NAVEED MERCHANT, PH.D

281-701-1544 | naveedmerchant2@gmail.com | github.com/naveedmerchant

MODELING & ANALYSIS | DATA QUERYING | COMMUNICATIONS | BUSINESS ACUMEN

QUALIFICATION SUMMARY

I am a statistician with experience in applying techniques to make speedy computations on very large data sets. My research involves looking at efficient screening methods to help large data set applications run quickly and effectively. I help companies make decisions using historical data, in support of decisions to develop profitable strategies. I enjoy collaborating with others to gain insight from data, as well as creating tools to make it easier for everyone to access the same data and methods in the future.

KEY SKILLS

✔ **Modeling and Analysis**	✔ **Database Querying**	✔ **Technical Communications**
✔ R (parallel computing, Repp)	✔ SQL	✔ LATEX
✔ Python (PyTorch)	✔ Spark	✔ Rmarkdown
✔ SAS	✔ AWS	✔ Microsoft Office, Excel VBA

RELEVANT PROFESSIONAL EXPERIENCE

Data Scientist, Climate Corporation **Jan 2022-Present**
COLABERRY Houston, TX

- Explore, query, manipulate and model big data using statistical and Bayesian modeling with Python;
- Extract, transform and load (ETL) abstract data structures to model the effect of disease on crop systems;
- Select appropriate modeling techniques and data visualization for large datasets to conduct hypothesis testing on the efficacy of disease on crop systems;
- Evaluate data quality and discover data limitations and report to the team and senior management.

Statistical Modeling Intern **Apr-Dec 2021**
MAXIS-IT Edison, NJ

- Applied semi-supervised and unsupervised machine learning techniques to detect a person's stress given real-time biological information;
- Processed large biological datasets in support of machine learning applications for the above;
- Responsible for recruiting new team members, including: coordinated with HR, reviewed resumes, drafted technical interview questions and conducted/evaluated technical interviews.

Consulting Center Research Assistant **Aug 2020-Jun 2021**
TEXAS A&M UNIVERSITY College Station, TX

- Using R, collaborated with A&M faculty in developing customized statistical models to analyze experimental data;
- Guided and adjusted experiment design of faculty's experiments to minimize cost and preserved power for inference with experiments; and
- Reviewed grant proposals by faculty to ensure funding budget supported sufficient data requirements to fulfill research goals in a variety of conditions.

Covid-19 Modeling Research Assistant **Jun-Dec 2020**
TEXAS A&M UNIVERSITY College Station, TX

- Queried, filtered, loaded and modeled real time COVID-19 data to inform Texas A&M University COVID-19 policy decisions;
- Assessed, selected, and tested different statistical models to examine impact of certain COVID-19 events on the incidence rate;
- Extended statistical models to make reliable short-term predictions of the number of COVID-19 incidences in the future;
- Leveraged Rshiny to deploy and broadcast predictions to the public and for reporting results.

Instructor **Aug-Dec 2019**
TEXAS A&M UNIVERSITY College Station, TX

- Developed and delivered STAT 201 introductory statistics course designed to introduce the concept of sampling to a class of 71 undergraduate students. Overall student rating: 4.22 out of 5;
- Increased engagement by leveraging online learning platforms

General Motors Research Assistant **Jan-Jun 2019, Jan-Jun 2020**
TEXAS A&M UNIVERSITY College Station, TX

- Optimized algorithms to enable calculations on huge volumes of data regarding car operations and its correlation to future mechanical flaws;
- Developed and delivered a tool with R and Python to query weather data from 3rd party sources and merge it within the General Motors database. The weather tool was documented for independent use in the future;
- Engineered informative features to use in large time series analysis to explain malfunctions in conveyor belts; and
- Collaborated with the Statistics and Computer Science departments to evaluate Matrix Profile's performance in unsupervised learning of anomalies in the conveyor belts.

