

# Zubia Mansoor

☎ (+1) 778-869-9829 | ✉ [zmansoor@sfu.ca](mailto:zmansoor@sfu.ca) | 🏠 [zubiamansoor.github.io](https://zubiamansoor.github.io) | 📷 [zubiamansoor](#) | 📺 [zubiamansoor](#)

## Skills Summary

- Programming Tools** R, Python (pandas, scikit-learn, keras, seaborn), SQL, Matlab, LaTeX, Git, Bash, Minitab
- Statistical Methods** Bayesian & frequentist inference, stochastic processes, survey methodology, monte carlo, ANOVA, sampling theory
- Machine Learning** Neural networks, ensembles, deep learning, random forest, bayesian nonparametrics

## Education

### Simon Fraser University

*Burnaby, Canada*

MASTER OF SCIENCE IN STATISTICS, CGPA : 4.27/4.33

*Sep 2019 - Jul 2021*

- Dean's Graduate Fellowship
- Statistical Society of Canada student travel award
- Community-Engaged Research Initiative Graduate Fellowship
- Mitacs Research Training Award

### St. Xavier's College

*Kolkata, India*

BACHELOR OF SCIENCE (HONS.) IN STATISTICS, MINOR IN MATHEMATICS AND ECONOMICS, FIRST CLASS WITH DISTINCTION

*Jul 2015 - Jun 2018*

## Work Experience

### Quantitative Analyst Intern

*Toronto, Ontario*

CAPITAL MARKETS | ONTARIO TEACHERS' PENSION PLAN

*Jan 2021 - Apr 2021*

- Work in a quantitative trading team to help make investment decisions using statistical models and data analytics
- Flagged 98% of the outliers in financial earnings data by developing an unsupervised anomaly detection system using Python and SQL queries

### Bioinformatics Research Intern

*Burnaby, Canada*

COVID-19 RAPID RESPONSE GRANT | CANADIAN STATISTICAL SCIENCES INSTITUTE (CANSSI)

*May 2020 - Sep 2020*

- Studied the effects of gene mutation on COVID-19 susceptibility after controlling for socioeconomic factors in UKBiobank human genome data
- Discovered significant association between dementia and COVID-19 susceptibility by constructing contingency tables
- Built pipeline to preprocess high dimensional data and compute the genetic correlation between two phenotypes
- Cut ~ 10 minutes per run by deploying a series of Python and Bash scripts on a parallel computing network

### Data Science Intern

*Kolkata, India*

STATISTICAL QUALITY CONTROL DIVISION | INDIAN STATISTICAL INSTITUTE

*May 2018 - Oct 2018*

- Collaborated on-site with statisticians and chemical engineers to improve the quality control process at a paper manufacturing company
- Reduced annual costs by 30K CAD by building a hybrid SVR-RT ensemble model to optimize chemical consumption
- Presented technical findings conducted using Python and R to various stakeholders and research staff
- Co-author in publication: "A hybrid regression model for water quality prediction" - Chakraborty, T., Chakraborty, A.K. & Mansoor, Z. Springer (2019)

## Relevant Academic Projects

### Image Classification using WasteNet - 1st Place [View Project](#)

*Vancouver, Canada*

VANCOUVER DATAJAM HACKATHON 2020

*Sep 2020 - Sep 2020*

- Led a 9-member team at a one-day virtual hackathon to correctly classify images of waste items for proper garbage disposal in Greater Vancouver
- Compared tuned SVM and CNN's ability to predict waste classification and achieved a test accuracy of 93% using tuned ResNet-15 in PyTorch
- Supplemented testing data with pictures of trash from diverse environments (e.g. street, beach) to assess the chosen model's generalization

### Predicting Mortgage Default Risk - Finalist [View Project](#)

*Burnaby, Canada*

SFU BUSINESS ANALYTICS HACKATHON

*Nov 2020 - Nov 2020*

- Solved a real-world business problem in under 3 hours by building an ensemble classifier to predict mortgage default for a fintech startup
- Achieved an AUC of 73% on a hold-out test set that averaged the predicted class probabilities by a neural net and random forest
- Used k-means clustering to identify key customer demographics to target in the upcoming advertising campaigns
- Presented findings to batches of judges sequentially, incorporating the judges' feedback before pitching to the next round

### Examining Out-of-Distribution Data using Deep Neural Networks

*Burnaby, Canada*

MASTER'S THESIS, SIMON FRASER UNIVERSITY

*Feb 2020 - Present*

- Compare the robustness of humans and current deep neural networks on out-of-distribution (OOD) data for image recognition tasks
- Employ transfer learning methods in deep neural networks to make OOD performance more human-like
- Parallelized jobs to create 8 different image distortions using 1.5 TB of ImageNet data to evaluate the robustness of different CNN architectures