Yiting Xu

(404)-319-0491 • yxu37@student.gsu.edu • www.linkedin.com/in/yiting-TIA • ATL, GA

OBJECTIVE & KEY COMPETENCIES

Creative and analytical problem-solver seeking a **Data Scientist**, **Data Analyst** or **Statistician** full-time position.
- Data Analysis Knowledges: Statistical Analysis, Data Mining, Machine Learning, NLP, Database System, ETLs, Modeling, Data Visualization.
- Software & Tools: Proficient in **R, SQL, Python,** Tableau, Power BI and MS office. Intermediate level of SAS**.** Experienced in **AWS**. Win/Linux

WORKING EXPERIENCE

Georgia State University – *Statistician*   *Jul 2020 – Present*
- Examining Mouse with liver metastasis and lung fibrosis to determine the disease features, treatment modalities and survival outcomes. Built data collection table and applying **descriptive statistical analysis** for the data in **R**. Visualizing the result and drawing conclusion with **Tableau**.
- Correlating diseased mouse MRI images to H&E images with Pearson Correlation Test and Linear Regression in **R**. Conducting pixel and **AUC analysis** for the image data and making the comparison btw groups to see the disease change through time.
- Classifying the tumor stage images with logistic regression and Creating **R shiny** dashboard. Organizing result and preparing for publication.

Boehringer Ingelheim – *Data Scientist CO-OP*   *Feb 2020 - Jun 2020*
- Collaborated with non-technical group. Understood business demand and assisted with data-related technical issues. Based on the need, designed and lead a protein classification big data AI project using different Machine Learning tools especially the Deep Learning Model to help experiment.
- Accomplished data extracting, transforming and loading (**ETLs**) with **Oracle** and **R** to over 20k raw data entries. Organized data into structure.
- Applied Naïve Bayes, SVM, K-means Clustering, Random Forest and other **Machine Learning** algorithms to datasets for data classification. Achieved highest prediction accuracy over 80% (with SVM). Set as a base line.
- Designed and implemented Convolutional Neural Network (**CNN**) model with Keras & TensorFlow in **R**.
- Set up **AWS EC2** cloud service to train the model. Raised the model's accuracy by big percentages from 65% to 91%. Automatically seek out Fales-positive & False-negative samples for scientists. Saved 3 months' experimental time.

Boehringer Ingelheim – *Data Analyst Summer Intern*   *Jun 2019 - Aug 2019*
- Extracted patients' clinical data with **MySQL**. Built data infrastructure in **Python**.
- Visualized the data to show the difference in patients' sex, age, race, disease stage and other factors distribution with **Power BI**.
- Selected variables with backward approach. Constructed Logistic model in **Python** to predict 10 years risk of heart disease and applied Chi-Square test as well as Wald Chi-Square test to estimate odd ratios.
- Created a report based on the visualization and model prediction result, built dashboard, made poster and gave a presentation to both the technical (IT department) and non-technical team (R&D department) to explain the result.

PROJECT & PUBLICATION

Project from CDC: Medical Service Cost & Utilization Analysis and Prediction
- Studied on the difference in utilization and cost of STD-related Medical Services between people who had anal sex, a risk factor of HIV, and those who didn't. Took charge of database management and data analysis for the project.
- Queried medical service's data and anal sex history's data from public and restricted dataset using **MySQL** based on project requirement.
- Merged and Organized the dataset in **SAS**. Cleaned and dealt with missing value. Used SAS **Macros** and procedures such as **PROC WTAHJUST, PROC REGRESS** and **PROC RLOGIST** to adjust the sample weights and conduct multivariate statistical analysis.

College Project: Manufacturing Company's Business Operating System (BOS) Building and Management
- Collected data from different tables including information of product, brand, customer, vendor employee and so on. Discovered relationship between tables and created database using **MySQL**. • Designed and implemented web for the system with **HTML** to make query straightforward.

EDUCATION

Georgia State University   *Atlanta, GA*
M.S. in Statistics and Computer Science   *Aug 2018-May 2020*
- main course: Linear Statistical Analysis, Database, Machine Learning, Data Mining, Big Data Programming, Nonparametric data analysis
- Leadership: Graduate Math Lab Assistant, Treasure & Secretary for American Mathematics Association (GSU) • Publication: Precise and non-invasive detection of liver metastases via imaging the biomarker, Structural Aspects and Prediction of Calmodulin-Binding Proteins

South China University of Technology   *Guangdong, China*
B.S. in Economics   *Sep 2014-June 2018*
- Honors: Second-class scholarship (top16%) in 2016, Third-class scholarship in 2015 and 2017, National College Student Business Plan Competition-Guangdong Province's Bronze Award (Project Leader), National Internet Financial Challenge Competition-top 20 in China (Leader)