Chen (Eric) Xue

(252) 955-4923 | eric.xue@emory.edu | ericxue64.github.io

EDUCATION

Emory University   Atlanta, GA
Bachelor of Science: Quantitative Science (Data Science), Computer Science   Expected: Dec. 2023
- Cumulative GPA**:** 4.0/4.0; 2022 Quantitative Theory and Method Award; Dean's List Scholar
- Coursework**:** Regression Analysis, Probability and Statistics, Data Visualization, Causal Inference & Machine Learning, Database System, Data Structure and Algorithms

Awards & Honors: 2022 ASA Emory DataFest Best Insight Award; 2021 ASA Fall Data Challenge Best Undergrad Overall
Tableau Desktop Specialist Certificate; MySQL Bootcamp; Python for Data Science & ML Bootcamp

SKILLS

Languages & Tools: Python (Scikit-learn, Keras & TensorFlow, PySpark), MySQL, Tableau, RStudio, AWS, Java, REDCap
Techniques: Machine Learning, A/B Testing, Visualization, Statistical Analysis, Big Data, Data Collection

WORK EXPERIENCE

HP Tech Ventures   Remote
VC & Business Analyst Summer Extern   Jun. 2022 – Aug. 2022
- Researched start-up metrics, industry, team, competition, and business model; made presentations on start-ups that have the potential for success, covering product design, market analysis, marketing strategy, and competitor product comparison
- Processed 20,000 start-up metrics data from Crunchbase for machine learning model and Tableau dashboard using Python
- Built an interactive dashboard using Tableau with 4 charts and a table to visualize start-up info and use as filter conditions
- Analyzed the correlation strength between various start-up metrics and total funding amount; split data and developed regression models using Python to predict start-ups' total funding amount with model evaluation metrics

Nell Hodgson Woodruff School of Nursing at Emory   Atlanta, GA
Queen Savvy Lab Data Manager   May. 2022 – Present
- Designed data collection instrument using REDCap; created survey scoring logic and distribution schedule
- Built Tableau dashboard to visualize App feedback; reported feature usage, user experience, and tech acceptance

Emory University   Atlanta, GA
Intro to Statistical Inference TA   Jan. 2022 – May. 2022
- Mentored students on statistical concepts, statistical coding, and interpretation of the results during class and office hour

Cloud Data Tech.   Zhengzhou, China
Technology Intern   Jun. 2021 – Aug. 2021

RELEVANT PROJECTS

Udacity Free Trial Screener A/B Testing   Aug. 2022
- Designed by selecting evaluation metrics, calculating size to achieve desired power, and setting % traffic and test duration
- Performed sanity check, effect size tests, and sign test to analyze the result; recommended not to launch

LendingClub Loan Default Prediction, Deep Learning   Aug. 2022
- Prepared by dropping unrelated variable, feature engineering, filling by mean imputation, scaling, converting to dummy
- Utilized Keras and Tensorflow packages in Python to build Neural Network Models; added Dropout layers and EarlyStopping callbacks to prevent overfit; visualized the loss of training and validation data across each epoch
- Evaluated model, achieving 0.89 accuracy; deployed model on a random customer to predict their loan status

Standard Bank Data Science Virtual Experience Programme, Machine Learning   Jul. 2022
- Prepared data by selecting variables, feature engineering, filling by model prediction, scaling, converting to dummy
- Fit the data to Logistic Regression, Random Forest, KNN, and SVC models; improved model performance by using the lowest error rate K and performing grid search on SVC; Achieved 0.8 accuracy and 0.97 recall
- Created visualizations to present the model to non-technical audiences; explained model, precision, and recall using the analogy of net, fish, and rock; emphasized recall of class 0 is important in the context of giving loan

2022 ASA DataFest@Emory: Elm City Stories Video Game, Best Insight Award (First Place)   Mar. 2022
- Cleaned 1.5G game log data to evaluate the game's ability to predict real-world drug usage using Python
- Analyzed the correlation of participants' performance on the three most relevant mini-games and their drug resistance index from the game log and survey data using Python; found the correlation strength decreased as more weeks into the study
- Proposed the Dunning-Kruger effect as a possible explanation for the observed patterns and used the external dataset to improve the credibility; Represented the team to present data-driven insights to the judges