Xiyu Yang

993 Memorial Drive • Cambridge, MA 02138 • xiyuyang@g.harvard.edu • 608-405-2394
LinkedIn: https://www.linkedin.com/in/xiyu-yang-61a065194/

EDUCATION

Harvard University   Cambridge, MA
MS, Data Science. GPA: 4.0/4.0   May 2023
Coursework: Introductory Data Science, Advanced Topics in Data Science, Linear Models, Generalized Linear Models, Probability, Systems Development, Critical Thinking in Data Science.

University of Wisconsin-Madison   Madison, WI
BA, Statistics, Sociology and Minor in Mathematics. GPA: 4.0/4.0   May 2021
Honors in the Liberal Arts, Ralph B. Abrams Scholarship, Thomas W. Parker Scholarship

SKILLS

Programming: R, Python (pandas, NumPy, scikit-learn, TensorFlow), Git
Visualization and database: Tableau, SQL

DATA SCIENCE EXPERIENCE

Graduate Researcher   Dec 2021 – Present
Political Science Research Group, Harvard University   Cambridge, MA
- Critiqued the R package structure and proposed efficient plans for improvements on Bayesian prediction of ethnic category using surname and geolocation.
- Improving package compatibility by incorporating Census data from recent years through Git Version Control.

Statistics Project Team Member   Oct 2021 – Dec 2021
Linear Models, Harvard University   Cambridge, MA
- Investigated 9 socio-economic factors closely related to happiness score at the country-level.
- Constructed 3 ordinary least square models and 2 linear mixed-effect models to evaluate the feature importance; conducted 2 ANOVA tests for model selection.
- Generated a 10-page summary report presenting the relationships between country-level happiness score and identified important features (e.g., GDP and life expectancy).

Data Science Project Team Member   Sep 2021 – Dec 2021
Introductory Data Science, Harvard University   Cambridge, MA
- Identified 6 socio-economic factors contributing to high crime rate in Boston and Cambridge areas.
- Merged data from 4 distinct sources and transformed over 200k raw crime data into tidy format in Python.
- Constructed a XGBoost model which outperformed the baseline model, a single decision tree, by 8%.

RESEARCH EXPERIENCE

Research Assistant   Oct 2020 – Aug 2021
Statistics Research Group, University of Wisconsin-Madison   Madison, WI
- Identified key factors that contribute to the growing public attention of QAnon community on Twitter.
- Collected 100k raw text data and transformed it to clean time series data for each Twitter user using R.
- Conducted imputation to fill in 10% missing values with linear interpolation and ARIMA modeling using R.
- Coauthored a paper accepted by the 72nd Annual International Communication Association Conference.

Undergraduate Researcher   Sep 2020 – Dec 2020
Statistics Department, University of Wisconsin-Madison   Madison, WI
- Collaborated with three PhDs to study the citation patterns of the first paper on false discovery rate (FDR).
- Clustered, and contextualized citation network data with 220 million papers and over 100Gb meta data by applying algorithms, Vintage Sparse PCA (VSP) and the best feature function using R.
- Discovered distinct diffusion patterns for FDR across research fields and presented the findings to a professor, PhDs and master's students.