

SRISHTI SAHA

srishtisaha.2809@gmail.com; +1-2065811905

Portfolio: <https://srishtis.github.io/>; GitHub: github.com/srishtis;

LinkedIn: [linkedin.com/in/srishtisaha28](https://www.linkedin.com/in/srishtisaha28)

EDUCATION

Masters' in Interdisciplinary Data Science, Duke University

Expected May 21

- **Courses:** Machine Learning, Data Management Systems, High Dimensional Data Analysis, Data Visualization, Data Ethics, Text Analysis, Data Analysis on Cloud, Bayesian Statistics, Deep Learning, Topological Data Analysis, Causal Inference, Social Network Analysis
- **Key Projects:** [Solar PV classification in Aerial Imagery](#), [Prediction of popularity of TedTalks from their transcript](#), [Application for image caption generator](#), [Classification using Topological Features of LEGO parts images](#), [Deep learning Approach for Question and Answering Systems](#), *Capstone Project (ongoing):* Participatory agent-based Bayesian modeling for community hunting mgmt.

B.Tech (Electronics & Communication), Manipal Institute of Technology

2011-15

- **Relevant Projects:** Create and evaluate ANFIS models using MATLAB for gender classification of speech signals ([link](#))

EXPERIENCE

Quantitative Analytics Intern, Wells Fargo

Jul 20 – Aug 20

- Worked with customer complaints data to study language drift and its effect on performance of classification models. Quantified the correlation between language complexity across open-source customer complaints data and model evaluation parameters derived from mortgage payments that helped future research within the organization (*NLP, Classification*)

Data Science Summer Project Member, Duke University (Fleet Mgt. Ltd.)

May 20 – Aug 20

- Helped Fleet Management Ltd. to establish its first set of analytical processes by predicting on-ship incident trends using their inspection reports (*NLP, Time-Series Forecasting*)

Decision Scientist, Mu Sigma Business Solutions Pvt. Ltd., India

May 19 - Jul 19

- Developed an interactive Marketing Mix module for enabling data scientists to develop and use models for simulating & analyzing marketing scenarios, thus reducing their average effort from about 40 hours to 2 hours resulting in time savings of 95% (*Panel Regression, RShiny, R*)
- Created Python modules for comparing models using A/B tests and multi-armed bandit tests, and building the data engineering pipeline for data streams from Kafka clusters on Kubernetes for a model operationalization product (*A/B testing, Multi-armed bandits, Kubernetes, Python*)
- Wrote a whitepaper on the application of Information Theory in Machine Learning ([link](#))

Trainee Decision Scientist, Mu Sigma Business Solutions Pvt. Ltd., India

Jul 15 – Jul 16; May 17 - May 19

- Lead a team of 3 to emulate pricing decisions of automotive parts by evaluating their price elasticity, competitive pricing strategies, product lifecycle, and creating an interactive & prescriptive price recommendation dashboard (pricing toolkit) using RShiny, reducing pricing cycle time of the North America pricing team by 40%. Built a test/control and lift analysis module in the toolkit to audit past price changes (*A/B testing, regression, clustering, RShiny, R, SQL*)
- Classified potential automotive dealer fraud using dealer sales trends and purchase loyalties, helping the Customer Care & Aftersales team redesign their dealer partnerships (*Benford's Law, R, SQL, Regression*)
- Used Topic Modelling to predict vehicle recalls for Safety Analytics Division using customer feedback data resulting in reduction of recall time by 57% (*NLP, MS Access, R*)
- Created a predictive model for a retail giant evaluating the performance of newly launched products and built a portfolio optimization module for an interactive toolkit used by purchase managers across 13 countries with quarterly total inventory of about \$2,000M (*Regression, Python, Hive, SQL, basic JavaScript, HTML*)
- Developed an ensemble of classification models as a POC for an insurance firm to identify misrepresented catastrophic claims using textual, temporal, and geographic data (*Classification, Python*)
- Worked on a POC to predict machine failures (rare events) for an energy firm using classification models to reduce failure instances by ~70% (*Deep Learning (Autoencoder), Classification, Oversampling techniques, Python, Tensorflow, Keras*)

Information Developer, SAP Labs, India

Aug 16 – Apr 17

- Created, built, tested, and documented workflows for SAP S/4HANA (*Python, HTML*)
- Automated functions for data flow from a content-management system to package products for the SAP marketplace (*DITA CMS, HTML, CSS*)

SKILLS

Technology:

- **Database Programming:** SQL (Oracle, PostgreSQL, MySQL), MS Access
- **Analytical Programming:** Python, R, MATLAB, MS Excel, SAS
- **Deep Learning:** PyTorch, Tensorflow
- **Data Visualization:** RShiny, Tableau
- **Cloud Platforms:** GCP, AWS
- **Big Data Tech:** Hive, Hadoop (basic)
- **Statistical Tools and Applications:** Regression, Classification, Panel Regression, Clustering, Time Series Forecasting, Text Mining and Analytics (NLP), Hierarchical Bayesian Modelling, Agent-based Modelling

ACHIEVEMENTS

Analytical:

- Spot awards for [Pricing Toolkit](#), [Emerging Analytics \(predicting vehicle-recalls\)](#) & [contribution to POCs](#)
- **1st position** in in-class Kaggle competition (Solar PV classification in Aerial Imagery)
- **Semi-finalist** in MIT COVID-19 Datathon (Project: How have public concerns evolved with the pandemic?)

Certification:

- Mu Sigma Certified Decision Scientist

Mentorship Experience:

- Trained Mu Sigma employees in Python & Machine Learning

Research/ Poster Presentations:

- [Predictive Modeling of Mechanical Failures at Sea at Rhodes Information Initiative at Duke](#) (2020)
- Evolution of Public Concerns in NY along the course of the COVID-19 pandemic ([NSF Student Conference 2021](#))

EDUCATIONAL EXPERIENCE

Research Assistantship:

- Built models for identifying text recycling in STEM research reports
- Analyzed research patterns at Duke Global Health Institute by designing metrics assessing past publications and co-authorships

Teaching Assistantship:

- Graduate Teaching Assistant for Fraud Analytics (using R)
- Teaching Assistant for Sourcing Data for Analytics (Python).
- Teaching Assistant for Modeling Process & Algorithms (Python)

