Huan Chen

2200 Waterview Pkwy, Richardson, TX, 75080

📱 413-210-8815 | ✉ huanchen321@gmail.com | 🔗 linkedin.com/in/huan-chen-574a6782/

Education

University of Texas at Dallas   *Richardson, TX, US*
Ph.D. in Statistics, GPA: 3.7/4.0   *August 2018 - May 2023*
- Research Interest: Statistical Inference, Hypothesis Testing, Survival Analysis, Hazard, Cox Model

University of Massachusetts Amherst   *Amherst, MA, US*
M.S. in Statistics, GPA: 3.9/4.0   *September 2016 - May 2018*
- Relevant Coursework: Machine Learning, Data Visualization and Exploration

Chongqing University   *Chongqing, Chongqing, China*
B.S. in Information and Computing Science, GPA: 3.5/4.0   *August 2011 - June 2015*
- Relevant Coursework: Data Structure and Algorithms, Graph Theory, Optimization Methods

Work Experience

Travelers   *Richardson, TX*
Data Science Leadership Development Program Intern   *June 2022 - Aug 2022*
- Developed a classification model to predict whether an agency can make high revenue (top 20%) to help the marketing team better understand the agency's potential and apply proper corporation strategy.
- Conducted feature engineering including generating agency characteristic features from a large dataset (millions of records) of demographic information, performed exploratory data analysis and feature selection using correlation coefficient, mutual information and PCA.
- Built XGBoost models and conducted extensive hyper-parameter tuning using Optuna. Interpreted feature importance through SHAP values to understand how the model made decisions.
- The model achieved more than 80% prediction accuracy while 70% of high-revenue agencies were identified. Compared with the baseline model, it achieved 20% increase in precision and 70% increase in recall.
  Technical Skills: Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Optuna), AWS SageMaker, Git.
  Soft Skills: Teamwork, Communication, Code Documentation, Leadership, Time Management, Presentation skills.

The Filbey Lab   *Dallas, TX*
Summer Intern   *June 2019 - August 2019*
- Conducted statistical modeling and programming for the project which analyzed the reward system dysfunction for cannabis users using experimental data.
- Quantified the task-modulated variability in dynamic functional connectivity and used ANOVA model to study the main effect of task conditions within subjects for each functional connectivity measure.
- As dynamic connectivity was sparsely studied, the results complemented previous knowledge on static functional connectivity and potentially added systematic evidence to the known models of substance use disorder from a specific case of cannabis use.
  Technical Skills: MATLAB, R, Excel, Statistical Methods.
  Soft Skills: Teamwork, Communication, Logical Thinking.

Projects

Goodness-of-Fit Test for Cox Model under Monotonicity Constraints   *Richardson, TX*
Research Project, University of Texas at Dallas   *June 2022 - Present*
- Proposed hypothesis test for the linear functional form of the covariate in the Cox model against the alternative that the hazard is monotone in covariate since there are only graphical approaches proposed until now.
- Performed simulation study to show the well-controlled type-I error and consistency of the test. Used the recently proposed most efficient pseudo iterative convex minorant algorithm for estimation, and used the bootstrap method to get the critical value of the test statistic.
- Compared with the test without monotonicity constraints, the proposed test has higher power. It is also the first one in literature to help researchers choose the appropriate one between Cox model and the monotone hazard model objectively based on their data.
  Technical Skills: R, RMarkdown, Git, Hypothesis Testing, Survival Analysis, Statistical Inference, Optimization.
  Soft Skills: Problem Solving, Presentation Skills, Code Documentation.

Customer Behavior Prediction   *Richardson, TX*
Machine Learning Project, University of Texas at Dallas   *March 2019*
- Developed models to predict whether an individual will purchase a caravan insurance policy.
- Applied KNN, random forest, logistic regression with ridge or lasso regularization on the labeled data for prediction. Compared methods according to sensitivity, specificity and overall misclassification rate.
- The logistic model using the maximum likelihood method outperformed since it had the smallest misclassification rate and the largest specificity.
  Technical Skills: R, RMarkdown, RShiny.
  Soft Skills: Logical Thinking, Report Writing.

Skills

Programming	Python (Pandas, NumPy, Scikit-learn, Optuna etc.), R(ggplot, parallel), MATLAB, SQL.
Miscellaneous	Linux, Git, AWS SageMaker, Microsoft Office, LaTeX(Overleaf/R Markdown).
Soft Skills	Time Management, Teamwork, Problem-solving, Documentation, Engaging Presentation.