

NAM HOAI NGUYEN

Houston, TX, 77030

832-758-3848

hn17@rice.edu

Linkedin: nam-nguyen-rice23

EDUCATION

- *PhD. Statistics, Rice University*, 2023 (Expected)
- *MPhil. Finance, University of Cambridge*, 2016
- *BS. Mathematics, Imperial College London*, 2015
Department rank: 7/~300; Awarded the IMA Prize for outstanding performance in Statistics.

TECHNICAL SKILLS

- **Programming languages:** R, Python, SQL, MATLAB, SAS, C++, Mathematica, Linux, git, LaTeX • **Software:** MS Office, PowerBI
- **Libraries/Frameworks:** Tensorflow, Numpy, Matplotlib, Pandas, Scikit-learn, folium, ggplot2, Rcpp, Shiny

EXPERIENCES

MD Anderson Cancer Center

Houston, TX

Graduate Research Assistant

May 2020 - present

- **Project:** *Bayesian estimation of a joint semi-parametric recurrent event model of multiple cancer types for individualized risk prediction (Advisor: Dr. Seung Jun Shin & Dr. Wenyi Wang)*
 - Developed a novel risk prediction model that allowed for recurrent cancers and competing risks from multiple cancer types.
 - Implemented the model in **R**. **Rcpp** was used for faster performance.
 - Performed MCMC on a high performance computer cluster (**Linux**) to estimate model parameters.
 - Integrated the new model to the lab's open-source LFSPRO library and the **Shiny** app for used by genetic counselors.
- **Project:** *Validation of Cancer-specific and Multiple-primary-cancer-specific Risk Prediction Models on Clinically Ascertained Family Data (Advisor: Dr. Wenyi Wang)*
 - Collaborated with genetic counselors to assess and mitigate the effect of missing patients' ages and incomplete cancer history.
 - Evaluated the performance of the lab's LFSPRO library on a clinically ascertained dataset on the basis of AUC and O/E ratio.

Department of Statistics, Rice University

Houston, TX

Graduate Research Assistant

Dec 2019 - present

- **Project:** *Mathematical Modeling of Stem Cells and Their Descendants through Multi-type Age-dependent Branching Process (Advisor: Dr. Marek Kimmel)* - submitted to **Stochastic Models with first authorship**
 - Derived closed-form expression for the evolution of cell population over time under different cell-type-specific dynamics.
 - Performed asymptotic analysis on the probability of extinction and the rate at which the population approaches extinction.
 - Performed simulation in **MATLAB** to confirm theoretical results.

Data to Knowledge (D2K) Lab, Rice University

Houston, TX

Graduate Student Researcher

Aug 2020 - Dec 2020

- **Project:** *Satellite-based estimation of air pollution response to the COVID-19 pandemic (Advisor: Dr. Dan Cohan & Dr. Dan Kowal)*
 - Produced interactive data visualizations using **R Shiny** to observe the changes in NO₂ in response to COVID-19 lockdown.
 - Built Generalized Additive Models and broken-stick models to describe how NO₂ changed over time and to make prediction.
 - Presented the findings in a data science workshop to industry professionals and to professors at Rice University.

Department of Statistics, Rice University

Houston, TX

Graduate Teaching Assistant

Aug 2019 - Dec 2020

- Courses: Introduction to Statistical Machine Learning; Probability and Statistics; Regression and Linear Models.
- Led weekly recitation sessions, graded homework and communicated with instructors to keep track of students' progress.

PROJECTS

Kernel Density Estimation on Decision Trees with Missing Values

- Developed and implemented in **R** a new density-based approach to deal with missing values when using a fitted decision tree for prediction. The method outperformed mean imputation and imputation by regression on a real data set.

Statistical Analysis of Astronomical Data

- Stored and managed a large database using **SQLite**. The primary data set from NASA JPL had more than 800,000 entries.
- Applied ML algorithms to predict orbital types with ~ 96% accuracy. Used **Shiny** to interactively visualize the decision boundaries.
- Built a ML model in **R** that predicted potentially hazardous asteroids. The model achieved an F1 score of 0.89.

Machine Learning for Loan Default Prediction

- Built a gradient boosted decision tree in **MATLAB** to predict the outcomes of new loans. Ranked 3rd out of 52 students on Kaggle on the basis of out-of-sample F1 score.

CERTIFICATES

IBM Professional Certificate in Data Science (IBM) • Deep Learning Specialization (deeplearning.ai) • Machine Learning (Stanford) • Fundamentals of Computing Specialization (Rice) • A/B Testing (Google) • SAS Programmer Professional Certificate (SAS Institute)