mining algorithms and libraries
- Proficient in database modeling and ETL; Proficient in using SPSS, MySQL, MongoDB, Tableau, and Excel
- Alpha Prize (2nd Place in the World) in AoCMM Mathematical Modeling Contest