Yile (Abby) He

Cell: (530)-760-8509 | Email: yileabbyhe0425@gmail.com | LinkedIn | GitHub | Seattle, Washington

EDUCATION

University of Washington, Seattle, United States September 2022 - June 2024
Master of Science, *major in Biostatistics*

University of California, Davis, United States September 2018 - June 2022
Bachelor of Science, *major in Statistics*
- Cumulative GPA: 3.98/4.00
- **Relevant Courses**: Linear Algebra, Nonparametric Statistics, Regression Analysis, Sampling Theory, Math Statistics, Statistical Data Science, Statistics Data Technologies, Analyze Categorical Data, Introduction to Programming, Math for Computer Science, Calculus

INTERNSHIP

Research Assistant, Advisor: Prof. Hao Cheng, University of California, Davis August 2021- June 2022
Learning Evidence of Conservation from Integrated Functional Genomic Annotations (LECIF)
- Improved the **LECIF** model by increasing the number of species and tissue types for efficient gene alignment.
- Performed a **deep learning model** on 3 species with 4 tissues to study species conservation at the genomic level.
- Built and trained an ensemble of **neural networks** using 1M+ training functional genomic annotations in **Python**.
- Used **ROC curve** with 100,000 testing functional genomic annotations, model achieving 85% accuracy.
- Performed the new model to study species conservation at the genomic level for pig, mouse, and human species.
- Performed a **neural network** method on 3 species with 4 tissues to study species conservation at the genomic level.

Research Assistant, Advisor: Prof. Hao Chen, University of California, Davis August 2021-April 2022
The Multivariate Probability Distributions Project
- Built functions to compare the performances of three selected **nonparametric** methods: MCM, MMCM, and a new model from the lab, on 70,377 T cell data extracted from three locations: normal tissue cells, blood cells and tumor cells for 86 pathways.
- Wrote a **Python package** using **API** to access genetic code information for each pathway on the KEGG website.
- Performed data visualization of the results via **ggplot2**
- Concluded the MMCM model has a better performance in differentiating T cells extracted from different locations.

PROJECT

New Approach to Quantify Covid-19 Policy Effectiveness and Related Factors August 2020-December 2021
- Built new **R** model to quantify policy effectiveness on COVID in 26 states for vaccination policy and 18 states for mask policy.
- Investigated the factors related to COVID-19 control policies through the **General Linear Model** in **R**
- Concluded factors that are related to mask and vaccination policies which can serve as reference for future covid policymaking.
- Present in **JSM 2022** public health policy section. (Link to the slides) (Link to the paper)

Various Sentiments Related to Covid January 2022- April 2022
- Accessed the database of YouTube, Reddit, CDC, and New York Times through their **API**.
- Applied **Sentimental Analysis** to 2k+ comments from YouTube and Reddit on Covid-19 vaccine through polarity test from SentimentIntensityAnalyzer in **python**.
- Visualized the correlation between COVID-19 and the number of articles reporting anti-Asian hate in **matplotlib,** and the positive and negative comments on the COVID-19 vaccine through **wordcloud**.

ANALYTICAL SKILLS

Programming Skills: Python (Numpy, Pandas, Matplotlib, Scipy, Tensorflow, JSON), R (dplyr, ggplot2), SQL, MATLAB
Application: MS Office, A/B test, Machine Learning (Logistic Regression, Random Forest, XGBoost), Tableau, Jupyter Notebook, HTML
Certification: Google Data Analytics Specialization, Specialization in Machine learning