ROCHITA DAS

|College Station, TX | +1 (979) 739 4890 | rochita.das.stat@gmail.com | LinkedIn | Website | GitHub | Medium |

EDUCATION

Texas A&M University, College Station, TX, USA — Aug 2018 - Aug 2023
Ph.D. in Statistics — GPA: 3.8/4
Research Interest: Bayesian, Multivariate, Big Data, Machine Learning, Stochastic Process, Biostatistics

University of Calcutta, Kolkata, WB, India — July 2014 - June 2016
M.Sc. in Statistics — GPA: 3.32/4

St. Xavier's College, Kolkata, WB, India — July 2011 - May 2014
B.Sc. in Statistics — GPA: 3.29/4

TECHNICAL SKILLS

Programming Skills	R (RCPP, Parallel computing, R-Package), Python(TensorFlow, Keras, PyTorch, Scikit-learn, Pandas, Numpy), SAS, C++, SPSS, Matlab, Minitab, JMP, Lindo, Git
Database and Visualization	SQL, NoSQL, MS Excel, MS Access, Tableau, Power BI
Clouds	Databricks, Google Cloud, Apache Spark, Docker

WORK EXPERIENCE

H-E-B — San Antonio, Texas
Planning & Analysis Data Science Intern — May 2021 - July 2021

- Developed a Predictive ensemble modelling using XGBoost, NeuralNet, GLM, Ridge, ElasticNet, Lasso, Support Vector Machine, K nearest neighbour in SQL, Python and Databricks.
- Improved the prediction error by 35%, reduced late deliveries by 33% and increased the number of deliveries per hour by 17%.
- Volunteered in the preparation of a R package related to AB testing using Bayesian approach.

Novartis Health Care Pvt. Ltd. — Hyderabad, India
Associate Statistician — Aug 2016 - May 2018

- Worked as Trial Statistician in three Oncology clinical trials of Regional Europe, supervised a team of 3 members and mentored new joinee for one year.
- Performed cross functionally and executed Statistical Analysis, supported in decision making by Clinical Trial Team, Medical Team and provided operational input in CSR (Clinical Study Report)
- Analyzed a retrospective study, proposed several date imputation rules and made the definitions of variables mutually exclusive prioritizing some conditions and considering some external factors.
- Achieved more than 95% customer satisfaction for each of the 3 projects in Novartis.

CONSULTANCY EXPERIENCE

Social Science Department, Texas A&M University — Aug 2021 - Dec 2021

- Analysed the project *'Public opinion survey on infrastructure in Texas'*.
- Variable selection is done using XGBoost, Lasso and an ensemble modelling is done with Bayesian decision trees and Weighted Generalised Linear Model (WGLM).

Nutrition Department, Texas A&M University — Jan 2020 - May 2020

- Analysed the project *'Obesity and Postmenopausal Breast Cancer'*.
- Analysed the data considering Zero-inflated mixed effect Gamma Regression and also considered Panel data analysis.

PROJECTS, POSTERS AND COMPETITIONS

Ph.D. Project 1: Bayesian Analysis for Multivariate Multinomial Probit Models

- Implemented Parameter-extended Metropolis-Hastings algorithm to sample covariance matrix and considered Tempered Gibbs method to draw sample form truncated marginal multivariate t distribution.

Ph.D. Project 2: Bayesian Nonparametric Causal Discovery

- Implemented Dirichlet Process mixture (DPM) of triangular densities as well as mixture of Asymmetric Laplace distribution Laplace distribution in Structural Equation Model (SEM). The model is extended to multivariate level to implement the Bayesian DAG modelling.

M.Sc. Project: A Study on Time-series and Regression Analysis of Yahoo share of Japan & USA

- Implemented VAR model considering some external factors and fitted Simultaneous Equation Model (SEM) to capture instantaneous causality and then predicted future values using Two Stage Least Squares(2SLS) model.

B.Sc. Project: A study on Prediction of Opening Weekend Box Office Collection

- Researched extensive regression techniques and implemented them to find prediction of collection of a movie using different covariates(factors) such as genres, production budget, releasing time etc.

Poster: Modelling Small Sample Discrete data using Exact Distributions

- Presented the poster in IISA 2017 to show with real life example how exact regression works better than usual Logistic/Poisson regression when the sample size is very small or the data are sparse, skewed or highly stratified.

Datathon: City Search Engine | Link for an interactive R Shiny app

- Identified important factors which play crucial role to select a city for move in and proposed fast, scalable search algorithm to recommend relevant cities based on user's input.

TEACHING EXPERIENCE

Department of Statistics, Texas A&M University, College Station, TX

- STAT 201 (Elementary Statistical Inference) Aug 2020 - Dec 2020
- STAT 302 (Statistical Methods) June 2020 - Aug 2020

AWARDS

- Stood 1st in all the four semesters in Masters in Statistics and recipient of Gold Medal.
- Recipient of Inspire Scholarship (Awarded by Govt. of India to only top 1% students).

LEADERSHIP

- Monitored mentally challenged children in IICP (organized by NSS) for 2 weeks.
- Organised blood donation camp with approx 100 doners in my locality.
- Mentored a first year Ph.D. student.
- Member of Diversity Committee, Texas A&M

ACTIVITIES AND INTERESTS

- Member of American Statistical Association (ASA)
- Member of Machine learning Group, Texas A&M Institute of Data Science (TAMIDS).
- Hiking, Triathlon, Adventure activities, Martial arts.