PAPERS, PRESENTATIONS AND RESEARCH PROJECTS

Screening Methods for Classification Based on Non-parametric Bayesian Tests | Aug 2022
Collaborator: Dr. Jeffery Hart, Texas A&M University

- A new screening method for choosing variables under classification is proposed and explored.
- The usefulness of the screening method (and others) are established by examining the improvement of common classification models after screening is applied.
- We use our method to detect important pixels for recognizing handwritten digits and detecting useful genes for discriminating between types of lung cancer.
- A pre-print is available upon request.

A Bayesian Motivated Two-sample Test Based on Kernel Density Estimates | Aug 2022
Collaborator: Dr. Jeffery Hart, Texas A&M University

- A Bayesian two-sample test that extends the framework in "Use of Cross-validation Bayes Factors to Test Equality of Two Densities".
- This test loses some interpretability of the Bayes factor in exchange for obtaining an exact frequentist type p-value, losing some hyper parameters, and an overall increase in computation speed.
- This paper has been accepted and published by Entropy.

COVID-19: Short Term Prediction Model Using Daily Incidence Data | Apr 2021
Main Collaborator: Dr. Hongwei Zhao, Texas A&M University

- A new method is proposed and used for forecasting the number of daily incidences of COVID-19, given the R_0 rate is roughly constant over short periods of time. We apply the method to estimate the number of future incidences in the Texas state and Brazos Valley county.
- Developed a dashboard that produces real-time predictions.
- This paper has been accepted and published by PLOS-ONE.

Effects of Gene Expression and LSO on Potato and Tobacco | Dec 2021
Collaborators: Julien G. Levy, Texas A&M University; Azucena Mendoza-Herrera, Texas A&M University; Cecilia Tamborindeguy, Texas A&M University; Zhiqian Pang, University of Florida; and Nian Wang, University of Florida

- This is a multiple department collaboration project – my primary role was consulting and guiding analysis of the data in a context where it was not clear how variables interacted.
- We characterize symptoms of different types of "LSO" (a plant pathogen) on tobacco, as well as explore the effect of a gene's expression in LSO infection, psyllid (insect) infestation and different plant responses.
- The full name of the paper is: "A salicylic acid hydroxylase gene from `Candidatus Liberibacter' manipulates plant responses and affects resistance to potato psyllid and `Candidatus Liberibacter solanacearum'". Currently in the process of being accepted and published.

Use of Cross-validation Bayes Factors to Test Quality of Two Densities | **Mar 2020**
Collaborators: Dr. Jeffery Hart, Texas A&M University and Dr. Taeryon Choi from Korea University

- Developed a non-parametric Bayesian test for testing equality of two distributions. Research has been implemented on increasing computation speed, and applying our method to feature selection for classification and fraud detection.
- A GitHub Repository with an R package that implements this research is available. A paper that details the procedure has been submitted to Bayesian Analysis.
- An invited presentation regarding this work was given at the *2021 World Meeting of the International Society for Bayesian Analysis.*

Experimental Rankings of Soybean Plants | **Apr 2020**
Collaborators: Myeongjong Kang, Zhao Tang Luo and Dr. Loren Vega, Texas A&M University

- We developed a new method to assign holistic ranks to the treatments, based on how they impacted several different factors of growth on soybean plants in collaboration with the AgriLife department.

Analysis of Taxi-Cab Trends Over Time | **Apr 2018**
Collaborators: Zhao Tang Luo, Dongbang Yuan, Anupam Kundu and Dr. Jianhua Huang, Texas A&M University

- Identify how taxi cab companies have evolved from 2013-2019, and how they have been impacted by Uber and Lyft's emergence in the city of Chicago.
- We analyzed four years of data (over 20 GB), and extracted information on which taxi cab companies survived Uber and Lyft and their successful techniques.
- Analysis is mostly exploratory, but was on a very large scale.

EDUCATION

DOCTOR OF PHILOSOPHY, Statistics *May 2022*
Texas A&M University College Station, TX

BACHELOR OF SCIENCE, Applied Mathematics *May 2017*
Texas A&M University College Station, TX