Yuxin Wu

206-218-5130 | xin1116@uw.edu | linkedin.com/in/yuxin-wu-154b3318b

EDUCATION

09/2021 - 08/2023 *(expected)* **M.Sc. in Biostatistics, University of Washington, Seattle** GPA: 3.68/4.00

- Major coursework: Biostatistics, Statistical inference, Software Development for Data Scientist, Database systems, Artificial Intelligence, Non-Parametric Regression & Classification, etc.

09/2019 - 06/2020 **Exchange student**, **University of Washington, Seattle** GPA: 3.80/4.00

- Major coursework: Java programming, Machine learning, Probability, Numeric computing, Resampling inference, Multivariable calculus, etc.

09/2016 - 06/2021 **B.Sc. in Science, Sichuan University, China** GPA: 3.64/4.00

SKILLS

Programming Skills: Java, Python, SQL, R, MATLAB, Git
Modeling Techniques: Natural Language Processing, Reinforcement Learning, Feature Engineering, Data Mining, Random Forest, PCA, KNN, SVM, Regression & Classification

EXPERIENCES

06/2021 - 09/2021 **iCarbonX** (*ranked top 1st in China AI Startups Ranking by Nanalyze 2017*)**, China**
Data Analyst Intern

- Applied PCA on peptide data and developed regression models to find reported and potential biomarkers
- Used correlation analysis to find similarly of gene info of clinical data to that from high-dimensional genomic data (more than 130,000 peptides) for new hypothesis generation
- Used random forest to build a classifier of the clinical metadata into hepatitis, cirrhosis, hepatocellular carcinoma (HCC), conducted feature engineering with H2O in R, identified unique patterns of HCC with high AUC (~0.75), improved the early detection of HCC

06/2020 - 08/2020 **University of Notre Dame**
International Summer Undergraduate Research Experience (iSURE) program, Lab for Big Data Methodology

- Developed an classifier to identify invalid responses in questionnaires by analyzing response time data
- Used change point analysis to determine number of change points and estimate the time of each change
- Employed cross correlation analysis to measure similarities of different respondents
- Utilized dynamic time warp to perform nonlinear warping on series where diffs in time are not penalized

06/2019 - 09/2019 **University of Alberta, Canada**
Visiting Researcher, Department of Biostatistics

- Resampled CancerSEEK patients data using prevalence rate of cancers and cancer's stage distribution
- Built logistic regression models and performed 10 times 10-fold x-validations on data after resampling
- Re-evaluated the CancerSEEK's diagnostic accuracy using AUC and Average Precision

SELECTED PROJECTS

09/2021 - 12/2021 **StateLegiscraper** *https://github.com/YuxinSIA/StateLegiscraper*

- Used Selenium to scrape websites for standing committee hearing text, audio, or video transcript data
- Employed NLTK and Textblob for text analysis and applied TF-IDF and Sentiment Analysis on raw text
- Applied semantic search with SentenceTransformers to let users decide the topic(s) of interest
- Built an interactive dashboard with Plotly Dash to build greater public transparency and academic knowledge about public policymaking and state-level politics